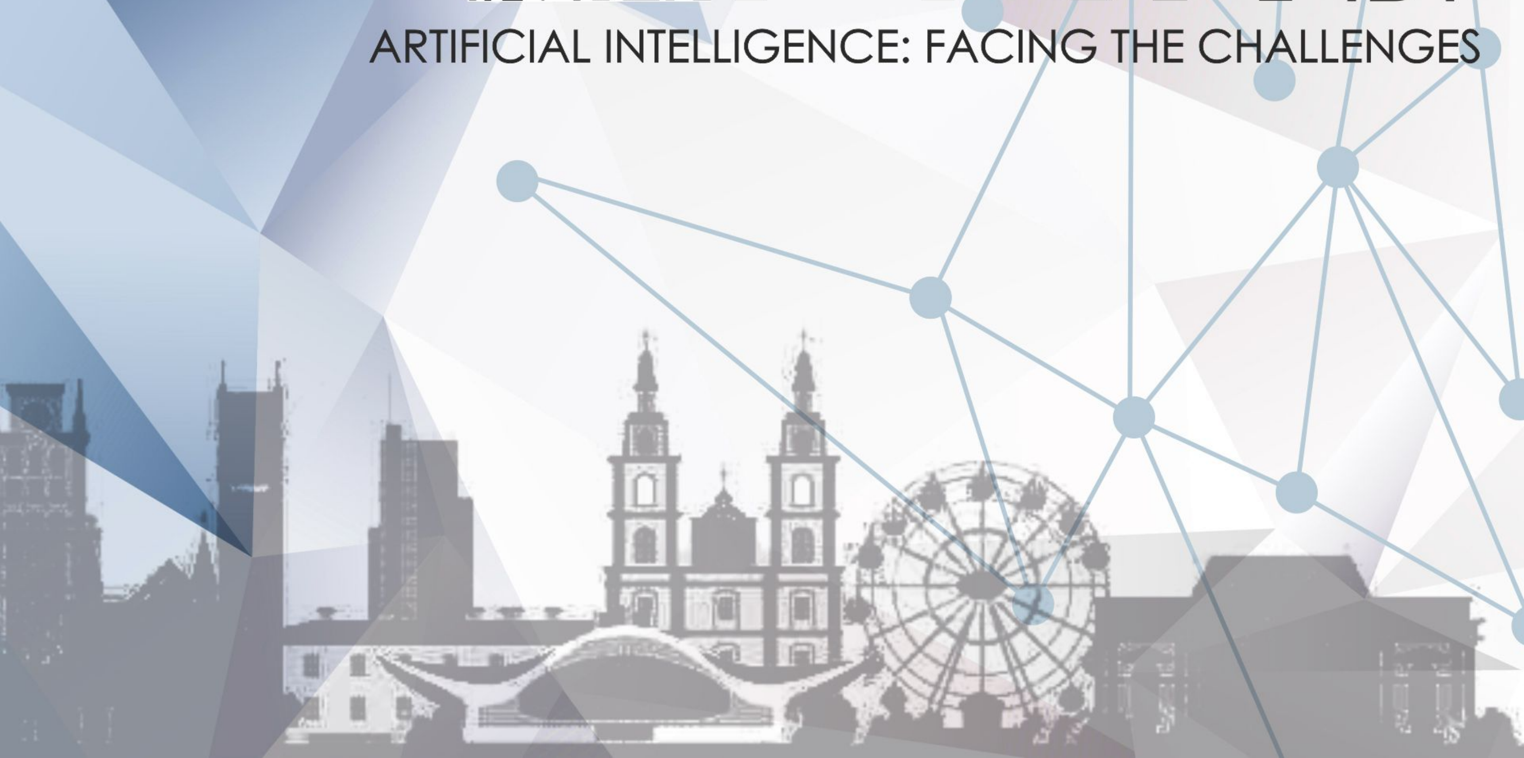
PROCEEDINGS OF THE 15TH INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION AND INFORMATION PROCESSING

21-24 SEPTEMBER • MINSK • ONLINE EDITION

2021 PRIP.BY

ARTIFICIAL INTELLIGENCE: FACING THE CHALLENGES



UNITED INSTITUTE OF INFORMATICS PROBLEMS
OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

PRIP'2021

PATTERN RECOGNITION AND INFORMATION PROCESSING

ARTIFICIAL INTELLIGENCE: FACING THE CHALLENGES

PROCEEDINGS
OF THE 15th INTERNATIONAL CONFERENCE

21–24 September 2021, Minsk, Belarus

Minsk
UIIP NASB
2021

UDC 004.93'1; 004.932

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The Proceedings include papers of the 15th International Conference on Pattern Recognition and Information Processing, which took place on September 21–24, 2021 in Minsk (Belarus) in an online format. Released by the organizers of the PRIP'2021 Conference and the Belarusian Association for Image Analysis and Recognition.

The Conference materials will be useful for students, undergraduates, and experts in the following areas: pattern recognition and image analysis, knowledge processing and representation, knowledge-based decision support systems, fuzzy mathematics and systems, pattern recognition and image analysis applications, Artificial Intelligence, including Neural Networks and Deep Learning.

The articles have passed multi-factor peer review and are printed in the form provided by the authors. Additional information, as well as online presentations, including keynotes and speeches of invited speakers in the framework of a special GÉANT Session, can be viewed on the Conference website https://www.prip.by/2021/ and the PRIP Conference Youtube channel “PRIP Conference”.

The Conference was held under the General Data Protection Regulation, all participants confirmed their consent to publicly display and record their speeches for the Conference's needs.

Сборник включает доклады 15-й Международной конференции «Распознавание образов и обработка информации», которая состоялась 21–24 сентября 2021 г. в Минске (Беларусь) в онлайн-формате. Подготовлен к выпуску организаторами конференции PRIP'2021 и Белорусской ассоциацией по анализу и распознаванию изображений.

Доклады конференции будут полезны студентам, магистрантам и специалистам в следующих областях: распознавание образов и анализ изображений, обработка и представление знаний, системы поддержки принятия решений, основанные на знаниях, нечеткая математика и системы, приложения распознавания образов и анализа изображений, искусственный интеллект, включая нейронные сети и глубокое обучение.

Статьи прошли многофакторное рецензирование и печатаются в виде, представленном авторами.

Дополнительную информацию, а также онлайн-презентации, включая пленарные доклады и выступления приглашенных докладчиков в рамках специальной секции (GÉANT Session), можно посмотреть на сайте конференции https://www.prip.by/2021/ и на Youtube-канале “PRIP Conference”.

Конференция была проведена в соответствии с Общим регламентом по защите данных, все участники конференции подтвердили согласие на публичную демонстрацию и запись своих выступлений для целей, предусмотренных конференцией.

Editors:

Alexander Tuzikov,
Alexei Belotserkovsky,
Marina Lukashevich

ISBN 978-985-7198-07-8

PRIP’2021 ORGANIZERS

- UNITED INSTITUTE OF INFORMATICS PROBLEMS OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

in cooperation with

- BELARUSIAN STATE UNIVERSITY
- BELARUSIAN STATE UNIVERSITY OF INFORMATICS AND RADIOELECTRONICS

SUPPORT AND ENDORSEMENT

- THE NATIONAL ACADEMY OF SCIENCES OF BELARUS
- THE INTERNATIONAL ASSOCIATION FOR PATTERN RECOGNITION (IAPR)
- THE BELARUSIAN ASSOCIATION FOR IMAGE ANALYSIS AND RECOGNITION (BAIAR)
- THE ASIA-PACIFIC ARTIFICIAL INTELLIGENCE ASSOCIATION (AAIA)
- BASNET, NATIONAL RESEARCH AND EDUCATION NETWORK
- GÉANT ASSOSIATION
- BUSINESS INTELLIGENCE QUANTS (BIQUANTS)

PRIP’2021 COMMITTEE

CONFERENCE CHAIRMAN

Prof. **Alexander Tuzikov** (Belarus)

CONFERENCE VICE-CHAIRMAN

Prof. **Sergey Ablameyko** (Belarus)

PROGRAM COMMITTEE CO-CHAIRMAN / CHIEF EVENT OFFICER

Dr. **Alexei Belotserkovsky** (Belarus)

PROGRAM COMMITTEE CO-CHAIRMAN

Dr. **Marina Lukashevich** (Belarus)

CONFERENCE INTERNATIONAL COMMITTEE

(in alphabetical order)

Astsatryan, Hrachya (Armenia)

Belokonov, Igor (Russia)

Bogonikolos, Nikos (Greece)

Chemeris, Alexander (Ukraine)

Deserno, Thomas M. (Germany)

Doudkin, Alexander (Belarus)

Dziech, Andrzej (Poland)

Frucci, Maria (Italy)

Gallo, Luigi (Italy)

Golenkov, Vladimir (Belarus)

Golovko, Vladimir (Belarus)

Gurevich, Igor (Russia)

Hiromoto, Robert (USA)

Kharin, Yuriy (Belarus)

Kovalev, Vassili (Belarus)

Krasnoproshin, Viktor (Belarus)

Madani, Kurosh (France)

Marcelli, Angelo (Italy)

Mariage, Jean-Jacques (France)

Nedzved, Alexander (Belarus)

Nystrom, Ingela (Sweden)

Piuri, Vincenzo (Italy)

Roth, Hubert (Germany)

Sachenko, Anatoly (Ukraine)

Sanniti di Baja, Gabriella (Italy)

Shmerko, Vlad (Canada)

Starovoitov, Valery (Belarus)

Subbotin, Sergey (Ukraine)

Tatur, Mikhail (Belarus)

Uchida, Seiichi (Japan)

Yanushkevich, Svetlana (Canada)

Ye, Shiping (China)

Yingke, Xu (China)

Zaitseva, Elena (Slovakia)

Zalesky, Boris (Belarus)

Zalewski, Janusz (USA)

Zhao, Qiangfu (Japan)

CHAIRMEN’S MESSAGE

It is our pleasure to present you the Proceedings of the XV International Scientific and Technical Conference on Pattern Recognition and Information Processing (PRIP’2021), which was held in September 2021 at the United Institute of Informatics Problems of the National Academy of Sciences of Belarus.

The PRIP'2021 Conference has a long history. It began in 1991 as the First All-Union USSR Conference on Pattern Recognition and Image Analysis in Minsk. PRIP was organized alternately by IEIP, Belarusian State University, and Belarusian State University of Informatics and Radioelectronics, and to date, 15 conferences were held.

2020 and 2021 have changed the international landscape affected by the Covid-19 pandemic. We are experiencing difficulties in meeting with our foreign colleagues, so we decided to hold PRIP'2021 purely online.

This year, for the first time, the Conference had a motto: "Artificial Intelligence: Facing the Challenges". It is difficult to find an area of Computer Science, where machine learning methods are not applied now. But the main challenge was the need to preserve the traditions of a scientific event while changing both the format and approach to the organization. We had to take into account the fact that the world scientific and educational community has already developed certain standards for large-scale online events.

This year, the Conference was held on a single track. 90 applications were submitted, but only 53 reports were accepted after peer-reviewing for presentation at the Conference. 75 speakers (including keynotes and invited speakers) from 18 countries took the floor.

The topic of the Conference usually includes theoretical and applied aspects of computer vision, recognition of signals and images, the use of distributed resources, and high-performance systems. This year it has been significantly expanded with the issues of Artificial Intelligence. Belarusian scientists have their achievements both nationally and globally within the framework of large projects in this area. Of course, a feature of this conference was the speeches of famous foreign scientists, including Professor Ruslan Salakhutdinov from Carnegie Mellon University (USA), Professor Jos Roerdink from the University of Groningen (Netherlands), Professor Frederick Leymarie from the University of London (UK), Professor Henning Müller from the University of Geneva (Switzerland), Professor Axil Mosig from the Ruhr-University of Bochum (Germany), Dr. Andrei Gabrielian from the National Institute of Allergy and Infectious Diseases of NIH (USA).

This year, the Conference partner was the GÉANT Association, which provides global connectivity for the needs of science and education. Within the framework of a GÉANT Session, participants were able to communicate with invited speakers from the Netherlands, Great Britain, Armenia, Italy, Greece, Poland, and Croatia. A highlight was the joint performance by Pierre-Philippe Mathieu, Director of the Φ-Lab Explore Office at the European Space Agency, and his colleagues.

The concept of Artificial Intelligence is actively entering our daily life, but there are still discussions about the effectiveness of using these methods, as well as about possible abuses of AI technologies. It formed the basis for a virtual panel organized on the last day of the Conference by participants from Belarus, Armenia, Great Britain, Germany, and Croatia.

We sincerely thank everyone who took part and helped us hold this large-scale online event in the scientific community of Belarus with international participation in the field of information processing and pattern recognition! We also would like to thank the organizing committee for the excellent work due to which this event took place!

Prof. Alexander Tuzikov

Prof. Sergey Ablameyko

Minsk, October 8, 2021

CONTENTS

Parallelizing by Smart Tiling

Sergii Sushko
Dept. of Modelling and Econometrics
Pukhov Institute for Modeling in Energy Engineering of NASU
Kyiv, Ukraine
sergii.sushko@gmail.com

Alexander Chemeris
Dept. of Modelling and Econometrics
Pukhov Institute for Modeling in Energy Engineering of NASU
Kyiv, Ukraine
a.a.chemeris@gmail.com

Svetlana Reznikova
Dept. of Modelling and Econometrics
Pukhov Institute for Modeling in Energy Engineering of NASU
Kyiv, Ukraine
svetlana.reznikova@gmail.com

Abstract. **The paper is devoted to the methods of automatic parallelization and software optimization. The authors focus on parallelizing of computational loops. The problem of quickly choosing a partitioning method and determining its parameters is an urgent task. Its solution provides a reduction in software's execution time for computing systems with multiprocessor architecture. To build an automated system for parallelizing programs, the authors propose to use Discrete Particle Swarm Optimization Method as an optimization method which allows to find a local or global minimum of program execution time regarding complicated relationship between tile sizes and execution time. The paper proposes an approach to optimizing the process of partitioning the iterative space of loop operators using the methods of Swarm Intellect. It uses partitioning by rectangular parts but has no fundamental restrictions on its use for other types of partitioning (triangles, parallelograms, rhombuses, etc.). Proposed method is called Smart Tiling Method.**

Keywords: **parallel programs, program parallelization, loop parallelization, particle swarm optimization, tiling, smart tiling**

I. INTRODUCTION

Hardware and software are constantly being improved during entire period of development of computer technologies. At the same time with an extensive type of development also an intensive type of development is used. Considerable attention is paid to more efficient usage of available resources. For example, more efficient usage of resources of data center can have a significant impact on energy consumption. Thereby, an amount of electricity which is required for air condition and as a result an amount of financial costs will reduce. Effective usage of mobile devices allows to work longer without recharging and to use a hardware with better computing capabilities. Effective usage of embedded systems allows to expand of functionality and service capabilities of devices.

Computational efficiency is a multifactorial characteristic. First, it is a measure of ability of hardware and compiler to implement high-level source code [1]. Second, computational efficiency can be defined as a fraction of actual processing performance relatively to peak processing performance available for given hardware. Third, the computational efficiency is based on efficiency of algorithms used by software. These factors can be applied to all computing systems – PCs, servers [2, 3], mobile devices [4], FPGAs [5], controllers, embedded systems [6] and so on.

Computational efficiency is also considered in context of integration of computational hardware. For example, estimates of effectiveness of various options for creating of complex of computing facilities as system with predefined composition of equipment are given in [7]. The author considers such indicator as a coefficient of decrease of real performance which characterizes cost of computing performance for organizing of common work of computers or processors.

$$K_k = \frac{P}{\sum_{i=1}^{N} V_i}, \qquad (1)$$

where V_i is effective performance of i-th computing device for individual functioning and specific class of problems; P — same parameter for whole system; N – number of computers or processors.

Efficiency of implementation of a certain algorithm or class of algorithms of computational system is almost completely determined by relation of structure of chosen algorithm and by structure of a system. For practical usage of computing systems, a flexibility of logical structure of algorithm and its ability to transform become especially important.

II. THE PROBLEM OF LOOP PARALLELIZATION

In operator of computational loop which consists of n nested loops, a set of distance vectors is approximated by an integer number l from the interval $[1, n] \cup \{\infty\}$, which is defined as the largest integer such that the first l–1 components of distance vectors are zeros. Dependence at level $l \leq n$ means that dependence is found on level l of nested loops, that is, on given

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

iteration of l–1 outer loops. In this case, dependence between iteration dependence and such dependences are called loop-carried on l level. If $l = \infty$, then dependence occurs inside of body of loop between two different operators. Such dependencies are called loop-independent. Value l is called level of dependence. Analysis of computational loops, each of which is defined by its own index variable, yields a set of index variables which creates an index vector of nested loops $I = (i_1, i_2 \ldots., i_n)$.

For any loop in which index of loop I changes from L to U with a step S, an iteration number i is equal to the value $(I - L + S)/S$, where I is index value for this iteration [7].

For n nested loops, iteration vector I for innermost loop is a vector containing an integer number of iterations for each loop in order of nested loop. In other words, iteration number of a multidimensional nested loop is determined in accordance with the form $I = (i1, i2 \ldots., in)$, where i_k, $1 \leq k \leq n$, is a loop iteration number for nested level k.

Definition. Iteration space is a set of all integer vectors $I = (i_1, i_2 \ldots., i_n)$ which satisfies the inequality:

$$L_i \leq x_i \leq U_i, i = 1..n., \tag{2}$$

Inequality (2) defines boundaries of loop which limit iteration space by a convex polytope.

Thus, a model for representing of iteration space is defined, which consists of constraints that define boundaries of space, a set of nodes that corresponds to loop iteration, and a set of dependencies between iterations. Such model of iteration space which represents it as a convex polyhedron, is called a polyhedral model [8].

Task of parallelization is to divide iteration space into separate blocks, under which following conditions should be met, if possible:

1) compliance with structure of computing system, for example, take into account architecture and number of processors;

2) ensure of same loading on system processors;

3) minimize connections between blocks of space, ensuring their parallel execution.

Considering variety of approaches for modification of computational loops, tiling method and its modifications should be noted. This method introduces additional loops which break entire iteration space into small blocks so called tiles. Calculations are performed first for each tile. Tile size is chosen in each case individually. This approach improves data locality which leads to more efficient use of cache and internal registers of processor and this can reduce loading on memory bus, faster computations and lower power consumption.

In general, implementation of parallelization and optimization process is performed by converters, algorithms, methods that analyze a program code and convert it to semantically equivalent version which is more efficient by some set of optimization goals. It was shown in [1] some code optimization problems are NP-complete or even such which cannot be solved. In practice, many of them are solved by heuristic methods and give a result in a satisfactory processing time for program code.

Conversion of software to minimize or maximize some goal function, which in general form is represented by expression (3):

$$\underset{x}{argmin}\, f(x) \in \{x \mid \forall y : f(y) \leq f(x)\}. \tag{3}$$

Then, considering parallelization and software optimization as a process of directed application or selection of a finite number of methods with a finite set of parameters, it can be described by the expression (4):

$$F_i = argmin(f(\overrightarrow{\boldsymbol{M}_l}, \overrightarrow{\boldsymbol{P}_k})), \tag{4}$$

where F is a target function, which can be one or several parameters requiring improvement – reduction of execution time, power consumption, program memory size, data, etc.; $\boldsymbol{M}$ is a set of possible methods for parallelization and optimization; $\boldsymbol{P}$ is a set of parameter values for these methods.

Expression (4) means that for tiling of iteration space, it is necessary to use several appropriate methods with their own optimization parameters. The choice of methods can be preselected (defined explicitly or set by default) or determined by evaluating a code by using various metrics.

Thus, process of dividing of iteration space into separate tiles is complicated by fact that a choice of a set of specific methods and their specific parameters is not predetermined and requires additional experiments to test effectiveness and their combinations. Through architectural features and limitations, selection of methods and their parameters is unique for each practical case. The authors propose Smart Tiling Method which is based on an optimization method by using swarm intelligence.

III. SMART TILING METHOD FOR LOOP PARALLELIZATION

The development is based on a typical process that is used in Pluto software package [8]. Its main stages are shown in Fig. 1. This package includes several software modules. These modules are used one by one. First one creates a polyhedral model based on source code of C/C++ languages, then Pluto package itself modifies this polyhedral model according to the specified methods of tiling and parallelization, then other software modules simplify resulting model without changing of lexicographic sequence, and at the last stage program modules create a source code of C or C++ programming languages according to obtained model.

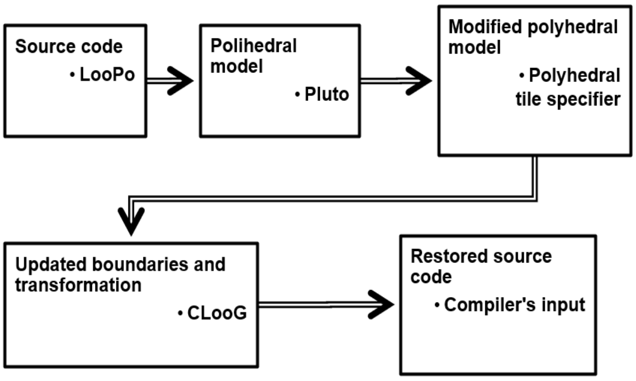


Fig. 1. Parallelization and optimization in Pluto

To efficiently divide of iteration space of loop operators, it is necessary to choose appropriate method that will give the maximum computational performance and find corresponding parameters of this method. For the approach which is shown in Fig. 1 it depends on efficiency of partitioning methods implemented and on experience of software developer. Nevertheless, automation of this process brings its additional effects on inefficiency of transformations. Therefore, it’s necessary to investigate and develop some approaches to implementation of parallelization and program optimization.

Process of choosing of parameters for tiling of iterative space is performed on the stage of modifying of polyhedral model. The experiments presented in [9] show complex dependence of computational efficiency on tiling method and size of tiles into which space is partitioned. After determination of minimum of this function a method of tiling and tiles’ sizes can be obtained.

The authors, in order to find the minimum of the goal function, proposed to use optimization method of swarm intelligence, namely Particle Swarm Optimization Method [10]. By using of discrete version of the method, the authors propose an intelligent method for partitioning of iterative space of loop operators in software programs. This will reduce searching time for optimal solution when parallelization algorithms for multiprocessor computing systems. The authors performed the experiments for programs written in C/C++, but there are no significant restrictions to use for other languages.

Smart Tiling Method, as shown in Fig. 2, uses Discrete Particle Swarm Optimization Method to find the best tiling block sizes by iterative size choice. Next, effectiveness of chosen tile sizes is evaluated by compiling and measuring of execution time of computer program. According to result of evaluation, parameters of particle swarm are modified. This algorithm is quite flexible and allows to use different versions of tiling methods. There are possible two approaches for tiling – standalone and with parallelization together.

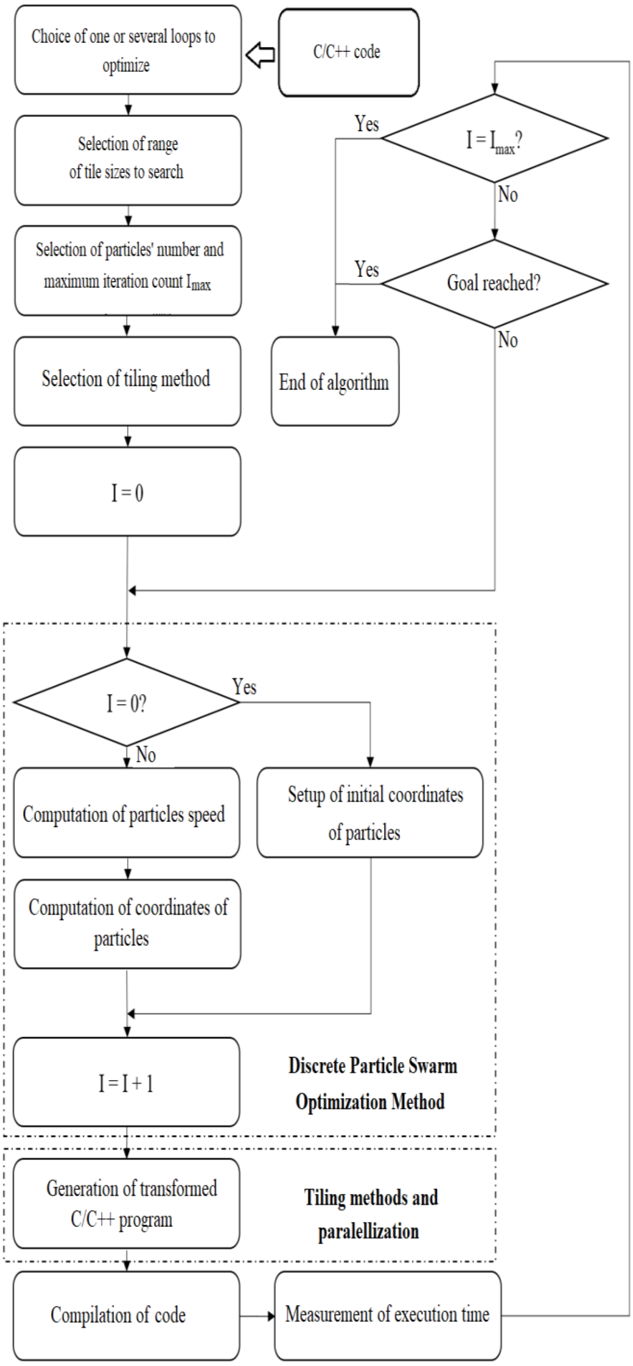


Fig. 2. Smart Tiling Method based on Discrete Particle Swarm Optimization Method

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

IV. THE EXPERIMENTS

For the evaluation of impact of various methods of tiling of iteration space on execution time, a lot of experiments were performed on various hardware platforms. The first experiments were performed on a single board computer Raspberry Pi 3 (quad-core ARM Cortex-A53 CPU, 1.2 GHz, 32KB L1 cache, L2 512KB, 1GB LPDDR2 900 MHz memory). Another series of experiments was verified on the desktop PC with Intel processor (Intel® Core ™ i5-4670K quad-core CPU, 3.4 GHz, 4x64KB L1 cache, L2 4x25.26KB, L3 6MB, 16GB DDR3 1333 MHz memory). The package of functions Polybench 4.1 [11] was chosen as set of the test programs. This package contains about 30 small test programs written in C, in which various classes of algorithms are implemented – operations with vectors, matrices, linear algebra, signal processing. Some experimental results are given in [9].

A series of experiments was performed to test an efficiency of optimizing on execution time of test programs by using of Smart Tiling Method. Smart Tiling Method was used for several tiling methods of Pluto package, namely “tile”, “innerpar”, “parallel”. The obtained data were normalized taking into account the initial execution time of the test programs and are shown in Fig. 3 and 4 as diagrams.

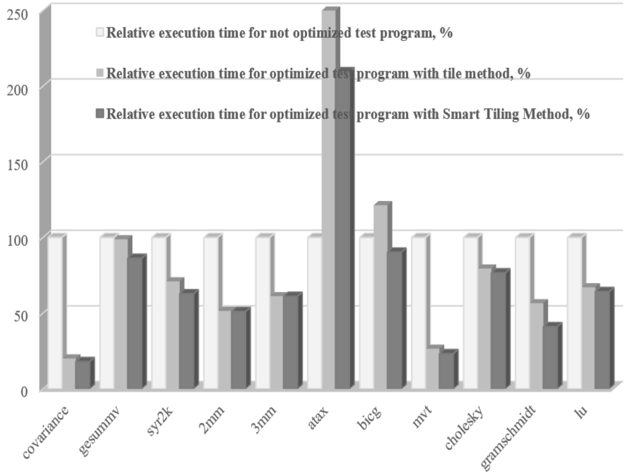


Fig. 3. Relative execution time with “tile” method in comparison with Smart Tiling Method

To accelerate preparation of software for computing systems with multiprocessor architecture, usage of automated parallelization and program optimization systems is an urgent and promising task. Potential effect on microprocessor control systems, IoT systems, embedded systems, mobile devices, and so on should be emphasized.

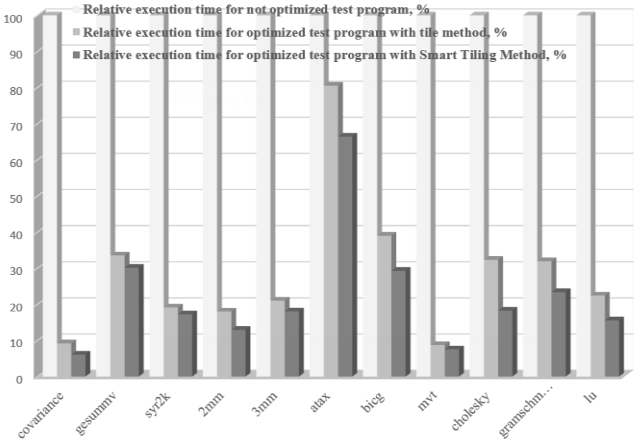


Fig. 4. Relative execution time with “tile”, “innerpar” and “parallel” methods in comparison with Smart Tiling Method

Polyhedral model for representing of iteration space of loop operators is formed by a system of constraints on loop variables and is N-dimensional convex polyhedron. This is of many approaches to analysis and parallelization of program loop operators. It is promising for building an effective system for parallelizing of loop parts of algorithms.

V. CONCLUSIONS

Analysis of methods for dividing of iteration space into parts or tiling method shows a complex dependence of program execution time and, thus, computational efficiency on parameters of tiling. Usage of optimization methods to find minimum of goal function makes it possible to speed up a process of program parallelization. For this Smart Tiling Method based on Discrete Particle Swarm Optimization Method is proposed. The method allows to speed up searching a solution by choosing the parameters of tiling method. This allows to improve performance of programs due to better choosing of parameters and thus faster program execution. Experiments show the effectiveness of the method for 2-dimensional case. For certain classes of problems, it’s possible to get up to 15 times improvement in performance.

The developed method for optimizing of partitioning of iteration space of program loop operators, which the authors investigated for the 2-dimensional case of tiling of iteration space into rectangular parts, has no fundamental restrictions on its to use for other types of partition (triangles, parallelograms, rhombuses, etc.).

REFERENCES

[1] V. V. Voevodin, V. V. Voevodin, Parallel computing, St. Petersburg, BHV-Petersburg, 2002, 608 p. (in rus.)

[2] O. O. Druzhinina, R. N. Kvetny, Increasing the efficiency of web servers with the use of time series forecasting technology based on neural networks, Information technology and computer engineering, 2013, № 1, pp. 15-21. (in ukr.)

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

[3] G. M. Lutskiy, I. S. Raizin, Increasing the efficiency of clusters based on Infiniband, Bulletin of the University "Ukraine", Informatics, computer science and cybernetics, 2011, № 8, pp. 133. (in rus.)

[4] P. Havinga, G. Smit, Low power systems design techniques for mobile computers. Centre for Telematics and Information Technology University of Twente, Enschede, 1997.

[5] I. A. Klimenko, Methods and means to increase the efficiency of information processing in reconfigured computer systems based on FPGA, Dr. Tech. Science thesis: 05.13.05. Kyiv, 2017, 377 p. (in ukr.)

[6] N. V. Borisova, L. V. Shabanova-Kushnarenko, Effective resource management of embedded systems for real-time computing, Information processing systems, 2018, № 1(152), pp. 87-93. (in ukr)

[7] V. K. Dushin, Theoretical foundations of information processes and systems, Moscow, Dashkov and Co., 2003, pp. 348 (in rus).

[8] Uday Bondhugula, Albert Hartono, J. Ramanujam, Ponnuswamy Sadayappan A practical automatic polyhedral parallelizer and locality optimizer, In Conference: PLDI '08: Proceedings of the 2008 ACM SIGPLAN conference on Programming language design and implementation, May 2008, ACM SIGPLAN Notices 43(6), DOI: 10.1145/ 1375581.1375595

[9] S. Sushko, A. Chemeris, Dependency between Tiles' Sizes and Program Execution Time, Proceedings: Reconfigurable Ubiquitous Computing (RUC-2018), 12 October 2018, Dzivnuv, Poland.

[10] A. Chemeris, S. Sushko, Usage of Discrete Particle Swarm Optimization Method for the Searching of Optimal Tile Size, 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T-2019), October 8-11, 2019, Kyiv, P. 202-206.

[11] L. N. Pouchet, PolyBench/C the Polyhedral Benchmark suite http://web.cse.ohio-state.edu/~pouchet/software/polybench/ #description.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Graph-Based Recognition of High-Level Structures in Transistor Circuits

Liudmila Cheremisinova
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
cld@newman.bas-net.by

Dmitry Cheremisinov
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
cher@newman.bas-net.by

Abstract. **The problem of converting a flat transistor circuit into a hierarchical circuit of logical gatesis considered. The problem arises in layout versus schematic verification and reverse engineering of integrated circuits. The offered subcircuit recognition algorithm collects transistors into gates without using any predefined cell library. Graph-based methods are proposed for solving some key problems of subcircuit (CMOS gates) recognizing and logical network extraction. The presented graph methods have been implemented in C++ as a part of a decompilation program, which was tested using practical transistor-level circuits.**

Keywords: **VLSI layout verification, reverse engineering, subcircuit extraction, graph matching**

I. INTRODUCTION

Modern digital circuits contain up to a billion primitive elements at the transistor level, and the circuit complexity is rapidly increases while time-to-market is imposed to decrease. Typically, digital system designers move from a gate-level netlist to a physical layout and mask and rarely proceed in the opposite direction. However, in recent years the study of reverse engineering of digital circuits has become increasingly important.

The step toward raising the level of circuit description is performed by decompiling transistor circuit to replace its representation at a low (transistor) level with a higher-level representation (logic gate level). Tools for solving the task can be used for supporting many tasks of designing integrated circuits, such as functional verification [1], fault simulation and automatic test pattern generation [2], hardware Trojan detection [3], circuit reengineering [4], static timing analysis, etc. At first, the main application of the means of transistor circuit decompilation was verification of the software implementations and finding logical bugs. In recent years, the validation of the integrity of untrusted design becomes a pressing issue. It is recognized, reverse engineering techniques can help detect hardware Trojans and malicious design changes [3].

In the paper we consider the problem of extraction of logical networks from transistor-level circuit netlists in SPICE. In graph interpretation, the problem is formulated as recognition of subgraphs corresponding to logical gates and other subgraphs that are often encountered in a given graph. The problem complexity was thought to be tremendous, but VLSI transistor netlists tend to be sparse enough and have the specific structure, so runtimes did not grow unreasonably, because a sensible data structures and data processing methods were adopted.

There were many attempts to solve the problem of extracting the hierarchy of large-scale subcircuits from a transistor circuits for various VLSI technologies, restrictions, solution methods. An overview of known approaches can be found in [5, 6]. Some methods of logical network extraction are based on structural recognition and use rule-based methods in which CMOS gate structures are recognized as channel connected sequences of MOS transistors [5, 7]. The other approaches [8] are based on mapping transistorlevel circuit into a graph and treating subcircuit pattern matching problem as subgraph isomorphism one. Some methods suppose that subgraphs to be found are known and the problem is reducible to pattern recognition [8].

The proposed paper presents methods and decompilation program for the most general case with no predefined cell library. Moreover, the methods make it possible to recognize subcircuits that implement the same logical functions but are not topologically isomorphic. The method is based on solving well-known graph problems, which are modified to process large transistor-level descriptions in a short time. The presented graph methods have been implemented in C++ as a part of a decompilation program, which was tested using practical transistor-level circuits.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

II. TRANSISTOR CIRCUIT REPRESENTATION

The source and resulting circuit netlists are presented in SPICE (Simulation Program with Integrated Circuit Emphasis) format [2]. The main part of the circuit netlist in the format is the list of transistors, in which each transistor terminal (drain, gate, source and substrate) is indicated by the label of the net connecting it with the rest of the circuit: M <name><nd><ng><ns><nb><model-name>,where M and «model-name» are the title and the type of the transistor; nd, ng, ns and nb are the labels of nets connected with its drain, gate, source and substrate terminals. For example, the transistor instance description «mp 2 1 3 3 P» is an abbreviated notation for the pairs (mp.d, 2), (mp.g, 1), (mp.s, 3), (mp.b, 3), in which the name «mp» of p-MOS transistor is taken out and the names of its terminals are omitted.

This netlist format defines an electrical circuit as consisting of elements connected to each other by nets. The convenient natural way to represent such circuits is to use an undirected bipartite graph $G = (V_1, V_2, E)$, $V_1 \cap V_2 = \varnothing$, were vertices are divided into classes V_1 and V_2. The vertices fromV_1 correspond to transistor terminals and circuit ports (primary inputs and outputs), and the vertices fromV_2 correspond to nets, i.e. connections between the terminals. Each edge $e \in E$ has one end in V_1 and the other in V_2.

III. DEFINITIONS AND NOTATION

As stated above, transistor circuits are modeled as bipartite graph $G = (V_1, V_2, E)$, $V_1 \cap V_2 = \varnothing$. Throughout this paper we assume that the graph is undirected and vertex-colored, i.e., each vertex has a color associated with it, that is drawn from a predefined set of vertex colors $L(V)$. Transistor-level circuits made by CMOS technology have several types of their nodes: terminals (drain, gate, source and substrate) of n-MOS and p-MOS transistors, input/output ports (external nets), power supply terminals (Vdd and Gnd) and internal nets. So each graph vertex corresponding to n-MOS terminal is assigned by one of the first four colors, p-MOS terminal is assigned by one of next four colors. Then input/output ports, Vdd and Gnd nets, internal nets have unique colors.

The graph corresponding to a MOS circuit is connected (there is a path between any pair of vertices in the graph) and sparse. Two bipartite colored graphs, $G^1 = (V_1^1, V_2^1, E^1)$ and $G^2 = (V_1^2, V_2^2, E^2)$, are isomorphicif there is a one-to-one mapping f: $V_1^1 \leftrightarrow V_1^2$ and $V_2^1 \leftrightarrow V_2^2$ between vertices of graphs such, that for each $v \in V_1^1 \cup V_2^1 L(v) = L(f(v))$ and each edge in E^1 is mapped into a single edge in E^2 and vice versa, i.e. $(v, u) \in E^1$iff$(f(v),f(u)) \in E^2$.

Given graph $G_s = (V_s, E_s)$ is a subgraph of $G = (V, E)$if $V_s \subseteq V$ and $E_s \subseteq E$. Two subgraphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ of a graph G are called edge-disjoint if they do not share edges, i.e., they use different sets of edges from E: $E_1 \cap E_2 = \varnothing$.

IV. GRAPH-BASED FORMULATION OF SUBCIRCUIT EXTRACTION PROBLEM

The proposed subcircuit extraction application begins with the construction of a graph model from the SPICE description and hierarchical hash tables for storing the syntax elements of the analyzed circuit [9]. After that, a preprocessing of the circuit is performed, during which some standard fragments are searched for. For example, identification of groups of identical MOS transistors (with the same signals supplying their terminals), connected in series or in parallel, pass gates is fulfilled.

The goal of the transistor circuit decompilation is to recognize subcircuits, which implement logic gates, or, if we cannot, to split the circuit into sufficiently large subcircuits that look like as logic gates In graph interpretation the problem is solved by partitioning a graph into sufficiently large edge-disjoint subgraphs in such a way, that they can be partitioned into the minimum number of classes of isomorphic graphs.

The algorithm realizes two-step process. First, it uses rule-based structural approach in which CMOS gate structures are recognized as channel connected sequences of transistors. Then frequent subcircuit pattern recognition is done to gather the rest transistors into restricted number of identical functional blocks. Finally, the set of all subcircuits, both implementing and not implementing CMOS gates, is partitioned into classes of topologically identical. Subcircuits of the same class represent the same functional block in resulting hierarchical description of two-level decompiled circuit. In graph interpretation the task is to classify subgraphs into classes of isomorphic.

As result of the mentioned steps performing, a hierarchical mixed gate-block-transistor netlist is generated. In the next step, the extraction of logic network from the hierarchical transistor-level circuit is done. That makes it possible to recognize more complex elements than gates. In graph interpretation the task is to extract (out of undirected graph) connected subgraphs only with those vertices that correspond to CMOS gates, and to convert the resulting undirected subgraphs into oriented ones.

V. PARTITIONING A GRAPH INTO CONNECTED SUBGRAPHS

In MOS transistor circuit, correct subcircuits are among channel connected sequences of transistors. So,

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

first, the proposed method of the subcircuit recognition algorithm uses the structural approach to divide the transistor-level circuit into subcircuits, that are channel connected sequences of transistors, as well as in [5, 7]. A group of channel connected transistors is a cluster with three types of external connections: the cluster inputs are fed only to the transistor gates; the outputs are supplied only outside the cluster; there are connections to Vdd and Gnd terminals. Fig. 1 shows an example of grouping transistors into two channel connected components.

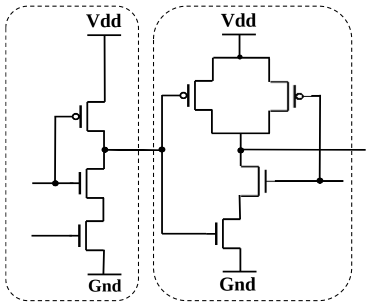


Fig. 1. Channel connected components of MOS transistor circuit

The task of recognition of the clusters is solved on a graph *H*, obtained from the introduced graph $G = (V_1, V_2, E)$ by removing gate terminals of transistors; by introducing local Vdd and Gndnets (with unit vertex degrees) instead of global ones and by shorting the drain and source terminals for each transistor.

In graph interpretation, channel connected groups of MOS transistors correspond to edge-disjoint connectivity component of the graph *H*. The search for graph *H* connectivity components is done by using the well-known depth-first search (DFS) algorithm that starts at some arbitrary unconsidered vertex and explores paths from it as far as possible along each branch before backtracking. Reaching a backtracking results in a new connectivity component. When implementing the DFS algorithm, the initial graph $G = (V_1, V_2, E)$ is not transformed explicitly into the graph *H*. Instead, the DFS algorithm was tuned to the modification of data structure for storing the graph *G*.

VI. LOGIC GATE STRUCTURAL RECOGNITION

The CMOS gate consists of two blocks separated by a connection net (output net) (Fig. 2). The first block is formed by n-MOS transistors (pull-down network), which are connected in series by their source/drain terminals. The second block is formed by p-MOS transistors (pull-up network), which are connected in parallel. The pull-down network is placed between the connection node and Gnd, and the pull-up network between Vdd and the connection node. The conductivities of the blocks are complementary, no matter what the input signals (on transistors gates) are, there is a valid path to output node either from Gnd or from Vdd.

A CMOS gate is a group of channel connected transistors; the opposite is not always true. The necessary conditions for the group to belong to the class of CMOS gates are the following ones: the only chain connecting the pull-down and pull-up groups is the output (connection) node; all paths from the connection node go to Gnd or Vdd; pull-down and pull-up networks have the same number of transistors and implement mutually inverse functions. For instance, the right group of channel connected transistors in Fig. 1 is a NAND gate, but the left one is not.

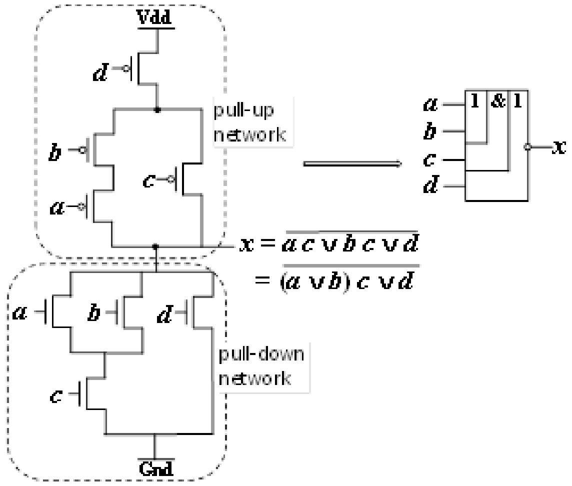


Fig. 2. CMOS gate: its transistor structure and implemented function

Thus, among channel connected components, there are those that implement standard CMOS gates. Their pull-down and pull-up networks have the same number of transistors and implement mutually inverse functions. The task is to find out such subcircuits and their functionality.

In graph interpretation it consists in tracing all paths between vertices corresponding to connection node and Gnd (or Vdd). Each path gives a conjunction of the conductivity variables fed to gate terminals of the transistors from the path. The OR of all such conjunctions yields disjunctive normal form (DNF) for the expression. If the conductivity functions f_n and f_p of pull-down or pull-up networks are complementary ($f_n = \overline{f_p}$) then the analyzed channel connected group is a CMOS gate.

To classify CMOS gates extracted from the transistor circuit, it is convenient to represent the recognized functions as parenthesized algebraic expressions. Such a form is constructed by the algebraic factoring of the found DNF [9]. For the CMOS gate in Fig. 2 we have

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

$f_n = ac \vee bc \vee d = (a \vee b)c \vee d$, $f_p = \overline{a}\overline{b}\overline{d} \vee \overline{c}\overline{d} = (\overline{a}\overline{b} \vee c)\overline{d}$

and $f_n = \overline{f_p}$. Thus, it is standard NOAO2 CMOS gate.

Transistor subcircuits of CMOS gates are divided into classes of functionally identical according to the formulae of implemented logic functions.

Often it is important to take into account not only functional aspect but the topological one too. The aspect requires dividing a class of functionally equivalent CMOS gates into subclasses of topologically equivalent ones. Some features of the topological implementation of circuits at the transistor level, which must be taken into account are given in [9]. For example, we should take into account the interchangeability of drain and source terminals of MOS transistors which results in existence of topologically different subcircuits that implement the same logic function. For instance, there are four subcircuit variants for a CMOS inverter. If, in a decompiled circuit, all variants of a logic gate subcircuit are represented by the same subcircuit, then the decompiled and original circuits will be topologically not isomorphic. Topological equivalence of CMOS gate implementations can be established by checking whether the corresponding graphs are isomorphic or not.

VII. GRAPH ISOMORPHISM AND CANONICAL LABELING

One of the key operations required to partition the set of subgraphs into classes of isomorphic ones consists in checking whether two subgraphs are identical or not. One way of performing this check is to perform a graph isomorphism operation. But in our case, when many such checks are required for the same set of subgraphs, a better way of performing the task is vertex canonical labeling. It assigns to each graph a unique code (a sequence of bits) that is invariant on the ordering graph vertices. Comparing canonical labels allows to partition the set of graphs into classes of pairwise isomorphic graphs.

In general, calculating canonical labels is computationally hard, but in our case, the complexity of the task is reduced due to taking into account special properties of subgraphs under classification: they are vertex-colored, sparse and small enough. Canonical labels of graphs, makes it is possible to sort them in a unique and deterministic way.

Canonical labeling is done in an iterative manner in the process of building a sequence of vertex partitions that defines an ordering of the graph vertices. Assume we have an ordered collection of subsets of the vertices (V_1, V_2,…, V_k), whose union is V. They say that all vertices from the same subset V_i have the same label i. The set of these subsets represents the partition on the set of graph vertices, constructed from the initial partition that is specified by colors and degrees of vertices.

At first, the number and sizes of these subsets V_i must be the same for both compared graphs, i.e. the graphs have identical partitions of the set V. Then we repeatedly apply a relabeling step, which assigns to each vertex v a classifier: $C(v) = (n_1, n_2,…, n_k)$, where n_i is the number of vertices in subset V_i that are adjacent to v. Using these classifiers, each subset V_i can be partitioned into subsets, where each subset should include all vertices with the same classifier. These subsets are lexicographically ordered according to their classifiers. If a division has been fulfilled, then all classifiers are recalculated (and vertices are relabeled). No division will be done if all vertices in each subset V_i have identical classifiers.

It is clear from the description that the essential idea is to relabel vertices so that each new classifier reflects information about a gradually increasing region around the vertex. In an ideal situation, after exhaustive applying the relabeling process, all subsets in partition (V_1, V_2,…,V_k) will become singletons (containing exactly one member), such a graph canonical labeling is called discrete. If two compared graphs have the same canonical labeling then they are isomorphic with each other.

VIII. GRAPH-BASED SUBCIRCUIT RECOGNITION METHOD

After structural recognition of logic gates and pass gates there are two main unrecognized groups of transistors. They are separate ungrouped transistors and found channel connected components of MOS transistors, which have been not recognized as CMOS gates, so they are assigned to be pseudo gates. In our case when there is no cell library, all we can do is to classify remaining pseudo gates into classes of pairwise identical subcircuits.

In graph interpretation the task consists in testing isomorphism between graphs by means of comparing their canonical labelings. To simplify the canonization problem, the subcircuit graphs are complemented with edges connecting all four terminals for each transistor. As the prototype of the program for computing canonical isomorphs, the program «bliss» [10] has been modified that provides fast handling of large and sparse graphs. The experiments with the program of graph canonical labeling have shown that the canonization of pseudo gate graphs results in becoming discrete canonical labelings.

The graphs of pseudo gates with the same initial partitions on the set of its vertices are considered one

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

by one. For each of them, a canonical labeling is generated and a hash of the canonized graph is computed. Graphs with equal hashes are isomorphic and they are changed in a hierarchical SPICE description with their canonical isomorph.

IX. LOGIC NETWORK CONSTRUCTION

The next step is to extract from the mixed circuit a logical network which consists only of CMOS and pass gates. In graph interpretation, logical network is directed connected graph $H = (W, A)$. The set of vertices W is partitioned into three subsets: network inputs and outputs, and internal vertices. Each vertex is labeled with input or output variable, or, if it is internal vertex, with the function realized by the corresponding gate. A directed edge $a = (u, v) \in A$ goes from the source vertex u to the target vertex v $(u, v \in W)$. Further we consider that graph $H = (W, A)$ is specified by the adjacency list, i.e. an array D of the length $|W|$ where each entry $D[i]$ is a pointer to a linked list of all the out-neighbors of vertex $w_i \in W$.

The connected graph $H = (W, A)$ is extracted from undirected bipartite graph $G = (V_1, V_2, E)$ corresponding to the object mixed circuit. Graph H is contained in G as the connected component C that includes only vertices corresponding to CMOS or pass gates. There can be more than one such a component in a graph G. Each undirected connected subgraph corresponding to a connected component in an undirected graph G is transformed into a directed connected graph $H_i = (W_i, A_i)$ of some logical network. The transformation is carried out in the process of traversing the subgraph along the paths in-going or out-going from the vertices labeled as gates.

The search of the next connected component C begins with any unconsidered vertex labeled as gate and is done by the breadth-first search (BFS) method considering only vertices labeled as gates. BFS allows not only to find out a connected component C, but also to get its topological sorting, which orders the vertices so, that the ordering respects reachability. In other words, if a vertex u is directly reachable from v, then the edge $(u, v) \in E$ generates $(v, u) \in A$, and if the vertex v belongs to the i-th graph rank then the vertex u belongs to the $(i+1)$-thrank. The proposed method provides to extract logic network that is ranked lexicographically. From a lexicographically ordered network of logical gates, it is easy to pass to the formulas of logical equations that specify the output functions of the network.

The next task connected with the logic network extraction is to determine its primary inputs and outputs. It is solved by considering fan-ins and fan-outs for all vertices of the graph $H = (W, A)$. If all vertices from both fan-in and fan-out of some vertex $v \in W$ are labeled as gates then the vertex v is an internal one. Non-internal vertices are referred to primary inputs or primary outputs, depending on which of the fan-in and fan-out sets contains a non-internal vertex.

After the gate-level networks are extracted it is possible to recognize more complex elements, than gates, when a cell library is known.

X. CONCLUSION

In the paper we present graph-based methods for solving the task of extracting gate-level circuits from transistor-level descriptions for the most general case when any predefined cell library of logic gates is unknown. We have used well-known graph methods, modifying them so that they process large data of special types in a short time. The proposed methods were implemented in C++ as a part of a decompilation program. The program was tested using practical and automatically designed transistor-level circuits. The tested circuits had up to 100000 transistors. Some results of experiments can be found in [11].

REFERENCES

[1] M. S. Abadir, J. Ferguson,“An improved layout verification algorithm (LAVA)”, Proc. of European Automation Conf., 1990, pp. 391–395.

[2] R. J. Baker,“CMOS circuit design, layout, and simulation”, 3rd ed., Wiley-IEEE Press, 2010.

[3] R. Torrance and D. James, “The state-of-the-art in IC reverse engineering”, Proc. of the 11th Int’l. Workshop on Cryptographic Hardware and Embedded Systems, CHES ’09, 2009, pp. 363–381.

[4] V. D. Hunt,“Reengineering: Leveraging the Power of Integrated Product Development”, Wiley, 1993.

[5] L. Yang, C-J.R. Shi, “FROSTY: A program for fast extraction of high-level structural representation from circuit description for industrial CMOS circuits”,Integration, the VLSI Journal, vol. 39(4), 2006, pp. 311–339.

[6] N. Zhang, D. C. Wunsch, F. Harary,“The subcircuit extraction problem”, Proc. IEEE Intern. Workshop on Behavioral Modeling and Simulation, 33(3),2003, pp. 23–25.

[7] “Logic Gate Recognition in Guardian LVS – Silvaco”. In https://www.silvaco.com/content/appNotes/iccad/2-003_ LogicGates.pdf (access date: 4.1.2021).

[8] M. Ohlrich, C. Ebeling, E. Ginting, L. Sather, “SubGemini: identifying subcircuits using a fast subgraph isomorphism algorithm”, Proc. IEEE/ACM Design Automation Conf., 1993, pp. 31–37.

[9] D. I. Cheremisinov, L. D. Cheremisinova, “Extracting a logic gate network from a transistor-level CMOS circuit”, Russian Microelectronics, vol. 48(3), 2019, pp. 187–196.

[10] T. Junttila, P. Kaski, “Engineering an efficient canonical labeling tool for large and sparse graphs”, Proc. of the Meeting on Algorithm Engineering & Expermiments, New Orleans, Louisiana, January 6 July 2007, pp. 135–149.

[11] D. Cheremisinov, L. Cheremisinova,”Subcircuit pattern recognition in transistor level circuits”, Pattern Recognition and Image Analysis, vol. 30(2), 2020, pp. 160–169.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Metric Correction of Similarities Based on Orthogonal Decomposition

Sergey Dvoenko
Institute of Applied Math and Computer Sciences
Tula State University
Tula, Russia
dvsrge@gmail.com

Denis Pshenichny
Institute of Applied Math and Computer Sciences
Tula State University
Tula, Russia
denispshenichy@yandex.ru

Abstract. **Raw data in modern machine learning usually appear as similarities or dissimilarities between members of a limited set. A positive definite similarity matrix represents a limited set of elements immersed in some metric space with dimensionality right up to the matrix size with similarities considered as scalar products. In a case of a nonpositive definite similarity matrix, it needs metric correction of similarities to be considered as scalar products. The known discrete Karhunen-Loeve expansion is usually used to reduce the dimensionality of the similarity matrix by removing eigenvectors corresponded to negative eigenvalues. As a result, a new similarity matrix of the reduced size is calculated to immerse members of a limited set in a reduced space of eigenvectors corresponded only to positive eigenvalues with data dispersion reduced. According to an orthogonal decomposition based metric correction here, it is proposed not to remove, but change negative eigenvalues to become positive ones. As a result, such an optimal correction preserves the dimensionality and dispersion of raw data.**

Keywords: **similarity matrix, orthogonal decomposition, eigenvector, eigenvalue, Karhunen-Loeve expansion**

I. INTRODUCTION

Let raw data be presented by a similarity matrix of a set of a limited size. It is known, the positive definite square matrix has the positive determinant and their eigenvalues are positive too [1]. In this case the set members can be represented in some multidimensional coordinate space by vectors with distances and scalar products between them calculated based on the cosine theorem. The end points of normalized vectors appear to be arranged on the hypersphere of the unit radius. For all positive scalar products all corresponding vectors are located in the positive quadrant of the coordinate space.

In the mathematical sense, paired comparisons must be immersed in some metric (Euclidean) space. This is the well-known theoretical problem [2]. Under modern conditions, this problem becomes practical in machine learning, data mining, image processing, etc. [3].

Nevertheless, empirical functions for paired comparisons are usually not correct mathematical functions of distances or similarities. Using of them usually results in so-called metric violations in the set configuration in some space. Hence, it needs to recover metric by correction of paired comparisons. Violations appear in negative eigenvalues of the similarity matrix of the set elements.

We have developed before and today we are improving the novel end-to-end correction technology for optimal recovering of a violated metric. As a result, the positive definiteness of the corrected matrix is achieved [4–8].

The originality of such approach consists in the following. Indeed, each metric violation is connected with some member of the set which is supposed to be responsible for the violation. This approach differs from the well-known multidimensional scaling problem, since it doesn't need to recover explicitly the feature space itself.

In this paper, another approach is proposed based on the known Karhunen-Loeve expansion in terms of a system of orthogonal functions. It is known, we face the problem of the spectral decomposition of a square matrix based on their eigenvectors [1].

II. DECOMPOSITION OF A MATRIX OF SCALAR PRODUCTS BASED ON ORTHOGONAL VECTORS

Let the set of n objects be represented by the normalized matrix $S(n,n)$ of paired comparisons with elements $s_{ii}=1$ for the main diagonal and values $0 \le s_{ij} < 1$ or $-1 < s_{ij} < 1$ for others.

The spectral decomposition of the nondegenerated matrix $S(n,n)$ has the form $S = A\Lambda A^T$, where

---

This work is supported by RFBR Grants 20-07-00055, 19-07-01178.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

$A(n,n)=(\mathbf{a}_1,\dots\mathbf{a}_n)$ is the orthogonal matrix $A^TA=A^{-1}A=E$ of eigenvectors-columns $\mathbf{a}_i=(a_{1i},\dots a_{ni})^T$, $|\mathbf{a}_i|=1$, $E(n,n)$ is the unity matrix, $\Lambda(n,n)$ is the diagonal matrix of eigenvalues sorted in the decreasing order.

It is known, the set represented by the matrix $S(n,n)$ is correctly immersed in some metric (Euclidean) space of the dimensionality not more than *n*, and the matrix itself consists of normalized scalar products of the set elements.

It needs to note immediately, the raw data matrix $X(n,m)$ and the corresponding feature space $\mathbb{R}^m$ with the dimensionality $m\le n$ are not presented here. We suppose they have been lost, otherwise scalar products of objects $S'(n,n)=(1/m)XX^T$ would be properly calculated.

It is easy to see that $\Lambda=A^TSA$ with $tr\,\Lambda=tr\,S=n$ based on the decomposition of the nondegenerated normalized matrix S of scalar products of objects.

III. Correction of Metric Violations

In data analysis problem, the decomposition in terms of orthogonal vectors usually used for a correlation matrix of features $R(m,m)=(1/m)X^TX$ targeted to reduce the dimensionality of the space of eigenvectors of the matrix $R(m,m)$ to the new value $m'<m$ in different tasks.

In particular, the noncorrect matrix $R(m,m)$ has negative eigenvalues. Hence, the projection of the initial data matrix $X(n,m)$ in the new orthogonal subspace $\mathbb{R}^{m'}$ is usually used. The new dimensionality is defined only by positive eigenvalues $\lambda'_1>\dots>\lambda'_{m'}>0$ of the initial decomposition (it is the so-called discrete Karhunen-Loeve expansion [9]). The new correlation matrix $\Lambda'(m',m')$ has the diagonal form, where $tr\,\Lambda'=\sum_{i=1}^{m'}\lambda'=m'<m$.

Note, such the projection requires the data matrix $X(n,m)$ to get the new one $X'(n,m')$. In this case, the dispersion of raw data (the total dispersion of normalized features) is reduced to $m'<m$.

Unlike the classical approach, the spectral decomposition in our investigation is used for the set members represented only by scalar products or non-negative similarities (scalar products in the positive quadrant of a metric space) $S(n,n)$. It doesn't matter, what they are: features or objects themselves.

In a case of metric violations in the set configuration in hypothetical space (we have not it), the spectral decomposition $S=A\Lambda A^T$ of the matrix $S(n,n)$ has negative eigenvalues.

In this paper, it is proposed not to reduce initial data, but replace negative eigenvalues in the decomposition by the appropriate positive ones to get as a result the new matrix $\tilde{\Lambda}(n,n)$ of the same dimensionality. After that, the matrix $\tilde{S}(n,n)$ is recovered in the form of $\tilde{S}=A\tilde{\Lambda}A^T$.

Note, the new matrix $\tilde{S}$ appears to be nonnormalized, since their diagonal elements appear to be more, than 1. Therefore, $tr\,\tilde{\Lambda}=tr\,\tilde{S}>n$. After the transformation $\hat{s}_{ij}=\tilde{s}_{ij}/\sqrt{\tilde{s}_{ii}\tilde{s}_{jj}}$, the decomposition of the corrected matrix $\hat{S}(n,n)$ is specified as $\hat{S}=\hat{A}\hat{\Lambda}\hat{A}^T$, $tr\,\hat{\Lambda}=tr\,\hat{S}=n$, where eigenvalues and eigenvectors take the final form.

In fact, based on this approach any set of eigenvalues can be modified by any other values to eliminate not only negative eigenvalues.

Note, based on such the approach, it is possible to formulate any suitable problems to find the corresponding set of eigenvalues for similarity matrices, for example, to provide the right level of the conditionality for the corrected matrix (we demonstrate it below), etc.

Additionally, the raw similarity matrix can have other type of violations, where non-diagonal elements exceed diagonal ones by the module. In this case, the correction recovers the matrix too. Naturally, the raw matrix must be of the full rank. In other case, the matrix dimensionality needs to be reduced before the correction.

IV. Using the Proposed Approach on Real Data

Let the correlation matrix $S(11,11)$ be given. It represents statistical interconnections between power of biorhythms of 11 frequencies in a brain like alpha-, beta-, and theta-rhythms in electroencephalograms.

Psychologist V.D. Nebylytsin obtained such data [10] during his investigations of light-sound sensations imposed by rhythms

$$\begin{vmatrix}
1 & 0.562 & 0.568 & 0.152 & 0.347 & 0.250 & 0.264 & -0.020 & -0.212 & -0.086 & -0.076 \\
0.562 & 1 & 0.784 & 0.057 & 0.196 & 0.218 & 0.009 & -0.017 & -0.002 & 0.163 & 0.284 \\
0.568 & 0.784 & 1 & 0.288 & 0.475 & 0.264 & 0.066 & 0.144 & 0.114 & 0.228 & 0.151 \\
0.152 & 0.057 & 0.288 & 1 & 0.686 & 0.293 & 0.034 & 0.048 & -0.069 & -0.064 & 0.175 \\
0.347 & 0.196 & 0.475 & 0.686 & 1 & 0.429 & 0.070 & 0.152 & 0.036 & 0.028 & 0.216 \\
0.250 & 0.218 & 0.264 & 0.293 & 0.429 & 1 & 0.788 & 0.197 & 0.154 & 0.109 & 0.035 \\
0.264 & 0.009 & 0.066 & 0.034 & 0.070 & 0.788 & 1 & 0.109 & 0.054 & -0.002 & -0.018 \\
-0.020 & -0.017 & 0.144 & 0.048 & 0.152 & 0.197 & 0.109 & 1 & 0.807 & 0.830 & 0.699 \\
-0.212 & -0.002 & 0.114 & -0.069 & 0.036 & 0.154 & 0.054 & 0.807 & 1 & 0.904 & 0.728 \\
-0.086 & 0.163 & 0.228 & -0.064 & 0.028 & 0.109 & -0.002 & 0.830 & 0.904 & 1 & 0.768 \\
-0.076 & 0.284 & 0.151 & 0.175 & 0.216 & 0.035 & -0.018 & 0.699 & 0.728 & 0.768 & 1
\end{vmatrix}.$$

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

This matrix contains ten positive eigenvalues 3.636340, 2.827085, 1.611613, 1.358204, 0.515165, 0.412792, 0.278171, 0.164165, 0.151054, 0.069977, and the last negative eigenvalue -0.024566. We have already studied this matrix in order to correct it and determine the optimal conditionality based on the end-to-end metric correction technology [6-8].

Replacing a negative eigenvalue with a practically zero value 10^{-5} gives, after normalization, eigenvalues 3.629217, 2.821157, 1.605834, 1.355937, 0.514139, 0.411444, 0.277702, 0.163727, 0.150881, 0.069932, 0.00003, and the corrected matrix

$$\begin{vmatrix}
1 & 0.558 & 0.568 & 0.152 & 0.346 & 0.251 & 0.263 & -0.021 & -0.212 & -0.086 & -0.074 \\
0.558 & 1 & 0.776 & 0.058 & 0.198 & 0.212 & 0.013 & -0.015 & -0.0001 & 0.164 & 0.277 \\
0.568 & 0.776 & 1 & 0.286 & 0.472 & 0.267 & 0.063 & 0.142 & 0.112 & 0.227 & 0.154 \\
0.152 & 0.058 & 0.286 & 1 & 0.686 & 0.291 & 0.035 & 0.048 & -0.069 & -0.064 & 0.173 \\
0.346 & 0.198 & 0.472 & 0.686 & 1 & 0.425 & 0.072 & 0.153 & 0.037 & 0.029 & 0.212 \\
0.251 & 0.212 & 0.267 & 0.291 & 0.425 & 1 & 0.782 & 0.195 & 0.152 & 0.108 & 0.039 \\
0.263 & 0.013 & 0.063 & 0.035 & 0.072 & 0.782 & 1 & 0.110 & 0.055 & -0.001 & -0.021 \\
-0.021 & -0.015 & 0.142 & 0.048 & 0.153 & 0.195 & 0.110 & 1 & 0.807 & 0.830 & 0.696 \\
-0.212 & -0.0001 & 0.112 & -0.069 & 0.037 & 0.152 & 0.055 & 0.807 & 1 & 0.904 & 0.724 \\
-0.086 & 0.164 & 0.227 & -0.064 & 0.029 & 0.108 & -0.001 & 0.830 & 0.904 & 1 & 0.765 \\
-0.074 & 0.277 & 0.154 & 0.173 & 0.212 & 0.039 & -0.021 & 0.696 & 0.724 & 0.765 & 1
\end{vmatrix}$$

with the large conditionality value 122035.7, which is calculated as the ratio of the maximal and minimal eigenvalues under the assumption that they are nonnegative ones.

Earlier in [7, 8], a statistically inspired conditionality value of 59.409 was found, which can be taken as a basis here, although it was found for another correction method.

Replacing a negative eigenvalue with a positive value 0.0612 gives after normalization eigenvalues 3.611811, 2.806524, 1.591663, 1.350369, 0.511535, 0.408003, 0.276734, 0.162535, 0.150530, 0.069870, 0.060426, and the corrected matrix

$$\begin{vmatrix}
1 & 0.550 & 0.569 & 0.150 & 0.343 & 0.253 & 0.259 & -0.022 & -0.214 & -0.087 & -0.070 \\
0.550 & 1 & 0.757 & 0.061 & 0.204 & 0.197 & 0.022 & -0.011 & 0.005 & 0.166 & 0.260 \\
0.569 & 0.757 & 1 & 0.283 & 0.463 & 0.273 & 0.056 & 0.139 & 0.108 & 0.224 & 0.162 \\
0.150 & 0.061 & 0.283 & 1 & 0.686 & 0.287 & 0.037 & 0.049 & -0.067 & -0.063 & 0.169 \\
0.343 & 0.204 & 0.463 & 0.686 & 1 & 0.416 & 0.076 & 0.155 & 0.039 & 0.030 & 0.204 \\
0.253 & 0.197 & 0.273 & 0.287 & 0.416 & 1 & 0.768 & 0.190 & 0.147 & 0.105 & 0.050 \\
0.259 & 0.022 & 0.056 & 0.037 & 0.076 & 0.768 & 1 & 0.113 & 0.058 & 0.001 & -0.030 \\
-0.022 & -0.011 & 0.139 & 0.049 & 0.155 & 0.190 & 0.113 & 1 & 0.807 & 0.830 & 0.687 \\
-0.214 & 0.005 & 0.108 & -0.067 & 0.039 & 0.147 & 0.058 & 0.807 & 1 & 0.904 & 0.715 \\
-0.087 & 0.166 & 0.224 & -0.063 & 0.030 & 0.105 & 0.001 & 0.830 & 0.904 & 1 & 0.757 \\
-0.070 & 0.260 & 0.162 & 0.169 & 0.204 & 0.050 & -0.030 & 0.687 & 0.715 & 0.757 & 1
\end{vmatrix}$$

with the practically optimal conditionality 59.77235484.

It is easy to see, all matrices are practically the same. The optimality of the last correction case is supported by the Karhunen-Loeve expansion properties.

V. CONCLUSION

In this paper, we propose a new approach to the metric correction of matrices of paired comparisons ased on the spectral decomposition of the square matrix of scalar products in terms of its eigenvectors.

Optimality of the correction is supported by the properties of the Karhunen-Loeve expansion and the correct selecting of the new value of the corresponding eigenvalue.

From the other side, here we face the well-known perturbation theory of eigenvalue problems [11, 12]. In the framework of this general theory, the novelty of our approach consists in the attempt to make a perturbation in eigenvalues first to recover then similarity matrices used in machine learning.

In further research, all developed techniques is planned to implement for some actual problems, for example, for using different quality metrics in machine learning, for multimodal analysis of heterogeneous data in formal concept analysis [13], etc.

REFERENCES

[1] J. N. Franklin, Matrix Theory. Mineola, N.Y., Dover Publications, 2000. 292 p.

[2] R. L. Bishop, R. J. Crittenden, Geometry of manifolds. N.Y., Acade-mic Press, 1964. 273 p.

[3] E. Pekalska, R. P. W. Duin, The dissimilarity representation for pattern recognition. Foundations and applications. Singapore: World Scientific, 2005. 607 p.

[4] S. D. Dvoenko, D. O. Pshenichny, “A recovering of violated metric in machine learning”, Proc. of SoICT’16. ACM, N.Y., 2016, pp. 15-21. DOI: 10.1145/3011077.3011084

[5] S. D. Dvoenko, D. O. Pshenichny, “On metric correction and conditi-onality of raw featureless data in machine learning”, Pattern Recognition and Image Analysis, vol. 28(4), 2018, pp. 595–604. DOI: 10.1134/S1054661818040089.

[6] S. D. Dvoenko, D. O. Pshenichny, “Optimal correction of metrical violations in matrices of pairwise comparisons”, Machine Learning and Data Analysis, 1 (7), 2014, pp. 885-890 (in Russian).

[7] S. D. Dvoenko, D. O. Pshenichny, “The conditionality of matrices of pairwise comparisons after metric corrections”, Machine Learning and Data Analysis, 3(1), 2017, pp. 50-60. DOI: 10.21469/ 22233792.3.1.04 (in Russian).

[8] S. D. Dvoenko, D, O. Pshenichny, “A corrrection of metric violations based on statistical hypotheses testing”, Bulletin of the Tula State University. Technical science, 10, 2018, pp. 100-107 (in Russian).

[9] J. T. Tou, R. C. Gonzalez, Pattern Recognition Principles, Addison-Wesley, 2nd ed., 1974. 377 p.

[10] V. D. Nebylitsyn, Selected psychological proceedings. Moscow, Pedagogika, 1990, 408 p. (in Russian).

[11] R. Bellman, Introduction to Matrix Analysis, N.Y., McGrow-Hill, 1970, 440 p.

[12] F. Rellich, Perturbation Theory of Eigenvalue Problems, Routledge, 1969, 138 p.

[13] M. Bogatyrev, D. Orlov, “Application of formal contexts in the analysis of heterogeneous biomedical data”, Proc. of CEUR Workshop, RCAI, 2020, Oct. 10-16, Moscow, vol. 2648, pp. 315-329.

Pattern Recognition Based on Classes Distinctive Features

Viktor Krasnoproshin
Faculty of Applied Mathematics
and Computer Science
Belarusian State University
Minsk, Belarus
krasnoproshin@bsu.by

Vadim Rodchenko
Faculty of Mathematics
and Informatics, Yanka Kupala
State University of Grodno
Grodno, Belarus
rovar@grsu.by

Anna Karkanitsa
Faculty of Mathematics
and Informatics, Yanka Kupala
State University of Grodno
Grodno, Belarus
a.karkanica@grsu.by

***Abstract.* In pattern recognition, the approach where Supervised Learning is reduced to the construction of decision rules is considered to be classical. These rules should ensure an extremum of some criterion. The paper proposes an alternative solution based on the search for combinations of features that ensure classes separation. The results of a numerical experiment on model data confirm the effectiveness of the proposed approach.**

***Keywords:* pattern recognition, classification, instance-based learning, discovering patterns**

I. INTRODUCTION

Pattern recognition is one of the most important fundamental problems in computer science. The progress in developing artificial intelligence technologies largely depends on theoretical achievements in this area [1–4].

According to classical definition, pattern recognition is the assignment of initial data to a certain class. This is met by the selection of significant distinctive features from the general data set [5].

The growing computing power, progress in big data sets availability and the accumulated practical experience have led to a new perspective on the pattern recognition problem. In particular, C.M. Bishop writes in his monograph: "the field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories" [6].

We propose to refrain from the traditional approach when training is focused on building a classification algorithm. An alternative approach is proposed in the paper. The combinations of features that ensure classes distinction are identified in the learning process.

The effectiveness of the approach is confirmed by the results of a numerical experiment on model data.

II. ABOUT STATEMENT OF THE RECOGNITION PROBLEM

The classical description of the recognition (classification) problem is as follows:

Let the set of objects is given. The set is divided into subsets called classes. Partial information about classes, descriptions of the entire set and an object are given. Does object belong to a certain class is unknown. It is required to establish object's belonging to one of the classes according to the given information [7].

Currently, this problem is traditionally being solved in the following statement:

Let the objects' descriptions X and the acceptable answers Y for objects classification are given. It is assumed that there is an unknown target dependency $y^* : X \rightarrow Y$*, which values* $X^m=\{(x_1, y_1),\ldots, (x_m, y_m)\}$ *are known only for a finite set of objects (training set objects).*

It is required to construct an algorithm $a : X \rightarrow Y$*, that would approximate this target dependency on the entire set X* [8].

To solve such a problem, a family of algorithms is firstly set (up to parameters). Then, in the learning process (Supervised Learning), such parameters values are found that provide the extremum of the selected criterion.

Practical experience of usage the described scenario has revealed a number of serious shortcomings:

1) Selecting a family (model) of algorithms is a nontrivial problem. So, we are dealing not so much with a science as with the art of building algorithms;

2) Learning can be implemented only in an automated mode. The resulting algorithm a: $X \rightarrow Y$ is a «black box» approximating the unknown target dependency. The obtained dependency can hardly if any be interpreted in terms of the subject domain;

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

3) Finding a solution (based on the training set X^m) is carried out only in the initial space of object description. The question of the existence of subspaces in which the problem is solved more efficiently remains open [9].

To overcome the mentioned disadvantages an alternative approach is proposed. It is based on the hypothesis of compactness: «in the objects space, classes form compactly localized subsets» [10, 11].

In this case, the mathematical statement of recognition problem can be formulated as follows:

Let the objects' descriptions X and the acceptable answers of objects classification Y are given. There is an unknown target dependency $y^*: X \to Y$, *which values* $X^m=\{(x_1, y_1),\dots,(x_m, y_m)\}$ *are only known for the training set objects.*

It is required to find feature subspaces where class patterns do not intersect. Based on these patterns it is necessary to construct an algorithm $a: X \to Y$ *that would approximate the target dependency on the entire set X.*

To solve the recognition problem in such statement, it is firstly proposed to find feature spaces in which the classes do not intersect. After that, the construction of the classification algorithm becomes a trivial procedure.

III. SUPERVISED LEARNING ALGORITHM

Let the training set $X^m=\{(x_1, y_1),\dots,(x_m, y_m)\}$ is formed on the basis of an a priory dictionary of features $F=\{f_1, \dots, f_n\}$.

Let $V=\{v_1, \dots, v_q\}$ denote the set of all possible combinations of features from F. Then V contains $q=2^n-1$ subsets.

The searching algorithm for combinations of features on the set $V=\{v_1, \dots, v_q\}$ for which the class patterns do not intersect is as follows.

Algorithm

Step 1. Select from V a subset $V^+=\{v^+_1, \dots, v^+_n\}$, where v^+_i contains only one feature.

Step 2. For each v^+_i build the class patterns and compare their mutual placement.

Step 3. If the class patterns do not intersect, then the feature v^+_i is included in the set $V^*=\{v^*_1, \dots, v^*_k\}$.

Step 4. Exclude the subset $V^+=\{v^+_1, \dots, v^+_n\}$ from the set $V=\{v_1, \dots, v_q\}$ and obtain $V^\wedge=\{v^\wedge_1, \dots, v^\wedge_p\}$.

Step 5. Exclude from $V^\wedge$ all combinations $v^\wedge_i$that contain any combination from $V^*=\{v^*_1, \dots, v^*_k\}$.

Step 6. Pick the next combination $v^\wedge_i$ from $V^\wedge$ and, on its basis, construct the feature subspace.

Step 7. In this feature subspace, build class patterns and compare their mutual placement.

Step 8. If class patterns do not intersect then the combination of features $v^\wedge_i$ is included in the set V^*, and from $V^\wedge$ we exclude all combinations that contain $v^\wedge_i$.

Step 9. The process is repeated until $V^\wedge$ becomes empty.

As a result of the described algorithm, the set $V^*= =\{v^*_1, \dots, v^*_t\}$ where $0 \le t \le q$, will be constructed.

Based on the combinations $v^*_i \in V^*$, we formulate the previously hidden and empirically revealed pattern: «in the feature space of the subset v^*_i the classes do not intersect».

It should be noted that within a specific applied problem the revealed combinations of features v^*_i can be interpreted. This, in turn, means the possibility to interpret all the revealed patterns.

The combinations of features $v^*_i \in V^*$ define decision spaces where class patterns do not intersect. For class patterns inside such spaces, the condition of the compactness hypothesis is hold. Therefore, the construction of classification algorithms does not cause any difficulties.

IV. RESULTS OF THE NUMERICAL EXPERIMENT

Let's show the results of solving the recognition problem on the model data example.

Example. Let the given:

– two classes of objects: **M5 (multiples of 5)** and **NM5 (not multiples of 5)**;

– a priori dictionary of features F = {units, tens, hundreds, thousands, tens of thousands, hundreds of thousands, millions};

– a training set consisting of 20000 seven-bit integers, including 10000 multiples and 10000 not multiples of 5.

Table I shows the research results of class patterns intersection based on the features *units* and *tens*, where $NM5_i$ = Number of $M5_i$ for the i-th digit; $NNM5_i$ = Number of $NM5_i$ for the i-th digit;

TABLE I. RESULTS FOR FEATURES UNITS, TENS

Digit	**Units**		**Tens**	
	M5	***NM5***	***M5***	***NM5***
0	5043	0	1004	1043
1	0	1253	1003	1049
2	0	1252	966	939
3	0	1238	1020	942
4	0	1235	1034	1010
5	4957	0	1095	992
6	0	1218	933	1018
7	0	1226	992	991
8	0	1266	984	1003
9	0	1312	969	1013

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

$$a_i = \begin{cases} NM5_i + NNM5_i, NM5_i = 0 \lor NNM5_i = 0 \\ 0, NM5_i > 0 \land NNM5_i > 0 \end{cases}$$

$$Intersection = \frac{20000 - \sum_{i=0}^{9} a_i}{20000} * 100\%$$

Table II shows the study results of class patterns intersection based on the features *hundreds* and *thousands*.

TABLE II. RESULTS FOR FEATURES HUNDREDS, THOUSANDS

Digit	**Hundreds**		**Thousands**	
	M5	**NM5**	**M5**	**NM5**
0	988	989	1064	1005
1	987	986	974	1020
2	999	1006	1048	968
3	1034	989	956	991
4	980	972	994	979
5	1043	1045	978	1002
6	935	996	1006	999
7	994	980	1042	1036
8	1020	981	966	1025
9	1020	1056	972	975

Table III shows the study results of class patterns intersection based on the features *tens of thousands*, *hundreds of thousands* and *millions*.

TABLE III. RESULTS FOR FEATURES TENS OF THOUSANDS, HUNDREDS OF THOUSANDS, MILLIONS

Digit	**Tens of thousands**		**Hundreds of thousands**		**Millions**	
	M5	***NM5***	***M5***	***NM5***	***M5***	***NM5***
0	1042	1070	998	1020	1072	1012
1	1013	978	975	1018	1014	1026
2	970	990	997	994	948	1029
3	1009	972	957	987	950	1003
4	1026	961	1009	1027	999	994
5	1008	1028	1005	943	1039	989
6	990	973	942	992	979	978
7	988	1017	1045	964	961	980
8	953	1009	1025	1032	1022	1001
9	1001	1002	1047	1023	1016	988

Table IV shows the analysis results for all features from the a priori dictionary.

TABLE IV. EXPERIMENT RESULT FOR ALL FEATURES

Feature name	**Intersection (%)**
units	0.0
tens	100.0
hundreds	100.0
thousands	100.0
tens of thousands	100.0
hundreds of thousands	100.0
millions	100.0

Tables 1 and 4 show that:

- multiples of 5 have no digits 1, 2, 3, 4, 6, 7, 8, 9 at the units' place; not multiples of 5 – have no digits 0, 5;
- the feature *units*provides an absolute separation of classes M5 and NM5, since *Intersection*=0.0%.

In conclusion, let's note that the algorithm running time is 0.35 seconds.

V. CONCLUSION

An alternative option of the recognition problem statement and solution are considered in the paper.

In the present case, in the learning process we determine combinations of features of a priori dictionary that ensure classes distinction. Then, on the base of revealed feature combinations a classification algorithm is constructed.

The results of solving the recognition problem are shown on the model data example.

REFERENCES

[1] N. G. Zagoruiko, Applied methods of data analysis and knowledge. Izdatel'stvo Instituta matematiki SO RAN. Novosibirsk, 1999, pp. 268 (in Russian).

[2] P. A. Flach, Machine learning. The art and science of algorithms that make sense of data. Cambridge University Press, 2012, pp. 396.

[3] V. I. Vasilev, The problem of training for pattern recognition. Vishashkola. Golovnoe izdatel'stvo. Kiev, 1989, pp. 64 (in Russian).

[4] V. V. Krasnoproshin and V. G. Rodchenko, "Learning by Precedents Based on the Analysis of the Features Properties", Doklady BGUIR, №6 (2017), p. 35–41 (in Russian).

[5] (2021, June) Pattern recognition – Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Pattern_recognition.

[6] C. M. Bishop, Pattern Recognition and Machine Learning. Springer-Verlag, New York, 2006, pp. 738.

[7] Yu. I. Zhuravlev, "On an algebraic approach to solving problems of recognition and classification", Problems of Cybernetics, № 33 (1978), pp. 5–68.

[8] (2021, June) Supervised learning – Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Supervised_learning.

[9] V. V. Krasnoproshin, V. G. Rodchenko, "Classification Based on Decision Spaces", Doklady BGUIR, №6 (2019), pp. 20–25 (in Russian).

[10] M. M. Bongard, The problem of recognition. Nauka, Moscow, 1967, pp. 320 (in Russian).

[11] V. V. Krasnoproshin, V. G. Rodchenko, "Cluster Structures and Their Applications in Data Mining", Informatics, № 2, 2016, pp. 71–77 (in Russian).

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

New Algorithm for Building Effective Model from Prediction Models Using Parallel Data

Zurab Gasitashvili
Deputy Rector for Science of Georgian Technical University, Professor
Tbilisi, Georgia
zur_gas@gtu.ge

Merab Phkhovelishvili
N. Muskhelishvili Computation Mathematic Institute of Georgian Technical University
Tbilisi, Georgia
merab5@list.ru

Natela Archvadze
Dept. of Computer Sciences
Ivane Javakhishvili Tbilisi State University
Tbilisi, Georgia
natela.archvadze@tsu.ge

Abstract. **Building much more effective new hybrid models from prediction models, using parallel data is discussed. The algorithm for selection of model pairs and its advantage over any best prediction model is provided. The advantage of prediction models with higher number of pairs over lower number of pairs is shown and the algorithm of taking into consideration the "approximate coincidence" of predictions is discussed when selecting pairs.**

Keywords: **parallel data, prediction models, approximate accuracy, probablity of prediction success**

I. INTRODUCTION

Some prediction models are based on using of "parallel data" [1-4], although it must be noted that the term "parallel data" is differently explained in each of them.

In practice, parallel data is used during prediction of various events, including natural disasters: earthquake, landslide, tsunami, mudflow, etc., for prediction economical (business, macro economy), political events (elections, positions of political forces), for effective solving of prediction tasks in the sphere of medicine and other fields.

The definition of parallel data is based on introduction of new type of dependence between the data, which is called "parallelism between the data" [1, 5-8]. Parallelism between the data is mutual dependence between those data, which are used for prediction of the same event. Various data affecting the same event may exist in different periods (parallel by time) or locations (parallel by location) and/or provide other additional information on prediction of the same event [9, 10].

The main idea of algorithms for building of prediction models is reviewed by us through parallel data and is the following: Let us assume that there are several models of prediction. From them, it is necessary to select such pairs, triplets, etc. from several models, which give much better result than a single best model from them or two models separately.

This algorithm was the following: such models were found, for which the number of coincidences of unsuccessful predictions for some given event was as low as possible, but successful predictions were necessary for them.

In this paper we first review static prediction models for natural disasters, when a result(s) of prediction should be guessed, for example, when, where and with which specifications occurred the event of interest.

Unlike static predictions, a prediction is dynamic, when for each time interval it is necessary to forecast an event of certain value. Such is, for example, a daily forecast of exchange rate, forecast of oil price, monthly subsistence level, annual income, human health condition, scope of coronavirus spread, etc.

The distinctive sign, by which the static prediction is different from the dynamic one, is its dependence on the time of prediction event. Actually it means that we should distinguish, how a result, i.e. prediction values, are declared. If it occurs continuously, with some predefined time interval, then this is dynamic prediction, but if time is one of prediction elements, then it is static prediction. For example, earthquake prediction implies declaring that date as one of the results, when earthquake is expected, therefore, it belongs to static prediction, and currency exchange rate is forecasted daily, therefore, it is dynamic prediction.

In this article we will establish 4 lemmas for the task of static prediction and show, how the accuracy of such models is increased through our algorithm. Specific data are taken for earthquake prediction task. Each prediction model is build based on certain predecessors. For earthquake the predecessor is geophysical phenomenon (mainly), which precedes the actual earthquake. For their part, geophysical precursors are divided into the following categories: seismic, hydro geodynamic, deformation, geochemical, thermal, gravitational, electromagnetic and, precursors obtained via remote monitoring by means of satellite technologies developed recently [11].

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Despite the fact that quite high number of predecessors exist, not any of them ensures high-accuracy prediction for time, place and magnitude of future earthquake. The probability of successful prediction of each predecessor (ratio of number of successful predictions to the number of all given predictions) does not exceed 0.5% [12]. One of the ways for overcoming this situation is to use several prediction predecessors simultaneously, although for each of them it is necessary to perform observation for a long time and process vast amount of data, which is not done in many models till now. “The practice of recent years show that their simultaneous use would improve the reliability and efficiency of prediction assessment, at least in medium-term (first years) prediction”.

II. “SUCCESS PROBABILITY” OF PREDICTION

Assume that we have several prediction models, which provide some predictions through their predecessors (for example, for earthquakes - when it would occur, at which location and with which magnitude). These predecessors should be “necessary predecessors” that means that if earthquake occurs, they will inevitably provide the prediction. If some predecessors do not provide prediction on actually occurred earthquake, it will be no longer considered.

We study history, let's assume that there is plenty of data and it is necessary to calculate, based on predecessors, how many times the prediction of earthquake occurrence was given and how many times actual earthquake occurred. Assume that we consider the necessary predecessors and the models created for them: $A_1, A_2, ..., A_n$, where n is the number of considered predecessors. t denotes time, during which we perform analysis and the number of actually occurred earthquakes is *m*. We calculated the number of earthquakes predicted by each predecessor: $p_1, p_2, ..., p_n$. For example, A_i model, which was based on *i* predecessor, predicted earthquake occurrence p_i - times.

For each p_i let’s calculate quotients of *m*, the number of actually occurred earthquakes, write it in % and designate with K_i:

$$K_i = \frac{m}{p_i} 100\%.$$

For example, if earthquake actually occurred 4 times, and we calculate $K_i = \frac{4}{20} 100\% = 20\%$ then the probability of A_i success will be 20%.

Put the sequence of model success in descending order and this sequence denote as: $k_1, k_2, ...k_n$ sequence. k_i is a model created for i -th predecessors. We get that k_1 highest value, which was determined by the prediction, the value of k_2 is less than that of k_1 and so on.

It is necessary to consider a combination of models (two, three, etc.) and assessment of the probability of their combined success. The assessment and selection of combinations is done according to the parallel probabilities [13].

Lemma 1 - If such pairs of models are selected, for which the number of coincidences of unsuccessful predictions for some given event was as low as possible, but successful predictions were necessary for them, then the success probability calculated for combination of any pair so selected is always higher than or equal to the success probability of best from them.

If we take best k_i and pair it with any, even the worst value k_j, then their combined result is not worse than k_i. Proof is based on fact that for the pair k_i and k_j (ki< k_j), then the intersection of their successes, of course, is less than or equal to k_i. For example, if k_j gives conclusion that earthquake occurs 5 times, even if others give values 10 or 7, the intersection of their successes cannot exceed 5. Therefore, whether pairs, triplets, quads are selected, their combination always give better result than best of them.

Lemma 2 - The higher number of intersections of prediction models, the better prediction we would get.

For example, the best triplet - combination of three predictions would give better result than the best pair of prediction (deuce), the best quad gives better result than the best triplet and so on.

This follows from the fact that intersection of any pairs with third is lower number than each pair.

III. SELECTION OF BEST PAIRS

Lemma 3 - The best pair is one that does not have intersection between each, except actual, occurred predictions.

For example, we calculated the pair or k_i and k_j predictions and let's calculate, by their combined prediction, which number of coincidences we have with the actual situation (for example, coincided 10 times). For example, the event occurred actually 2 times, if we calculate success %, we get $\frac{2}{10} 100\% = 20\%$, but if it

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

turned out that k_i and k_j jointly only two times had prediction success for the event, then it means that k_i and k_j is the best pair.

It is an interesting metamorphose - it turns out that if we consider all those models jointly, which have many errors, their combination may give the best result.

IV. MODIFICATION OF PREDICTION MODEL

Geophysical characteristics of environment constantly change, for example, an average temperature, erosion of ocean shores, etc. The question is, how to plan the change of selected scheme of prediction, from which time pairs, triplets, etc. should be selected and success probabilities recalculated, what we should do, recalculate everything that occurred till today?

Lemma 4 - Modification of prediction model, i.e. recalculation should be done from the day of last earthquake occurrence.

For consideration of relevant pairs of models, it is necessary to take corresponding figures from such moment, when we have the aggregate of all input data. Of course, it is possible that new prediction models may be introduced with the data of relevant predecessors, and additional regulations are required for consideration, because of search for the relevant pair. It is possible that after each actually occurred event, the selected pairs of prediction models would be changed and other pairs would become better for prediction. Therefore, selection of each new pair should be done after occurrence of each event during static prediction, and for dynamic prediction, the process of determination of such pairs should be regulated within certain time periods. For example, if we have daily prediction data, new pairs should be selected at least once a week.

V. "APPROXIMATE COINCIDENCES" OF PREDICTIONS

In accordance with Lemma 2, the prediction pairs are selected. When selecting them, we determined the number of "accurate" coincidences of predictions. Now this is not sufficient and a "coincidence accuracy" should be determined. Of course, prediction data should not directly coincide with each other, but coincide within certain intervals of time, place or other characteristic.

Lemma 5 - When selecting the pairs of prediction models, an "approximate coincidence" should be taken into account.

An interval of "approximate coincidence" is determined with the help of experts. It may be prediction of such time period, when occurrence of given event is expected, or determination of certain radius from the epicenter. This task is faced in cases of earthquake, virus origin, beginning of military conflict, etc. Of course, time interval has great importance. For earthquakes, a short-term forecast - where (with radius of 50 km), when (with interval of 24 hours), and with which magnitude (with difference of 0.5) the earthquake is expected.

SPATIAL MODELS

If we have only 3 data and build given prediction points in the relevant 3-dimensional space: in this case, x is location, t is time and v is power. Assume, that each has its own dimension. For example: Location - plain. In this case, 4, 5 of more dimensional model will be built, depending on how much parameters are in prediction. Prediction data are presented in 3-dimensional space on Fig. 1. Here the distance between two points (predictions) is the error between their predictions. The value of "approximate coincidence" (i.e. this distance) is determined by an expert.

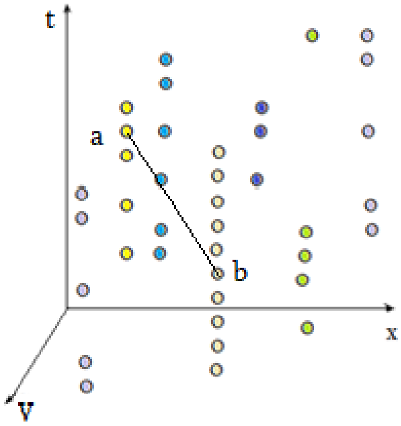


Fig. 1. Presentation of prediction data in space

Introduction o "approximate coincidences" sharply increases the volume of calculations needed for selection of pairs of prediction models. It is necessary to utilize higher computational capacities, technical capabilities of supercomputers and use the algorithms of parallel computations and relevant programs. It should be noted that, in our opinion, this can be realized by using powerful parallel and recursive computations of programming language F#, as we prove it in [14].

VI. CONCLUSION

We considered the possibility to build much more effective new hybrid models from prediction models, using parallel data, and by means of lemmas we state that: The advantage of selection of model pairs over any best model of prediction, it was shown that the more is the number of model pairs, the more is advantage over the lesser numbers of pairs.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The best pair is one that does not have intersection between each, except actual, occurred predictions. When to reselect the pairs of prediction models and that it is necessary to consider “approximate coincidences” of predictions in this case.

REFERENCES

[1] Z. Gasitashvili, M. Pkhovelishvili, N. Archvadze, Prediction of events means of data parallelism. Proceedings - Mathematics and Computers in Science and Engineering, MACISE 2019, pp. 32–35, 8944725, 2019. https://ieeexplore.ieee.org/abstract/document/8944725.

[2] Y Chen, Y Lv, FY Wang. Traffic flow imputation using parallel data and generative adversarial networks - IEEE Transactions on Intelligent , 2019.

[3] J. Bhimani, N Mi, M Leeser, Z Yang. New performance modeling methods for parallel data processing applications - ACM Transactions on Modeling and 2019.

[4] D Skillicorn. Strategies for parallel data mining. - IEEE concurrency, 1999 .

[5] N. Archvadze, M. Pkhovelishvili. Prediction of Events by Means of Data Parallelism.Proceedings of International Conference on Matematics, Informatics and Informtional Technologies (MITI2018). pp.120-121, 2018.

[6] Z. Gasitashvili, M. Pkhovelishvili, N. Archvadze. Usage on Different Types of Data to Solve Complex Mathematical Problems. WSEAS Transactions on Computers, vo. 18, Art. no. 7, pp. 62-69, 2019.

[7] M. Phkhovelishvili, N. Jorjiashvili, N. Archvadze. Usage of heterogeneous data and other parallel data for prediction problems. PRIP’2019. Pattern Recognition and Information Processing (Proceedings of 14th International Conference (21-23 May, Minsk, Belarus). pp. 178–181, Minsk,”Bestprint”, 2019.

[8] M. Phkhovelishvili, N. Jorjiashvili, N. Archvadze. Using Different Types Data Operations for Solving Complex Mathematical Tasks. Computer Science and Information Technologies. Proceedings of the conference (September 23-27, 2019), Yerevan, Armenia, pp. 187–190, 2019.

[9] M. Pkhovelisvili, M. Giorgobiani, N. Archvadze, G. Pkhovelishvili. Modern Forecasting Models in Economy. Proceedings of Materials of International Scientific Conference „Modern Tendencies of Development of Economy and Economic Science“. Ivane Javakhishvili Tbilisi State University Paata Gugushvili Institute Of Economics. pp. 219-224, 2018.

[10] N. Archvadze, M. Pkhovelisvili. Modern Forecasting Models in Economy X International Conference of the Georgian Mathematical Union. Book of abstracts, pp. 55, 2019.

[11] A. D. Zav'yalov Prognoz zemletryaseniy: sostoyaniye problemy i puti resheniya, v zhurnale Zemlya i vselennaya, № 5, pp. 66–79, 2018 (in Russian).

[12] A. D. Zav'yalov Srednesrochnyy prognoz zemletryaseniy: osnovy, metodologiya, realizatsiya, Izdatel'stvo Nauka; 2006 (in Russian).

[13] Z. Gasitashvili, M. Phkhovelishvili, N. Archvadze, N. Jorjiashvili. An Algorithm of Improved Prediction from Existing Risk Predictions. Published by AIJR Publisher in “Abstracts of The Second Eurasian RISK-2020 Conference and Symposium” April 12-19, 2020, Tbilisi, Georgia, pp. 31, 2020.

[14] N. Archvadze**,** M. Pkhovelishvili. Reforming the Trees – C# and F# comporation. International Conference on “Problems of Cybernetics and Informatics" (PCI'2012). pp. 93-96, 2012.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Search for Genomic Mutations Associated with Drug-resistant Tuberculosis

Katsiaryna Rumiantsava
DMA FAMCS of
Belarussian State University
Minsk, Belarus
kattytsurikova@gmail.com

Andrei Gabrielian
Office of Cyber Infrastructure &
Computational Biology
National Institute of Allergy and
Infectious Disease, National
Institutes of Health
Bethesda, MD, USA
gabr@niaid.nih.gov

Alex Rothenthal
Office of Cyber Infrastructure &
Computational Biology
National Institute of Allergy and
Infectious Disease, National
Institutes of Health
Bethesda, MD, USA
alexr@niaid.nih.gov

Roman Sergeev
United Institute of Informatics
Problems of NAS of Belarus
Minsk, Belarus
roma.sergeev@gmail.com

Alexander Tuzikov
United Institute of Informatics
Problems of NAS of Belarus
Minsk, Belarus
tuzikov@newman.bas-net.by

***Abstract.* The problem of drug-resistant tuberculosis, its diagnosis and treatment, is especially relevant today. Every year the causative agent of the disease becomes more and more resistant to existing drugs. Here we analyzed 1244 tuberculosis cases with available results of phenotypical assays for drug resistance as well as tuberculosis genome sequences using single-marker and multi-marker tests.**

***Keywords:* drug-resistance, tuberculosis, phylogenetic tree, single-marker tests, multi-marker tests**

I. INTRODUCTION

The emergence of high-throughput sequencing methods for determining the DNA nucleotide sequences of living organisms has become a driving force for biological research. However, the genetic code itself does not have great practical value until the necessary information is extracted from it. The analysis of decoded genomic sequences often leads to large-scale problems, where the number of unknown parameters is measured in tens of thousands with a relatively small number of available observations. One of these tasks is the search for mutations in the genomes of microorganisms of bacterial nature, which are associated with the presence of drug resistance.

Mathematical analysis of genome-wide data on Mycobacterium tuberculosis allows predicting resistance to first-line drugs with a high probability. At the same time, resistance to second-line drugs is poorly explained only by genomic mutations. This requires a deeper study of all available data. In particular, a comparative analysis of genomic data will provide relevant information with already known cases.

II. METHODS

A. Strain collection and phylogeny

In this paper we used data collected from the Drug Resistant Tuberculosis Project [9], https://tbportals.niaid.nih.gov. At the first stage, duplicates and cases with conflicting results for drug resistance were removed and 944 cases were further analyzed. We investigated resistance to first-line drugs: isoniazid, rifampicin, pyrazinamide, ethambutol and streptomycin, as well as second-line drugs: fluoroquinolones and aminoglycosides. For each drug, we formed the two case-control group.

Case-control studies are sensitive to population separability of samples [1]. To identify population subgroups phylogenetic trees were constructed with two different methods: neighbour joining and maximum likelihood. As a result, initial data were divided into two subgroups.

To reduce the dimension of tasks and improve quality we used minor allele frequency (MAF) equal to 0.01.

B. Single-marker tests

Single-marker tests are used to test associations between observed drug resistance and individual mutations [2]. We used modifications of classical statistical tests as single-marker tests: Fisher's exact test and Cochran-Mantel-Haenszel test. Both methods are based on building contingency tables of the following form.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

TABLE I. CONTINGENCY TABLE CONSIDERED IN SINGLE-MARKER TESTS TO SEARCH FOR MUTATIONS ASSOCIATED WITH DRUG RESISTANCE

drug susceptibility	**Presence of mutation**		
	present	***absent***	***total***
sensitive	n_{00}	n_{01}	n_{0*}
resistant	n_{10}	n_{11}	n_{1*}
total	n_{*0}	n_{*1}	n_{**}

The Cochran-Mantel-Haenszel test, in contrast to the Fisher test, takes into account population subgroups. We need to make adjustments for the population subdivision because the variation in the frequency of occurrence of some mutations can be explained by their belonging to different populations not by drug susceptibility. To reduce the likelihood of errors in multiple hypothesis testing, we used the Bonferroni correction.

C. Multi-marker tests

Multi-marker tests unlike single-marker consider additive effects between mutations. In this study, we used an algorithm for searching for combinations of mutations boosting [1] and factor analysis of mixed data with linear mixed model.

In addition to genomic data, we also considered phenotypic traits of samples. Such data include both categorical and continuous features. Therefore, we used factor analysis of mixed data (FAMD) to transform them into a set of uncorrelated features.

The linear mixed model is used for regression on hierarchical data [4]. It considers population subgroups in original data. The main idea of the linear mixed model is that it takes into account both fixed and random effects. The fixed effects are the basic regression on the data. Each subgroup may have its own unique characteristics or traits, which can be expressed in presence of common mutations that are not associated with drug resistance, and vice versa, the peculiarities of diagnosis and treatment of tuberculosis and the health system itself. There may also be similarities in the closest subgroups. The linear mixed model considers such subgroup effects as random effects. The model can be described with the equation:

$$Y = X\beta + Z\upsilon + \varepsilon. \tag{1}$$

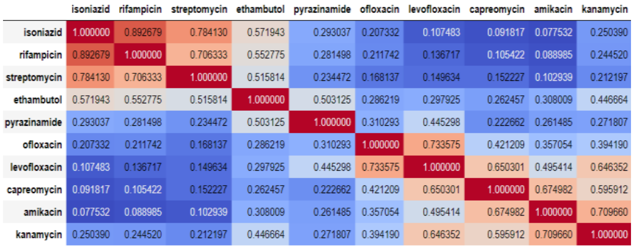
Where Y is a phenotype vector, X – matrix with data, β – vector of effects that determine the significance of a set of mutations, Z – matrix of random effects of subgroups, υ – vector of random effects for subgroups, ε – vector of residues that cannot be explained by the model.

To control how well the model works we used cross-validation for a given number of blocks [5].

Another approach we used in the study is the search for combinations of mutations. The main feature of this task is that total number of considered mutations significantly exceeds the number of observations. The task of enumerating all the possible combinations of mutations requires large computational resources. To avoid this problem, one can switch from the task of enumerating mutations to the task of enumerating samples, since their number is much smaller. An exhaustive description of the algorithm is given in [1, 3].

III. RESULTS

First, we carried out analysis of the results of biological tests for drug resistance that revealed cross-resistance between drugs. Selected correlation coefficients calculated from the results of biological tests are presented in Fig. 1.



	isoniazid	rifampicin	streptomycin	ethambutol	pyrazinamide	ofloxacin	levofloxacin	capreomycin	amikacin	kanamycin
isoniazid	1.000000	0.892679	0.784130	0.571943	0.293037	0.207332	0.107483	0.091817	0.077532	0.250390
rifampicin	0.892679	1.000000	0.706333	0.552775	0.281498	0.211742	0.136717	0.105422	0.088985	0.244520
streptomycin	0.784130	0.706333	1.000000	0.515814	0.234472	0.168137	0.149634	0.152227	0.102939	0.212197
ethambutol	0.571943	0.552775	0.515814	1.000000	0.503125	0.286219	0.297925	0.262457	0.308009	0.446664
pyrazinamide	0.293037	0.281498	0.234472	0.503125	1.000000	0.310293	0.445298	0.222662	0.261485	0.271807
ofloxacin	0.207332	0.211742	0.168137	0.286219	0.310293	1.000000	0.733575	0.421209	0.357054	0.394190
levofloxacin	0.107483	0.136717	0.149634	0.297925	0.445298	0.733575	1.000000	0.650301	0.495414	0.646352
capreomycin	0.091817	0.105422	0.152227	0.262457	0.222662	0.421209	0.650301	1.000000	0.674982	0.595912
amikacin	0.077532	0.088985	0.102939	0.308009	0.261485	0.357054	0.495414	0.674982	1.000000	0.709660
kanamycin	0.250390	0.244520	0.212197	0.446664	0.271807	0.394190	0.646352	0.595912	0.709660	1.000000

Fig. 1. Pairwise correlations between biological test results

In this work, we constructed two phylogenetic trees for 46 genes of 944 sequences using different approaches: neighbour joining [6] and maximum likelihood [7]. Despite the difference in the lengths of branches and variation in the locations of the end nodes, identical subgroups are identified in both trees.

We used the R software functions from the stats package: fisher_test for Fisher's exact test and mantelhaen_test for the Cochran-Mantel-Haenszel test. We calculated p-values for all corresponding mutations in the data matrix and consider only those with p-value greater than 5×10^{-8}. To visualize test results Manhattan charts were used. The diagram displays a set of points, the coordinate of which on the abscissa is the position of the mutation, and on the ordinate is the negative logarithm of the p-value. Most of the points on the graph are close to abscissa axis. The higher points along the ordinate axis, the lower are their p-values, which mean the greater statistical significance.

To validate the results obtained, we compared them with the drug resistance mutations published in [8].

Single-marker tests revealed a set of mutations associated with drug resistance to individual first-line drugs, their combinations, as well as groups of second-line drugs. However, they were unable to identify

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

mutations associated with resistance to individual second-line drugs, but only their combinations.

For factor analysis of mixed data, we used the prince python package, the implementation of a mixed linear model from the statsmodel package; cross-validation of the generalizability of the model was performed using StratifiedKFold over five partitions from the sklearn package. To validate classification results we used four classical metrics in machine learning: precision, recall, f-score and accuracy.

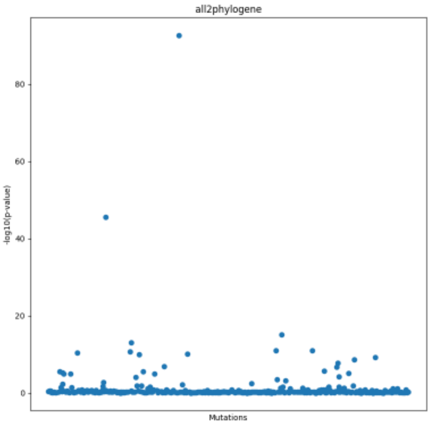


Fig. 2. Manhattan Plot of Cochran-Mantel-Haenszel test results (example) for first-line drugs

For first-line drugs and their combinations, except ethambutol, all metrics values were greater than 0.8. The percentage of true positives among all positives is high (95% and above) for all such samples. The proportion of predicted drug resistance among all is about 80-90%. The proportion of correctly classified samples is slightly higher than the previous one. It is important that the classifier covers as many cases of drug resistance as possible and only few sensitive.

For the second-line drugs, more or less good metric values were obtained for fluoroquinolones, including ofloxacin. The accuracy of the classification is about 80%; the recall is a bit lower, about 75%. For aminoglycosides, the classification results are much worse. For capreomecin the algorithm failed to build the model. For amikacin, the percentage of true positives among all positives turned out to be below 50%, which is equivalent to random labeling. The classification results are slightly better for kanamycin, on average, all metrics values were in the region of 0.66 - 0.71, but these are also low estimates. Based on the classification results, it can be concluded that it was not possible to build a model for aminoglycosides that could predict drug resistance to these drugs with good probability.

In the study, we used H37Rv NC_000962.3 reference sequence; all mutations mentioned below correspond to that sequence. For all first-line drugs and their combinations, the mutation combination search algorithm found one dominant mutation, C2155175G, associated with isoniazid resistance. When boosting on these sets, the algorithm added to the dominant mutations new ones related to the markers of phylogenetic lines, while the accuracy of the classifiers did not increase.

Considering all mutations occurring in one position as one mutation, the algorithm has identified a set of mutations: 7570, 7572, 7581, 7582 associated with resistance to fluoroquinolones (according to information from TBDreamDB [8]). Despite the low accuracy, the result for fluoroquinolones, except levofloxacin, is considered satisfactory.

For aminoglycosides, with the exception of kanamycin, scores were rather low. Each set of mutations contains A1473252 associated with resistance to this group of drugs. This is consistent with the results of single-marker tests. However, this mutation, even when combined with others, is not enough to build a good classifier. For kanamycin, the algorithm constructed a set of mutations, which included the G2715356A mutation associated with resistance to this particular drug.

IV. DISCUSSION

In this paper, we considered the problem of drug-resistant tuberculosis and solutions for the comparative analysis of mycobacterial genomes. Using the phylogenetic tree, two population subgroups were identified in the original dataset. We found cross-resistance between the individual drugs.

To search for mutations associated with drug resistance, we used single-marker and multi-marker tests. To validate the test results, the TBDreamDB database was used, which contains already known mutations associated with drug resistance. Single-marker tests revealed a set of mutations associated with resistance to individual first-line drugs, their combinations, as well as groups of second-line drugs. However, they were unable to identify mutations associated with resistance to individual second-line drugs. Multi-marker tests built good classifiers and identified combinations of mutations for individual first-line drugs and their groups. For second-line drugs, multi-marker tests have built combinations that include individual mutations associated with drug resistance, but they alone are not enough to build a good classifier.

ACKNOWLEDGMENT

We acknowledge support from TB Portals Consortium and the TB Portals Program.

REFERENCES

[1] R. Sergeev, Algorithms for analysis and search for associations in genetic data, PhD thesis, Minsk, 2019.

[2] X. Wang, N. J. Morris, D. J. Schaid, R. C. Elston, Power of single- vs. multi-marker tests of association, Genet Epidemiol, 2012, 36 (5), pp. 480-487.

[3] K. Tsurykava, R. Sergeev, Algorithm for searching combinations of mutations associated with drug-resistant tuberculosis, Minsk, BSU, 2021, pp. 130-132.

[4] J. C. Pinheiro and D. M. Bates, Mixed-effects models in S and S-PLUS, New York, NY u.a., Springer, 2000.

[5] J. Fox, Applied Regression Analysis and Generalized Linear Models. Sage Publications, Thousand Oaks, California, 2015.

[6] Rapid Neighbour Joining, Martin Simonsen, Thomas Mailund and Christian N. S. Pedersen, In Proceedings of the 8th Workshop in Algorithms in Bioinformatics (WABI), LNBI 5251, Springer Verlag, October 2008, pp. 113-122.

[7] M. Salemi, A.-M. Vandamme, Lemey, The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing, Cambridge, UK, Cambridge University Press, 2009.

[8] A. Sandgren, M. Strong, P. Muthukrishnan, Weiner BK, Church GM, Murray MB (2009) Tuberculosis Drug Resistance Mutation Database. PLoS Med 6(2): e1000002. https://doi.org/10.1371/journal.pmed.1000002.

[9] Alex Rosenthal, Andrei Gabrielian, Eric Engle, et. al. The TB Portals: an Open-Access, Web-Based Platform for Global Drug-Resistant-Tuberculosis Data Sharing and Analysis. Journal of Clinical Microbiology, vol. 77, no. 1, 2017, pp. 3261-3282. https://doi.org/10.1128/JCM.01013-17.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Joint Dataset for CNN-based Person Re-identification

Sviatlana Ihnatsyeva
Faculty of Information Technology
Polotsk State University
Novopolotsk, Belarus
ignateva604@gmail.com

Rykhard Bohush
Faculty of Information Technology
Polotsk State University
Novopolotsk, Belarus
r.bogush@psu.by

Sergey Ablameyko
United Institute of Informatics Problems of NAS of Belarus
Belarusian State University
Minsk, Belarus
ablameyko@bsu.by

***Abstract.* In this paper, we propose a joint dataset for person re-identification task that includes the existing public datasets CUHK02, CUHK03, Market, Duke, LPW and our collected PolReID. We investigate the training dataset size and composition effect on the re-identification accuracy. We carried out a number of experiments with different size of dataset to solve re-identification task. The results of experiments are presented.**

***Keywords:* large-scale dataset, cross domain, convolution neural network, PolReID dataset**

I. INTRODUCTION

Person re-identification (person ReID) is the process of identifying a person in another place or at different time using video surveillance systems. The ReID system extracts features of the query-image and compares them with features other persons in dataset's gallery. Convolutional neural networks (CNN) are most efficient for feature extraction.

Re-identification comes with a number of problems. People appearance may change in the course of movement, or different people may appear similar. There is also the problem of occlusion. At some points, a person part can be hidden by other people or landscape elements. Video cameras can have dissimilar resolutions, shooting at different times - different degrees of illumination, different camera positions will give different backgrounds, and this leads to the existence of such a problem as domain shift. This is of great importance when working with datasets, because each of them is a separate domain [1, 2]. Good increasing the accuracy value of the re-identification algorithm was shown by the random erasing method [3]. Random erasing is a method to increase dataset by adding images, in which an arbitrary image fragment is randomly deleted, which is filled with zero or random values. This method improves the algorithm's occlusions resistance. Currently, most re-identification systems use these augmentation methods.

The deep neural networks success makes it possible to achieve high results in the person re-identification problem [4] when the data for training and testing are independent and identically distributed. However, such models are well suited for a training set and will perform poorly in an invisible domain [5].

One of the approaches to increase the stability of the ReID system is to use a dataset that will have the maximum similarity with the data with which the re-identification algorithm will have to work. Another approach is to significantly increase the training dataset, which would include a huge number of identifiers and their images. Our paper discusses a problem of forming large dataset associations for re-identification systems.

II. EXISTED DATASETS

When training a re-identification system, a dataset is of great importance, and the more diverse the examples, the more robust the trained system will be. To increase the training set without using additional data, the simplest way is to add to the existing dataset images from the original dataset, which have undergone such manipulations as rotation, reflection vertically or horizontally, changes in brightness and contrast, color fluctuations.

In [6], a cross-domain mixup scheme is considered, and proposed scheme study is carried out, when training is carried out on the Market 1501 dataset, and testing on Duke, and then vice versa, is trained on Duke, and tested on the Market. The studies carried out have shown that the re-identification accuracy in the two considered examples is different, and it is impossible to say unambiguously how the system will behave on other data sets. A large experiments number with a different composition of training and test samples are carried out in [7], where the authors propose a new CNN framework for learn effective features, which allows to improve re-identification in the cross domain, and the authors conduct a study by training the model on one of the datasets.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Shinpuhkan2014dataset, CUHK02, CASPR, i-LIDS, PRID, and testing is performed on VIPeR, i-LIDS, Shinpuhkan2014dataset. The datasets used for training are small, which is probably one of the reasons for the low Rank1 scores. In [2], the authors propose an approach to generalize the subject area and consider training sample variants, with a different composition and number of various datasets included in its composition. Increased training set includes over 18000 IDs and almost 122000 bounding boxes by combining different datasets

In [8], authors strive to develop a universal framework for human ReID that can be generalized and work well on target domains. This work also uses an increasing strategy for the training sample at the expense of other datasets, and is conducting several cross-domain experiments, including a combined unified database that included Market, Duke, CUHK03, and MSMT17. This database includes almost 9000 identifiers and more than 220000 boxes. Dataset expansion made it possible to improve Rank1 from 33.9 when training on Duke and testing on the Market, to 82.3 when training on a combined database.

The most famous and significant in volume terms are datasets such as Market, Duke, CUHK02, CUHK03, LPW.

Market-1501 was assembled at Tsinhua University in supermarket front and includes 32668 hand-crafted bounding boxes for 1501 people. 12936 bounding boxes for 751 people are used for training, and 19732 bounding boxes in galleries for 750 people to test the re-identification algorithm. In addition, the dataset contains 2793 bounding boxes-distractors [9]. Duke MTMC-ReID is a subset of the Duke MTMC dataset acquired in March 2014 from the campus Duke University. The images were taken from 8 CCTV cameras located between the buildings, and include 36411 bounding boxes. To train the re-identification algorithm, 16522 images for 702 people are used. The remaining 17661 bounding boxes for 702 people are used for testing [10]. CUHK02 contains 1816 people and five camera views pairs. Each of them contains 971, 306, 107, 193 and 239 people, respectively. There are 4 images for each person - 2 from one camera, 2 from the another. The dataset was collected from The Chinese University of Hong Kong (CUHK) campus. The dataset contains sensitive data and the authors ask that the privacy of CUHK students be respected [11]. CUHK03 was obtained from the same campus of the Chinese University of Hong Kong, and contains 1467 people, each person has 5 images from 2 angles. This set, like CUHK02, is available only for academic research and its distribution is available only by agreement with the authors [12]. LPW (Labeled Pedestrian in the Wild) is obtained from three different scenes. The first scene includes images from 3 CCTV cameras, the other two scenes include 4 cameras. The full dataset contains 2731 people captured by at least two cameras. 7694 image sequences were generated, with an average of 77 frames per sequence, thus the total LPW dataset contains 592438 bounding boxes [13].

III. LARGE JOINT REID DATASET

We used two approaches to form a large ReID dataset. The first one was that we combined the existing datasets, which are presented in different formats and their structure is different. Our second approach involves the formation of a new images set, which is included in the joint database being created. Thus, the joint database consists of Market, Duke, CUHK02, CUHK03, LPW and PolReID.

We developed our own dataset, called PolReID [14]. For its formation, video sequences received from volunteers were used. For each person, there are from 1 to 7 video sequences with different locations, illumination levels, image quality and distance from the CCTV camera. Thus, images for most people also correspond to different domains. To extract bounding boxes from frames, the YOLOv4 detection algorithm implemented in pyTorch [15] was used. Incorrect bounding boxes were removed manually. For each person in the dataset, there are images with partial overlap, both horizontal and vertical. The same person is presented from different angles. In total, the dataset contains images for 54 people and includes 5609 bounding boxes. PolReID is split into training and testing data. For training, 30 people (3603 bounding boxes) are used, for testing - 24 people (2006 bounding boxes). Examples of images are shown in Fig. 1.

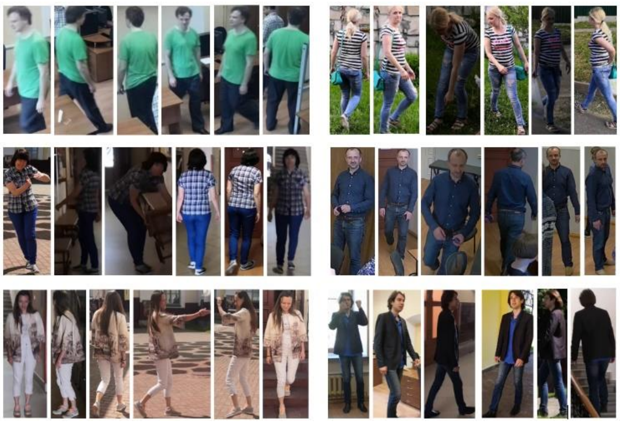


Fig. 1. Some images from PolReID dataset

When joining these datasets, considering the need to train algorithms and test them, the databases were also splinted. For Market and Duke, this task was accomplished in accordance with the protocols of the original documents.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

In LPW, 666 people (141466 bounding boxes) were randomly selected for testing, the rest were added to training data. When joining sets with CUHK02, CUHK03, Market and Duke were not divided into test and training sets, as in [16, 17], and all images were used for training.

Joining dataset is challenging. This is due to the fact that different datasets have a different way of writing names, different file locations in directories hierarchy. Directory names can contain useful information such as camera number from which the image was taken, scene number. In addition, identifiers and sequence camera number values can be the same in different datasets, but they will belong to different people. To avoid such a situation, when adding each new dataset, the maximum value that was used in the existing dataset was added to the ID value and camera number.

For correct re-identification algorithm operation, the image file names were brought to a single recording format: XXXXX_cYYsZZ_AAAAAA_BB.jpg, where XXXXX is the person's identifier, YY is camera number, ZZ is the video sequence number from this camera, AAAAAA is the frame number in video sequence, BB is the different people number whose images were obtained from this frame. If the dataset did not contain any required information (usually ZZ or BB), the value was set to 0. The capital letters number in the example corresponds to digit numbers.

The joint dataset includes 8690 identifiers and 537109 images.

IV. TRAINING AND RE-IDENTIFICATION

A. *Training model*

For re-identification, the algorithm proposed in [18] was used with hyper parameters specified in Table I.

TABLE I. HYPER PARAMETERS USED IN MODEL

Backbone network:	**DenseNet-121** [19], **ResNet-50** [20]
Droprate	0.5
Batch sise	32
Learning rate	0.05
Epochs	60

After epoch 40, decay learning rate by a factor of 0.1, and Fig. 2 shows that this has a positive effect on the convergence of the model.

We carried out experiments number with different consist increase the data for training the re-identification algorithm and testing with different data sets.

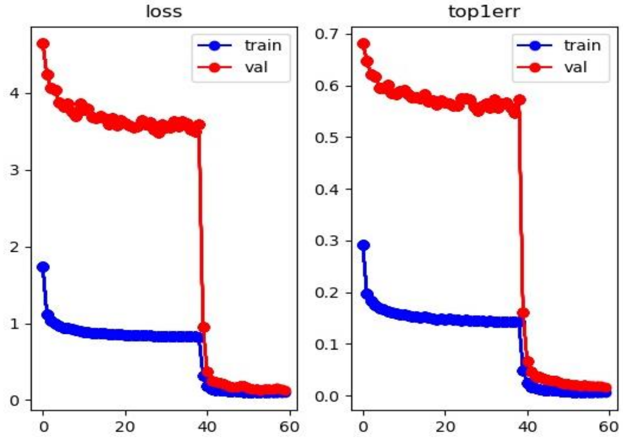


Fig. 2. Loss and top1 error graph during training re-identification model with backbone network DenseNet-121 on joint training sample

Model was trained for nine different training samples. Table II show the training sample consist and size.

This algorithm assumes that the trained neural network extracts features for each person located in the test sample gallery. Then, for each query, the all images feature table is ranked. The cosine distance is used as the similarity metric. The images obtained from the same camera as the request image are excluded from the ranked feature table. To assess the re-identification accuracy, we used metrics is Rank1, Rank5, Rank10 and mAP. The RankN metric is the ranking accuracy, i.e. the ratio of correctly obtained results to the total number of outputs among the N first issued results. mAP – this is mean Average Precision for a queries set is the mean of the average precision for each query.

TABLE II. PARAMETRS FOR TRAINING SAMPLES

Training datasets	**Size (ID/Bboxes)**
Duke	702 / 16522
Market	751 / 12936
LPW	2064 / 448568
Market, duke	1453 / 29188
Market, duke, PolReID	1483 / 32765
Market, duke, LPW, PolReID	3547 / 481333
CUHK02, CUHK03, Market, Duke	6596 / 84964
CUHK02, CUHK03, Market, Duke, PolReID	6626 / 88541
CUHK02, CUHK03, Market, Duke, PolReID, LPW	8690 / 537109

B. *Re-identification results*

The experimental results are presented in Table III. Samples for training and testing do not overlap.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

TABLE III. EXPERIMENTAL RESULTS

Dataset for test Dataset for train		***Market***		***Duke***		***LPW***		***PolReID***	
		DenseNet	ResNet	DenseNet	ResNet	DenseNet	ResNet	DenseNet	ResNet
Market	Rank1:	89.782	87.708	39.138	31.418	32.132	27.628	65.854	60.975
	Rank5:	96.259	97.952	55.207	48.070	40.691	38.438	65.854	63.415
	Rank10:	97.298	96.675	61.715	54.533	44.294	43.234	65.854	65.854
	mAP:	73.439	70.536	21.079	16.901	19.238	17.734	58.632	57.227
Duke	Rank1:	51.456	44.151	81.688	79.623	28.679	24.775	63.415	65.854
	Rank5:	70.042	62.084	90.260	89.632	36.937	34.535	63.415	68.293
	Rank10:	76.485	69.269	92.774	92.369	40.991	39.339	63.415	68.293
	mAP:	23.536	18.590	64.029	62.001	16.306	13.835	55.038	52.840
LPW	Rank1:	63.005	56.562	41.248	36.894	79.729	71.772	**68.293**	65.854
	Rank5:	79.365	74.822	57.092	51.706	84.535	79.580	**68.293**	68.293
	Rank10:	85.273	81.799	63.600	57.900	86.036	82.733	**68.293**	68.293
	mAP:	34.663	30.413	23.449	18.808	70.069	61.809	58.180	56.584
Market, Duke	Rank1:	**92.132**	89.608	82.406	81.373	44.294	39.940	65.854	68.293
	Rank5:	97.090	95.784	91.023	89.722	55.255	48.949	65.854	68.293
	Rank10:	98.248	97.595	93.312	92.684	59.610	53.303	68.293	68.293
	mAP:	77.698	74.461	67.766	65.772	29.135	26.094	64.384	64.731
Market, Duke, PolReID	Rank1:	91.716	88.955	82.982	79.488	44.294	35.586	**68.293**	65.854
	Rank5:	96.704	95.814	91.472	88.734	55.105	46.847	**68.293**	68.293
	Rank10:	98.070	97.565	93.896	92.011	59.309	52.402	**68.293**	68.293
	mAP:	78.004	71.840	68.065	63.765	29.817	22.709	**65.872**	63.077
Market, Duke, LPW PolReID	Rank1:	**92.132**	88.717	**83.079**	78.591	80.030	75.225	**68.293**	65.854
	Rank5:	96.615	95.814	**91.607**	88.330	84.535	81.081	**68.293**	65.854
	Rank10:	97.951	97.506	**93.537**	91.158	86.687	84.084	**68.293**	65.854
	mAP:	**80.470**	74.176	**69.568**	62.842	73.638	67.154	61.603	61.004
CUHK02, CUHK03, Market, Duke	Rank1:	-	-	-	-	50.751	44.895	**68.293**	68.293
	Rank5:					60.060	54.955	**68.293**	68.293
	Rank10:					65.466	59.910	**68.293**	68.293
	mAP:					38.260	33.153	**66.800**	65.367
CUHK02, CUHK03, Market, Duke, PolReID	Rank1:	-	-	-	-	50.900	42.042	**68.293**	65.854
	Rank5:					60.661	51.051	**68.293**	65.854
	Rank10:					64.865	57.207	**68.293**	65.854
	mAP:					37.370	30.338	64.700	64.350
CUHK02, CUHK03, Market, Duke, PolReID, LPW	Rank1:	-	-	-	-	**83.934**	78.679	**68.293**	68.293
	Rank5:					**87.838**	83.484	**68.293**	68.293
	Rank10:					**89.640**	85.886	**68.293**	68.293
	mAP:					**76.286**	69.815	64.114	61.070

Horizontally, the table can be divided into three parts, each of which three lines consist. The first part includes the testing results the model when it was trained on one of the Market, Duke and LPW datasets. The best value has examples when training and testing were carried out on the same dataset, i.e. training and test samples belong to the same domain. But this result is not objective for invisible domains. The best Rank1 and mAP value corresponds to the experiment when testing invisible datasets, if the training sample was an LPW dataset, and testing was carried out on Market and Rank1 = 63.005, mAP = 34.663. When tested in invisible datasets, the LPW dataset generally showed better training ability compared to Market and Duke, which is most likely due to the significantly larger size of the LPW training set (448568 bounding boxes for 2064 IDs). If we pay attention to the example where LPW acted as a test sample, we can see that the re-identification accuracy is higher when training on the Market than on Duke, which gives us reason to assume that the different identities number is more important than the number of bounding boxes.

The second table part reflects the test results with an increase in the training sample. Here we confirm that an increase in the dataset for training the used CNN leads to an increase in the re-identification accuracy. It was found that the best test results for re-identification can be obtained when the training sample includes data belonging to the same domain as the target one. The best accuracy was achieved when combining all 4 datasets for training during testing: on Market Rank1 did not change, but the mAP increased from 77.698 to 80.470; for duke increase all parameters; for LPW, Rank1 almost doubles, and mAP more than doubles when the dataset is increased from 32765 Bbox for 1483 ID when Market and Duke are combined to 481333 bbox for 3547 ID; for PolReID we can see an increase in Rank1, Rank5, Rank10, but the mAP has become smaller. The last table three rows show the test results when the approach was slightly changed when creating a training sample, i.e. the data from the Market, Duke, CUHK02 and CUHK03 sets were not divided into test and training sets, and this does not allow testing on Market and Duke. Testing on the LPW dataset showed an increase in the Rank1 accuracy to 50.900, in the case when the LPW is an invisible dataset, which is almost one and a half times higher than when using one cross dataset for training. Adding LPW to the training sample allowed us to obtain the maximum values for all estimated metrics, and Rank1 = 83.934, mAP = 76.289.

The maximum Rank1 accuracy achieved for the PolReID dataset is 68.293. The reason for this may be test sample size and composition. Some of the PolReID dataset images were obtained from only one camera, and the re-identification algorithm used cannot detect them. With the further dataset expansion, this will be taken into account.

V. CONCLUSION

Modern person re-identification systems use convolutional neural networks to efficiently extract features. With this approach, the training sample is of great importance. Dataset variety and size allows the re-identification system to have better generalizability and reliability. The built unified database includes 8690 identifiers and 537109 images. Such a large dataset allowed us to improve Rank 1 and / or mAP on all test sets. In further research, we plan to expand the PolReID database we have collected.

REFERENCES

[1] Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S. “Deep Learning for Person Re-identification: A Survey and Outlook”, IEEE transactions on pattern analysis and machine intelligence, 2021.

[2] J. Jieru, Q. Ruan and Timothy M. Hospedales. “Frustratingly Easy Person Re-Identification: Generalizing Person Re-ID in Practice”, BMVC, 2019.

[3] Zhong, Zhun, L. Zheng, Guoliang Kang, Shaozi Li, Y. Yang. “Random Erasing Data Augmentation”, 2020. URL: https://arxiv.org/pdf/1708.04896.pdf.

[4] Xiao, Tong, Hongsheng Li, Wanli Ouyang and Xiaogang Wang. “Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification”, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp.1249-1258, 2016.

[5] Bąk, Sławomir and Peter Carr. “One-Shot Metric Learning for Person Re-identification”, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp.1571-1580, 2017.

[6] Ch. Luo, Ch. Song, Zh. Zhang, “Generalizing Person Re-Identification by Camera-Aware Invariance Learning and Cross-Domain Mixup”, ECVA, 2020.

[7] Hu. Yang, Dong Yi, Shengcai Liao, Zhen Lei and S. Li. “Cross Dataset Person Re-identification”, ACCV Workshops, 2014.

[8] Jin, Xin, Cuiling Lan, Wenjun Zeng, Zhibo Chen and Li Zhang. “Style Normalization and Restitution for Generalizable Person Re-Identification”, IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 3140-3149, 2020.

[9] L. Zheng, L. Shen, L. Tian, Sh. Wang, J. Wang, Q. Tian. “Scalable Person Re-Identification: A Benchmark”, IEEE Int. Conf. on Computer Vision (ICCV), pp. 1116-1124, 2015.

[10] Zheng, Zhedong et al., “Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in Vitro”, IEEE Int. Conf. on Computer Vision (ICCV), pp. 3774-3782, 2017.

[11] W. Li, Xiaogang Wang. “Locally Aligned Feature Transforms across Views”, IEEE Conf. on Computer Vision and Pattern Recognition, pp. 3594-3601, 2013.

[12] W. Li, Rui Zhao, Tong Xiao and Xiaogang Wang. “DeepReID: Deep Filter Pairing Neural Network for Person Re-identification”, IEEE Conf. on Computer Vision and Pattern Recognition, pp. 152-159, 2014.

[13] Song, Guanglu, B. Leng, Y. Liu, Congrui Hetang and Shaofan Cai. “Region-based Quality Estimation Network for Large-scale Person Re-identification”, AAAI, 2018.

[14] PolReID. URL: https://github.com/SvetlanaIgn/PolReID.

[15] Pytorch-YOLOv4. URL: https://github.com/Tianxiaomo/pytorch-YOLOv4.

[16] S. Choi, T. Kim, M. Jeong, H. Park, C. Kim, “Meta Batch-Instance Normalization for Generalizable Person Re-Identification”, IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 3425-3435, 2021.

[17] Chen, Peixian, Pingyang Dai, Jianzhuang Liu, Feng Zheng, Q. Tian and Rongrong Ji. “Dual Distribution Alignment Network for Generalizable Person Re-Identification”, AAAI , 2021.

[18] Person reID baseline pytorch. URL: https://github.com/layumi/Person_reID_baseline_pytorch.

[19] Huang, Gao, Zhuang Liu and Kilian Q. Weinberger. “Densely Connected Convolutional Networks”, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2261-2269, 2017.

[20] He, Kaiming, X. Zhang, Shaoqing Ren and Jian Sun. “Deep Residual Learning for Image Recognition”, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Bag of Deep Features for Classification of Gigapixel Histological Images

Nadia Brancati
Institute for High Performance
Computing and Networking,
National Research Council of Italy
Naples, Italy
nadia.brancati@icar.cnr.it

Crispino Cicala
Computer Scientist, Milan, Italy
cr.cicala@gmail.com

Daniel Riccio
University of Naples Federico II
Naples, Italy
daniel.riccio@unina.it

Maria Frucci
Institute for High Performance
Computing and Networking,
National Research Council of Italy
Naples, Italy
maria.frucci@icar.cnr.it

Abstract. **Convolutional Neural Networks (CNNs) have proven to be one of the most powerful tools for solving complex problems in the field of pattern recognition and image analysis, even if serious challenges remain. Indeed, one of the main drawbacks of CNNs is their inability to cope with very high-resolution images. In areas other than digital pathology, image resizing is often the simplest and most effective solution. However, histopathological images not only show a very high resolution, but also contain a lot of information at the detail level, making this strategy completely ineffective. Other approaches partition the image into small patches and analyze them independently, losing the context information that is fundamental in digital pathology. In this paper, we present a method based on a compressed representation of the Whole Slide Image (WSI), by building a 3D tensor, that preserves the topological and morphological information relating to the proximity relationships between the patches of the WSI. Tensors are used to train a CNN to solve a binary classification task. This technique has been evaluated for the analysis of gigapixel Hematoxylin and Eosin (H&E) histological images with the aim of supporting the diagnosis of breast cancer. Several experiments have been performed on the Camelyon16 dataset by generating different types of 3D tensors. The results of the proposed approach on the breast cancer classification task have been compared with some state-of-the-art approaches.**

Keywords: **histological images, deep learning, clustering**

I. INTRODUCTION

In the field of Computer Aided Diagnosis (CAD), one of the main challenges concerns the analysis of WSIs obtained by scanning tumor tissues stained with H&E and commonly used for the diagnosis of tumor pathologies. Unfortunately, deep learning approaches cannot be applied directly to WSI because of their very high resolution. WSI are generally made up of trillions of pixels that cannot be managed by current deep learning systems. Over the years, different approaches have been proposed trying to meet computational needs but preserving the information needed to perform different tasks of analysis, including the classification of the disease. Most classification approaches are primarily based on partitioning the entire WSI image into patches small enough to be processed independently by a deep network. The class of the entire WSI is usually inferred by combining the decisions obtained for the individual patches [1–3]. Unfortunately, all these approaches neglect the information provided by relationships between patterns presented by individual patches, making the prediction of the CNN an isolated result. Recent methods [4–6] map the WSI into a new compressed and dense feature space by rearranging patch-wise feature vectors in a grid-based representation aiming to preserve spatial correlations of different patches. Although these methods save most of the discriminatory information and the grid representation can be used to train a CNN to classify the entire WSI, the contextual analysis of each point of the grid (i.e the feature vector of a patch) is limited to its 3x3 neighboring in the grid. Indeed, the analysis of relationships between patterns present in the single patches is performed through 3D convolutional operations.

This study proposes a solution for applying CNNs to histopathological images that works on the entire image, but preserving both detail and contextual information by widely extending the contextual analysis of each patch. A set of reference patches is mapped into a high dimensional deep features space, so that bags of deep features words are constructed using a clustering algorithm. A whole slide is partitioned into patches, which are projected into the same high dimensional deep features space and the co-occurrences of the deep features words are considered to build a 3D tensor that represent the entire image in a more compact way.

The experiments were conducted on the Camelyon16 dataset for the binary classification task of breast cancer. Comparisons with the state of the art confirm that the proposed method opens up to the future possibility of further extending this method aiming to further reduce the amount of data to be processed, while still obtaining good results for classification tasks.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

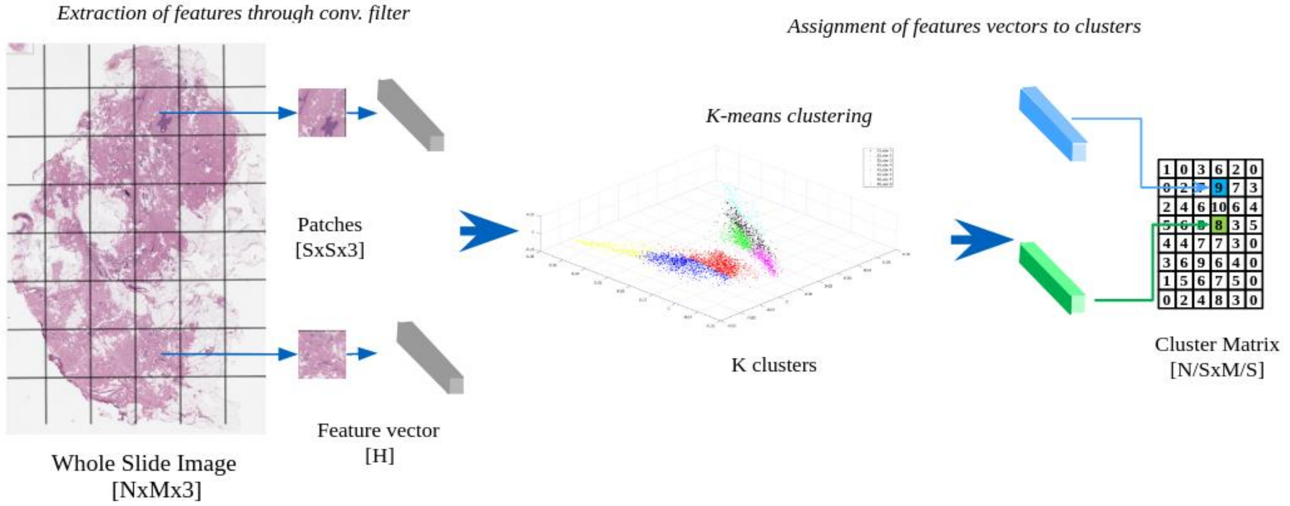


Fig. 1. Matrix cluster feature extraction. A WSI is divided into a set of patches and each of them is mapped to a feature vector using a pre-trained ResNet-18. To each feature vector is assigned a label cluster obtained computing its minimum distance from the feature vectors representing the centroids of the previously computed clusters. The set of label clusters is rearranged in a matrix according to the original spatial arrangement of the patches

II. METHOD

The system takes a WSI of any size as input and its pipeline is as follows. The image is partitioned into non-overlapping patches, which are projected into a high dimensional deep feature space by means of suitably fine-tuned CNN. Our strategy requires pre-processing to divide in clusters the feature vectors of the patches in which any WSI of the training set has been partitioned. The information about the clusters is used to map later each WSI to be analyzed in a dense feature space that is fed to the image-level classifier.

After the preprocessing phase, the overall framework is functionally divided into two main stages, namely Tensor-based Feature Extractor and Tensor-based Classifier.

A. Pre-processing

A pre-trained ResNet-18 [7] is fine-tuned for a binary classification, i.e. in order to distinguish between malignant and benign tissue. The fine-tuning is performed by considering patches extracted from a reference set of WSI that have been manually annotated by pathologists and are provided with the dataset adopted for the experiments. The trained network is then used to extract a feature vector of length *H* for each considered patch, that is then projected into a H-dimensional deep features space. The set of deep features vectors extracted from reference patches undergo a clustering process that is performed by K-means to form a set of K bags of deep features words.

Clusters might include irrelevant or redundant information, so that a post-process is applied aiming to balance data in each cluster and to remove data associated with patches including no tissue. Finally, for each cluster *i*, the corresponding centroid *V* is stored in the *i*-th row of a matrix *MV* of $K \times H$ size. The matrix *MV* will be used to perform the assigned patches extracted from a WSI to the corresponding cluster. In other words, *MV* allows assigning a patch to the corresponding bag of deep features words.

B. Tensor-based Feature Extractor and Classifier

This step is devoted to the generation of a 3D tensor which stores information on the relationships between each couple of different patches lying at distance less or equal to *D* in the WSI input. The distance *D* is the value of the proximity radius determining the contextual area considered for each patch, i.e. *D* is the maximum distance between two different patches of a WSI for which the relationship between the corresponding features can be taken into account.

In the following, two different patches of a WSI at distance less or equal to *D* will be indicated as adjacent patches. Each analyzed patch has size $S \times S \times 3$.

Given an input WSI, namely *W*, with size $N \times M \times 3$, it is partitioned in non overlapping patches of size $S \times S \times 3$. Each patch $p_{i,j}$ is projected in the H-dimensional deep features space by computing its deep feature vector and assigned with the cluster $k \in K$ with the minimum Euclidean distance. Distances are computed between the patch feature vector and the cluster centroids. A cluster matrix *CM* of size $N/S \times M/S$ is constructed, where each (i, j) corresponds to a patch in *W*, and stores the index *k* of the cluster the patch has been assigned with. The index *k* can be also thought of as the cluster label of the patch $p_{i,j}$ (see Fig. 1).

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The cluster matrix CM is then used to build a 3D tensor storing two different types of information about the relationship between feature vectors of adjacent patches of *W*. In more details, let *T* be a tensor, with size $K \times K \times D$, whose elements are initially set to 0. The tensor *T* is dealt as the union of two equivalent prisms P_{dis} *and* P_{cor} each of them of height *D* and with bases formed by orthogonal triangles with legs of length *K*. The prism P_{dis} contains the information related to the distribution of clusters among the adjacent patches of W, while P_{cor} includes information on the correlations between the feature vectors of the adjacent patches of *W*.

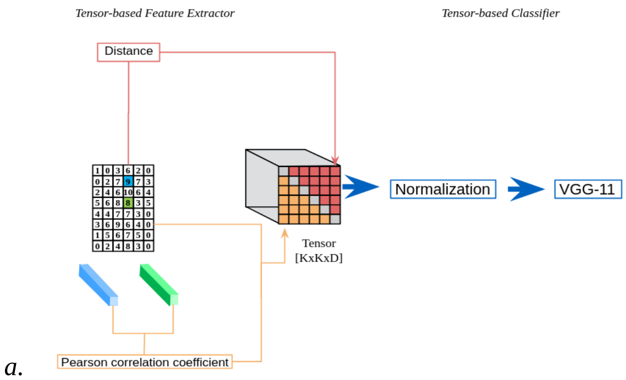


Fig. 2. Our tensor-based network. Two different sets of features are extracted independently on the basis of the information of the cluster matrix and feature vectors of the patches. These sets are stored in two different symmetric volumes of the tensor *T*. For each couple of different patches with a given distance $d \leq D$ and with feature vectors belonging to two defined clusters, the first volume (red part) specifies the occurrence in *W* of the selected pattern, while the second volume of *T* (orange part) includes the sum of correlation indexes between the feature vectors associated to the patches of the selected pattern in *W*. Then, the tensor *T* is normalized and is fed a deep network for the classification tasks

In more details, let be $p_{i,j}$ and $p_{i',j'}$ two patches in W lying at the distance $d \leq D$, which have been assigned to clusters *k* and *k*' respectively, that is $CM(i, j) = k$ and $CM(i', j') = k'$. The set $\{k, k', d\}$ (or also $\{k', k, d\}$) identifies a pattern *SP* in *W* represented by any couple of patches of *W* at distance *d* in *W* and with feature vectors belonging to *k-th* and *k'-th* clusters. The occurrence of *SP* in *W* is stored into the tensor P_{dis}. In particular, the point (m, n, d) of *T*, with $m = min(k, k')$ and $n = max(k, k')$, is incremented by 1 every time the pattern *SP* is detected in *W*. Concomitantly, the sum of correlation indexes between the feature vectors associated to the patches in W characterizing SP is stored in P_{cor}. In particular, for each detected SP in *W*, the point (n, m, d) of *T* is incremented by the value of the Pearson coefficient [8] computed between the feature vectors of the patches belonging to the current *SP*. See Fig. 2.

A normalization process is applied on each slice of *T* to obtain values between -1 and 1. This operation is performed adopting the mapping function [9] that is a quasi sigmoid normalization. Finally, the normalized tensor *T* is fed to a VGG-11 network [10].

III. EXPERIMENTS AND RESULTS

The performance of the proposed method has been evaluated on the publicly available histopathology image Camelyon16 dataset [11].

Different experiments have been performed to select: a) the deep networks both for feature vector extraction and for classification; b) the patches set for the clustering process; c) the normalization function and finally, d) the values of *K* and *D*. Moreover, different strategies have been considered for training of our network, considering either single parts or the whole tensor, aimed at assessing the potential contribution of different kinds of information in T. For the sake of brevity, only some of these experiments will be presented in this paper. Comparisons with recent state-of-the-art techniques are provided on the same task with respect to the same testing protocols.

A. *Dataset and cluster data preparation*

The Camelyon16 dataset contains 400 H&E WSIs of sentinel lymph nodes of breast cancer obtained from two independent sets collected in Radboud University Medical Center (Nijmegen, the Netherlands) and in the University Medical Center Utrecht (Utrecht, the Netherlands). The dataset is originally split into 270 WSIs (160 of normal tissue and 110 containing metastasis) for the training phase and 130 WSIs (80 of normal tissue and 50 containing metastasis) for the test phase; this original splitting was preserved in our experiments. All WSIs of the training set containing metastases are accompanied by manual annotations that have been used for both the training of the ResNet-18 and the selection of patches used for the clustering processes. In particular, 120156 patches have been extracted from the WSI training set to fine-tune the ResNet18 and 15000 patches coming from the WSI test set were used for clustering. The involved patches were appropriately selected from many different images, equally distributing them according to their type, normal or tumor tissue. On the basis of different experiments, the number of clusters *K* was set to 256.

B. *Experimental Setup and Results*

In this study, each analyzed patch has size $S \times S \times 3$, with *S* equal to 224 and ResNet-18 has been adopted as feature extractor for both the clustering process and the generation of the tensor *T*. The extracted features are one-dimensional vectors of length H=512 elements.

We propose three different scenarios for the classification, depending on whether only one part of *T* (P_{dis} or P_{cor}) or the whole tensor *T* is involved in the analysis.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

For each strategy, the results have been evaluated with tensors at different depths, in particular for $D = 4$, 8, 16 and 32.

The performance of different approaches has been compared in terms of standard metrics, namely Accuracy, F-Measure, Specificity and Sensitivity. The performance has been also measured in terms of the Area under the ROC Curve (AUC). The numerical results of these experiments are reported in Table I.

TABLE I. RESULTS

	D	***AUC***	***Acc.***	***F-score***	***Spec.***	***Sens.***
T	4	0,61	0,56	0,56	0,45	0,73
	8	0,72	**0,64**	0,66	**0,51**	0,79
	16	0,61	0,60	0,66	0,48	0,70
	32	0,61	0,62	0,69	0,50	0,70
$\boldsymbol{P_{dis}}$	4	0,51	0,62	**0,77**	–	0,62
	8	0,55	0,62	**0,77**	–	0,62
	16	0,50	0,62	0,73	0,50	0,66
	32	0,49	0,44	0,51	0,32	0,56
$\boldsymbol{P_{cor}}$	4	0,71	0,63	0,65	**0,51**	0,79
	8	**0,75**	0,63	0,65	**0,51**	0,78
	16	0,70	0,58	0,56	0,47	**0,80**
	32	0,72	0,59	0,60	0,47	0,75

In Table I, the best value for each measure is written in bold, while the best result for each type of tensor is written on a gray background. Considering the values for each strategy as a whole, setting $D = 8$ represents the best choice for the maximum distance between two different patches of a WSI for whose relationship between the relative features can be taken into account. The highest values of accuracy (0,68), specificity (0,51) and sensitivity (0,79) are obtained when the whole tensor T is considered. The highest value of F1 score (0,77) is provided by P_{dis} (0,66 for T and 0,65 for P_{cor}). The best performance in terms of AUC (0,75) is obtained considering P_{cor} (0,72 for T and 0,55 for P_{dis}). The remaining measures for P_{cor} show values similar to those obtained for T. Thus, the best strategy can be considered the one based on P_{cor} and for $D = 8$. For this configuration, Fig. 3 shows the confusion matrix and the ROC curve.

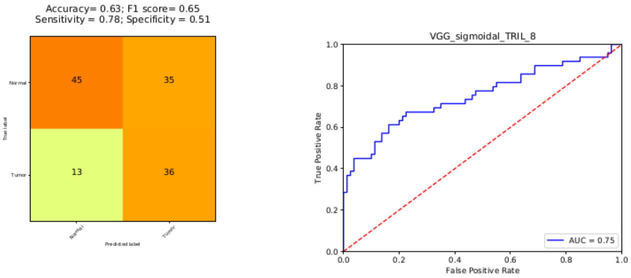


Fig. 3. Confusion matrix and ROC curve of the network, taking into account only P_{cor} with D = 8

The classification result in terms of AUC is comparable with those of the studies in [4] and [5]. The capacity of these methods to reduce the whole-slide images into a compact format was tested in [4] by using three different networks: the Bidirectional Generative Adversarial Network (BiGAN), a Variational AutoEncoder (VAE) and a discriminative model based on contrast training, while the method [5] is based on two Attention networks (AN). The AUC of the BiGAN, VAE and contrastive networks are respectively 0.70, 0.67 and 0.65, while AN provides an AUC equal to 0.71. The results are in line with many of those obtained from the method presented in this study, in which the level of abstraction of whole-slide image representation has increased. However, the result obtained by our model is quite relevant, since for depth levels 4, 8, 16 and 32, it has obtained an AUC of 0.71, 0.75, 0.70 and 0.72 respectively, which is equal to or higher than that obtained from [4] and [5] methods. Thus, the proposed method can represent giga-pixel images in an alternative way, while preserving the ability to discriminate images by classes even at a higher level of abstraction.

IV. CONCLUSIONS

In this paper, a methodology for the analysis of histological images has been proposed which extracts 3D tensors by constructing a grid of clustered deep features and extracting information related to the proximity of the patches. These tensors allow a compact representation of WSIs that can be analyzed by deep learning techniques. Results have shown that a tensor constructed by considering a proximity radius of 8 patches and the correlation measures between the different patches provides the best performance. With this type of image synthesis, the results obtained by the network exceed those obtained by recent studies proposed in the literature, opening up to the future possibility of extending this approach to further reduce the amount of data to be processed, while still obtaining good results in classification tasks.

ACKNOWLEDGMENTS

This work was supported by the project “Campania Oncotherapy - Fighting tumor resistance: an integrated

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

multidisciplinary platform for an innovative technological approach to oncotherapies – Regione Campania”.

REFERENCES

[1] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. Shih, ... and F. González, “High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: Application to invasive breast cancer detection”. PloS one, vol. 13, no. 5, 2018.

[2] K. Das, S. Conjeti, J. Chatterjee, and D. Sheet, “Detection of breast cancer from whole slide histopathological images using deep multiple instance cnn,” IEEE Access, vol. 8, pp. 213 502–213 511, 2020.

[3] Y. S. Vang, Z. Chen, and X. Xie, “Deep learning framework for multi-class breast cancer histology image classification,” in International Conference Image Analysis and Recognition. Springer, 2018, pp. 914–922.

[4] D. Tellez, G. Litjens, J. van der Laak, and F. Ciompi, “Neural image compression for gigapixel histopathology image analysis,” IEEE transactions on pattern analysis and machine intelligence, 2019.

[5] N. Tomita, B. Abdollahi, J. Wei, B. Ren, A. Suriawinata, and S. Hassanpour, “Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides,” JAMA network open, vol. 2, no. 11, pp. e1 914 645–e1 914 645, 2019.

[6] N. Brancati, G. De Pietro, D. Riccio, and M. Frucci “Gigapixel Histopathological Image Analysis using Attention-based Neural Networks”. arXiv preprint arXiv:2101.09992, 2021.

[7] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[8] S. Stigler, "Francis Galton's Account of the Invention of Correlation". Statistical Science, vol. 4, no. 2, pp. 73–79, 1989.

[9] M. De Marsico and D. Riccio “A new data normalization function for multibiometric contexts: A case study”, International Conference Image Analysis and Recognition. Springer. 2008, pp. 1033–1040.

[10] K. Simonyan and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv: 1409.1556, 2014.

[11] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol et al., “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” Jama, vol. 318, no. 22, pp. 2199–2210, 2017.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Chest Pathologies Intellectual Analysis

Viktor Krasnoproshin
Faculty of Applied Mathematics
and Computer Science
Belarussian State University
Minsk, Belarus
krasnoproshin@bsu.by

Andrew Naumovich
Faculty of Applied Mathematics
and Computer Science
Belarussian State University
Minsk, Belarus
andrew.naumovich@yandex.ru

***Abstract.* The paper addresses chest pathologies analysis problem. An algorithmic approach based on neural networks is proposed. An algorithm is implemented as a special software package.**

***Keywords:* neural networks, X-Ray image, chest pathologies, medical screening, supervised learning**

I. INTRODUCTION

In the conditions of a pandemic, the primary task of medicine is to detect the disease at an early stage. The solution of this problem is carried out by mass screening of the population. Doctors usually use all kinds of (stationary and portable) equipment to obtain primary data about patients. However, the use of various types of medical devices makes analysis of the images obtained quite difficult. To increase radiologists' performance, in this case, a flexible software tool is necessary. It should be as independent as possible from the types of technical devices. The data processing and diagnosis making a diagnosis can be simplified to a greater extent by using an appropriate set of algorithmic tools that automate all its stages.

The report discusses one of the possible approaches to solving the problem based on intelligent image analysis [1, 2].

II. PROBLEM ANALYSIS AND PROBLEM STATEMENT

Before proceeding to the solution of the problem, it is necessary to identify the main stages of obtaining and analyzing [3–5]. X-ray images and the problems associated with them.

The overall process of obtaining and analyzing radiographic images can be presented in the form of a sequence of the following main stages:

- Collecting data from various radiographic devices,
- Image preprocessing,
- Intellectual data analysis,
- Predictions calibration,
- The stage of diagnosis.

It is evident that the solution of a specific application problem is associated with each of these stages. In addition to that, it is necessary to take into account the specifics of the corresponding problem when solving them. Thus, at the data collection stage, it is necessary to support the transfer of images from various types of medical devices. At the image preprocessing stage, images of various devices must be brought to the common view. At predictions' calibration stage the specifics of the specific medical equipment should be taken into account. Finally, when making a final diagnosis it is necessary to diagnose not only the main, but also the accompanying diseases.

III. THE PROCESS OF SOLVING THE PROBLEM

Let's first consider the stage of preprocessing a set of images X.

Let some transformation $g: X \rightarrow X^{'}$ be given for performing image preprocessing. The resulting image $X^{'}$ has the following limitations: $x^{'} \in$ $\in [0 \dots 255]^{w \times h \times c}, \sim \forall x^{'} \in X^{'}$, where w and h are width and height of the image and c is the number of channels.

That is, the initial X-ray images are reduced to a common format with the same dimension, which simplifies their subsequent processing.

Let's consider the next stage of image analysis. Let $D = \{d_1, d_2, \dots, d_k\}$ be the set of possible pathologies. Notation $d_i \in x$ is used to show that pathology d_i is present at the image x. Since it is necessary to detect all diseases present in the image, then for each input image $x \in X^{'}$ a k -dimensional vector $y =$ $= (y_i | y_i = 1(d_i \in x))$ is assigned, where $1(x)$ is an indicator function.

Let $Y = \{y\}$ be the set of all possible pathologies. Thus, at the stage of image analysis, it is necessary to construct a transformation $C: X^{'} \rightarrow Y$.

To solve this problem, a convolutional neural networks approach [6–8] is proposed. Inputs for the neural network are images $x \in X^{'}$ and outputs are probability vectors $p(x) = p(p_i | p_i = P(d_i \in x))$.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

At the stage of converting the vector p into y, it is necessary to build an algorithm $I: P \rightarrow Y$. It is intended for results calibration and interpretation. At the same time, it is necessary to take into account the specifics of the X-ray equipment (the manufacturer, the shooting conditions and other factors). A simple threshold selection algorithm is proposed.

Let's assume that a vector of threshold values $T = (t_1, t_2, \ldots, t_k), \sim t_i \in (0,1)$ is given for each specific device. Interpretation transform is defined as following: $I(p) = (\hat{y}_i | \hat{y}_i = 1(p_i > t_i))$.

Thus, the process of chest pathologies analysis is described in general terms. The most interesting stage in this case is the image analysis stage, since the performance of the entire intellectual framework depends on the quality of its solution.

IV. DENSENET ARCHITECTURE

A well-known DenseNet architecture [9, 10] is chosen for image analysis. An image x of dimensionality $384 \times 384 \times 3$ is fed into the network. As a result of the transformation performed by the neural network, the resulting feature description of the image $f(x)$has dimensionality of $12 \times 12 \times 2048$.

With the help of linear layers composition feature description is projected into the vector $L = (l_1, l_2, \ldots, l_k | l_i \in \mathcal{R})$. Number of components k in that vector matches the total number of lung pathologies under analysis. To interpret the result in terms of probability, as required by the process, it is necessary to apply a sigmoid function to the vector L component-wise: $p(x) = (\sigma(l_1), \sigma(l_2), \ldots, \sigma(l_k))$, where $\sigma(x) = \frac{1}{1+e^{-x}}$. DenseNet architecture is presented in Fig. 1.

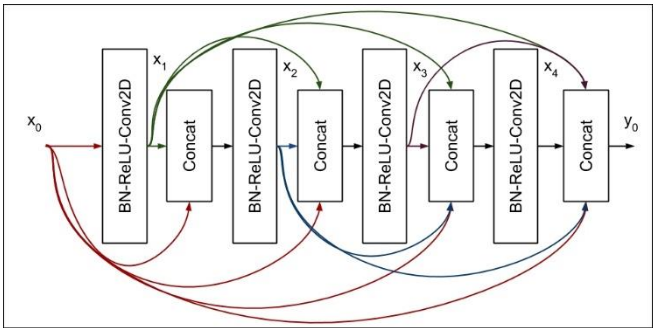


Fig. 1. DenseNet architecture

As a result of the intelligent image analysis, besides pathologies recognition, it is also vital to interpret the outputs of the neural network. That interpretation will be used later by the radiologist to validate or correct the diagnosis made by the algorithm.

To interpret the prediction, it is proposed to use a heat map constructed by the CAM algorithm.

Let $f(x)_{ij}^k$ be a pixel of k-th feature map of the resulting image vector description created by the neural network. The operation Global Average Pooling can be defined as follows: $F^k(x) = \frac{1}{144}\sum_{i=1}^{12}\sum_{j=1}^{12} f(x)_{ij}^k$. The heat map is defined by the equation $Y^c = \sum_{i=1}^{k} w_i^c F^i$, where c is the sequence number of the pathology under consideration, w_i^c is the weight of the connection between k −th feature map and c-th pathology inside the dense classification layer.

V. EXPERIMENTS

To test the operability of the algorithmic solution, experiments were conducted using images from the NIH [11] and RSNA radiographic image databases.

The RSNA database contains X-ray images with annotations. The total amount of data is about 30,000 images. There are 3 classes available for recognition: *Normal*, *Lung Opacity*, *Not Normal/No Opacity*. To simplify the experiment, only the first 2 classes were used. In addition to the class labels, the annotation of the images contains the coordinates of the areas in the image that the experts considered important when making the final diagnosis. Such data are not used in this study. The results of evaluation on the test set are presented in Table I.

TABLE I. RSNA EVALUATION RESULTS

	Precision	Recall	F1
Normal	0.947	0.971	0.959
Pneumonia	0.956	0.919	0.937

The second data set (NIH) consists of 112 thousand images. 14 different lung pathologies are available for classification. A distinctive feature of the data set is that pathologies can appear simultaneously. The data set is very diverse, it contains images that differ in brightness and contrast. Some images have labels that indicate the orientation of the X-ray image.

The NIH database is not balanced by the number of classes which significantly complicates the process of training a neural network. To overcome this complexity, the following approach is proposed.

Each training sample x is assigned a vector w which is used to scale loss function for that training sample. For each unique pathology label $c \in T$ two weight coefficients w_c^+ and w_c^- are calculated. The first one is used for positive images relatively to pathology c, the second one - for negatives.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

$$w_c^+ = \frac{\alpha_c(N_c - N_c^+)}{\alpha_c(N_c - N_c^+) + N_c^+},$$

$$w_c^- = \frac{N_c^+}{\alpha_c(N_c - N_c^+) + N_c^+}.$$

In that equation N is the total number of training samples, N_c^+ - the number of samples with pathology c, α_c is a hyperparameter. The resulting weight w is calculated as $w = (w_i | i \in T), w_i = 1(i \in x) * w_i^+ + 1(i \notin x) * w_i^-$.

To increase the variety of images and combat the retraining of the neural network, the augmentation of input data is conducted. The following operations are considered: *Center Crop*, *Random Horizontal Flip*, *Random rotation*.

VI. PRACTICAL IMPLEMENTATION AND THE EXAMPLES OF OPERATION

A software package has been developed based on the algorithmic solution of the problem. This complex automates the process of diagnosis, which greatly simplifies the work of radiologists.

The software package consists of the following modules:

- Algorithmic module,
- A database with verified images and diagnoses to them,
- Module for visualization and interpretation of results,
- Calibration module,
- User and Expert Interfaces.

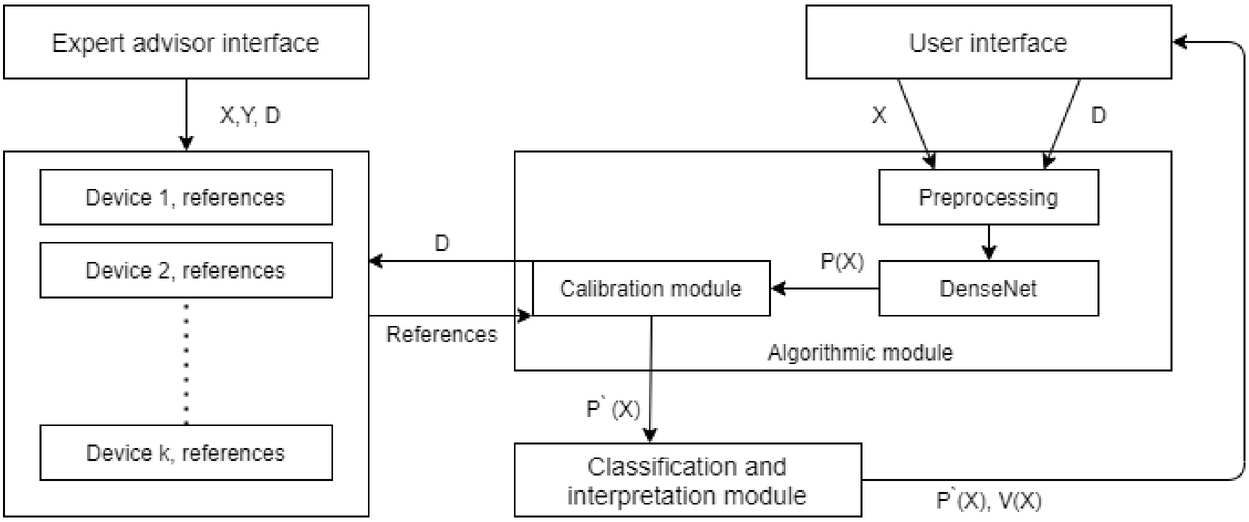


Fig. 2. Software package architecture

The results of the evaluation are presented in Table II. They confirm the efficiency of the proposed algorithm for its use in practice.

TABLE II. NIH EVALUATION RESULTS

	AUC-ROC	**F1**	**F1(calibrated)**	**Threshold**
No Finding	0.792	0.740	0.770	0.001
Atelectasis	0.816	0.397	0.420	0.7994
Consolidation	0.797	0.207	0.219	0.9790
Infiltration	0.708	0.387	0.426	0.001
Pneumothorax	0.885	0.383	0.422	0.959
Edema	0.883	0.237	0.237	0.9590
Emphysema	0.896	0.287	0.347	0.979
Fibrosis	0.799	0.092	0.110	0.979
Effusion	0.883	0.556	0.555	0.599
Pneumonia	0.791	0.068	0.096	0.979
Cardiomegaly	0.891	0.301	0.351	0.979

The architecture of the software package is presented in Fig. 2.

VII. CONCLUSION

The article addresses the problem of chest pathologies intellectual analysis. The process of the analysis is carried out using X-ray images. That images are obtained from different types of technical devices. An approach to solving the problem is proposed, including preprocessing of source images, interllectual analysis, calibration of results and diagnosis. Experiments have been conducted to confirm the efficiency of this approach. Based on the results of theoretical research, a software technology has been developed that provides an effective solution to the problem.

REFERENCES

[1] V. V. Krasnoproshin, H. E. R. M. Vissia, Decision-Making and Big Data, Las Nuevas Areas del Poder Economico Mundial. XII Acto Internacional de la Real Academia de Ciencias Economicas y Financieras, Barcelona, Royal Academy of Economy and Financial Sciences, 2017, pp. 105–120.

[2] V. Krasnoproshin, V. Obraztsov, H. Vissia, Solution of applied problems: formalization, methodology and justification, World Scientific Proceeding Series on Computer Engineering and Information Science, vol. 3. “Computational Intelligence in Business and Economics”, London, World Scientific, 2010, pp. 57–64.

[3] A. Naumovich, Program logs intellectual analysis, BEST YOUNG SCIENTIST – 2021: III International Book Edition of the countries of the Commonwealth of Independent States, Nur-Sultan, 2021, pp. 119–122.

[4] A. Naumovich, Neural Network Based log anomaly detection, Proceedings of the 63rd All-Russian Scientific Conference of MIPT, Section: applied mathematics and computer science, Moscow, MIPT, 2020, pp. 309–311.

[5] Rahaman, Md Mamunur et al., ‘Identification of COVID-19 Samples from Chest X-Ray Images Using Deep Learning: A Comparison of Transfer Learning Approaches’, 1 Jan. 2020, pp. 821–839.

[6] Pranav Rajpurkar et al., CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning [Electronic resource]. Available at https://arxiv.org/pdf/1711.05225.pdf (accessed 2021 March).

[7] V. Golenkov, N. Guliakina, V. Golovko, V. Krasnoproshin Artificial Intelligence Standardization Is a Key Challenge for the Technologies of the Future, Communications in Computer and Information Science, vol. 1282: “Open Semantic Technologies for Intelligent System”, Springer, 2020, pp. 1–21.

[8] V. A. Golovko, V. V. Krasnoproshin. Neural network technologies of data processing: a textbook, Minsk, BSU, 2017, 264 p.

[9] Gao Huang et al., Densely connected Convolutional Networks [Electronic resource]. Available at https://arxiv.org/abs/1608.06993 (accessed 2021 April).

[10] Hongyu Wang, Yong Xia. ChestNet: A Deep Neural Network for Classification of Thoracic Diseases on Chest Radiography [Electronic resource]. Available at https://arxiv.org/abs/1807.03058 (accessed 2021 May).

[11] Xiaosong Wang et al., ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-supervised Classification and Localization of Common Thorax Diseases, [Electronic resource]. Available at http://dx.doi.org/10.1109/CVPR.2017.369 (accessed 2021 April).

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

AI-based Retrospective Study for Revealing Diagnostic Errors in Chest X-ray Screening

Vitali Liauchuk
Biomedical Image Analysis Dept.
United Institute of Informatics
Problems of NAS of Belarus
Minsk, Belarus
vitali.liauchuk@gmail.com

Aleh Tarasau
Republican Research and
Practical Center for
Pulmonology and Tuberculosis
Minsk, Belarus
novoe1975@gmail.com

Vassili Kovalev
Biomedical Image Analysis Dept.
United Institute of Informatics
Problems of NAS of Belarus
Minsk, Belarus
vassili.kovalev@gmail.com

***Abstract.* In this paper, we explore the ability of an AI-based computer-aided diagnostic system (CAD) to help to reveal the early signs of probable lung diseases in X-ray images. We use a large screening database which contains natively-digital X-ray images acquired between 2001 and 2014 along with the corresponding diagnostic reports provided by the radiologists. We apply a Deep Learning-based CAD system to the cases from the database which were labeled by the radiologist as a norm and compare the CAD prediction results to the radiologists' diagnostic reports. Our experiments demonstrate the ability of an automated AI-based CAD to reveal discrepancies between the diagnostic reports and the actual state of lungs as conveyed by the X-Ray image. Additionally, in a number of cases the Deep Learning algorithm was able to detect early signs of lung diseases which progressed later according to the patient anamnesis.**

***Keywords:* X-ray CAD, AI, Deep Learning, Retrospective**

I. INTRODUCTION

With the recent emergence of Big Data and Deep Learning methods we observe a rapid development of algorithms which are often referred as Alternative Intelligence (AI). State-of-the-art AI-based solutions of different kinds find more and more applications in different business areas including marketing, industry, and modern software.

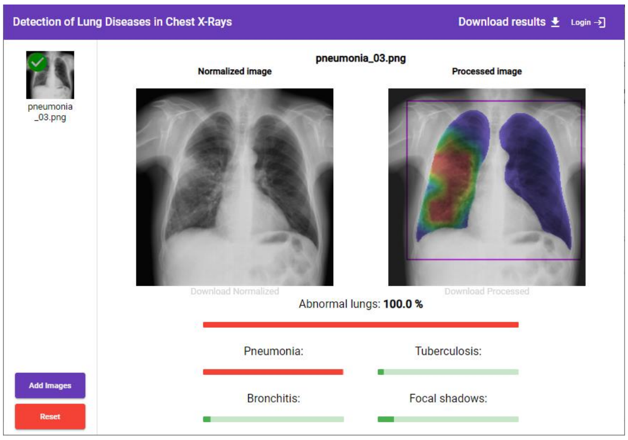
However, in the field of medicine the process of incorporation of AI-based solutions for diagnosis and treatment is not as rapid as in other fields. The use of AI in medicine is still rather limited due to high responsibility in the decision making, strict protocols and some skepticism with respect to the use of computerized methods [1]. The exact ways of using AI algorithms in medicine are under discussion [2].

In this paper, we assess the potential profit of using an AI-based X-ray CAD system in clinical practice during the screening. One way to do so is to start using an X-ray CAD system in daily clinical routine. After a while, the effect of using the AI algorithms during the preliminary diagnostic process can be quantitatively evaluated in terms of increased sensitivity, reduced time spent for diagnosing, etc. However, such prospective study requires significant efforts and takes a substantial amount of time to gather representative statistics. This is especially true in case of population screening scenarios where most of the examined cases are expected to be normal. Therefore, here we consider a retrospective study which uses the X-ray images and the corresponding diagnostic reports made in the past.

In this work, we present the results of a retrospective study with the use of a large archive X-ray screening database and a domestic X-ray CAD system. Here we address the following two questions.

1) Is a CAD system able to alarm potential miss-classifications of X-ray images?

2) Is a CAD system able to help in detection of early signs of lung diseases?



Abnormal lungs: **100.0 %**

Pneumonia:

Tuberculosis:

Bronchitis:

Focal shadows:

Fig. 1. Visual appearance of the X-ray CAD web-service (https://image.org.by/xray_lung)

To this end, we first automatically detect the cases with potential discrepancy between the radiological description and the X-ray image content. Then, the most probable candidates are manually analyzed by a

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

qualified radiologist to approve or disapprove the revealed discrepancy. Additionally, for those patients which had multiple records in the database we explore whether the automatic CAD could detect the lung abnormalities earlier than the radiologist did according to the anamnesis.

The web-version of the X-ray CAD system used with this study is available for testing at [3]. Visual appearance of the web-service is shown in Fig. 1.

II. X-RAY CAD SYSTEM EMPLOYED

A. *CNN Training data*

The majority of the X-ray images used in this study were originally digital images taken from the population screening data storage system. The X-ray images were presented as single-channel 16-bit DICOM images with resolution varying from 520×576 to 2800×2531 pixels. Each X-ray scan from the database had its corresponding textual description provided by the radiologist during the image assessment process. In order to extract the image class labels ("normal", "tuberculosis", "pneumonia", etc.) the textual descriptions were parsed with use of keyword and keyphrase matching. The textual descriptions were assigned by qualified radiologists within the screening procedure.

A relatively small portion of the database images was used to train the Convolutional Neural Network (CNN) model for X-ray image classification. The entire database contained 1,908,926 records. From the screening database, a total number of 33,089 cases were selected to compose the study group. Out of those, 16,594 cases represented healthy subjects ("normal"), and 16,495 represented X-ray cases with visible signs of at least one lung disease including tuberculosis, pneumonia, focal shadows and bronchitis ("abnormal"). The screening-based study group was randomly split into training, validation and testing subsets with ratio 70%, 20% and 10% respectively. The split was performed so that all images that were known to belong to a single patient appeared all together in one subset.

To increase the robustness of the resultant prediction model, data from the 3rd party datasets was added into the training subset. The 3rd party datasets included Montgomery and Shenzhen datasets [4], and an additional subset of normal X-ray images taken from "Normal CXR Module: Train Your Eye" [5]. In total, the training subset was extended by 657 normal and 295 abnormal images from 3rd party sources.

B. *CNN model training*

The CNN model used for classification of X-ray images and localization of abnormal regions was composed of a convolutional part of VGG16 network as backbone appended by several additional layers including a special Heatmap layer. The details about the CNN architecture employed are described in [6]. During inference, the network outputs the overall abnormality confidence score, partial confidence scores for presence of signs of tuberculosis, pneumonia, focal shadows and bronchitis, and a heatmap indicating the localization of the abnormal and suspicious regions in the target image (see Fig. 1).

The CNN training process included two stages. At the first stage, the CNN backbone was initialized with ImageNet-trained weights and the training was performed for the binary classification task ("normal" vs. "abnormal"). This allows using all the available data including 3rd party images which lack meta-information on the specific diseases (tuberculosis, pneumonia, etc.). The second stage is continued from the checkpoint from the first training stage which minimizes the validation loss value. The second stage considers training the CNN in a multi-class multi-label mode with use of all the data except for abnormal 3rd party cases. The experiments showed that using data from different sources makes the final trained model better usable for 3rd party data.

The classification performance of the trained model is assessed with use of ROC-curves. Results of the performance assessment evaluated on the testing subset are shown in Fig. 2.

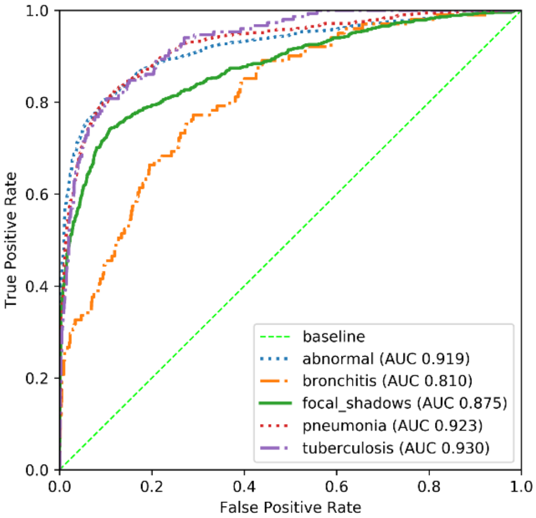


Fig. 2. Roc-curves for prediction of X-ray image with use of the domestic CAD system evaluated on the testing subset

III. RETROSPECTIVE DATABASE ANALYSIS

A. *Automatic evaluation of normal X-rays*

A total number of 563,495 screening database records were recognized as normal according to their textual descriptions. All the selected images were

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

evaluated with use of the domestic CAD system. The overall abnormality score was used for the subsequent triage of X-ray cases. The statistical results of the evaluation of cases which were marked as normal in the database are shown in Fig. 3. As it can be seen from the histogram, the vast majority of cases had the abnormality score below 0.5, which is well expected for such selection of cases. Still, a large number of images had big abnormality score values. Specifically, 10,007 cases had scores above 0.9, out of which 17 had the maximum possible score of 1.0.

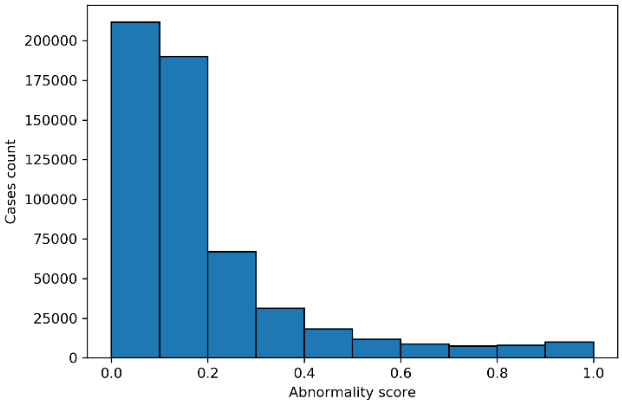


Fig. 3. Abnormality scores automatically evaluated on the screening X-ray images which were labeled as normal

B. Manual evaluation

Due to high costs of manual analysis, only the top-500 X-ray images with the highest abnormality scores were selected for further analysis. Thus, in this preliminary study only a small portion of the potentially interesting data was evaluated. The subsequent analysis included two stages. At the first stage, the automatically selected cases were filtered by a non-radiologist with use of visual analysis to exclude the obvious false-positives (wrong image orientation, unsuitable projection, scanning failures, artifacts, CNN reaction to nipple shadows, etc.). As a result of this stage, 124 out of 500 X-ray cases were selected for the subsequent visual analysis by a qualified radiologist. At the second stage of visual analysis the radiologist had access to all X-ray images of the selected patients available in the screening database. For each image, the corresponding textual description was available. For each analyzed image, the radiologist was to answer two questions: (1) “Is there a discrepancy between the X-ray image content and its description in the database?”, and (2) “Was the X-ray CAD helpful in revealing such discrepancy?”. An additional task was to find those cases in which the automatic CAD could reveal signs of a lung disease earlier than it was done by the radiologists according to the anamnesis.

C. Results

The visual analysis procedure described above revealed 54 cases which had discrepancy between the X-ray image content and its description in the database. In all these cases the CAD-provided heatmaps were helpful in finding the suspicious findings in the X-ray images. In 5 cases the CNN revealed findings which were confirmed later in the anamnesis, in some cases the findings developed into a lung disease.

Fig. 4 shows the examples of X-ray images which have visual signs of abnormalities but were marked as normal in the screening database. Here, the left column shows the original X-ray images, and the right column depicts the corresponding images analyzed with use of the developed X-ray CAD with lung mask and heatmap overlayed on top of the original image.

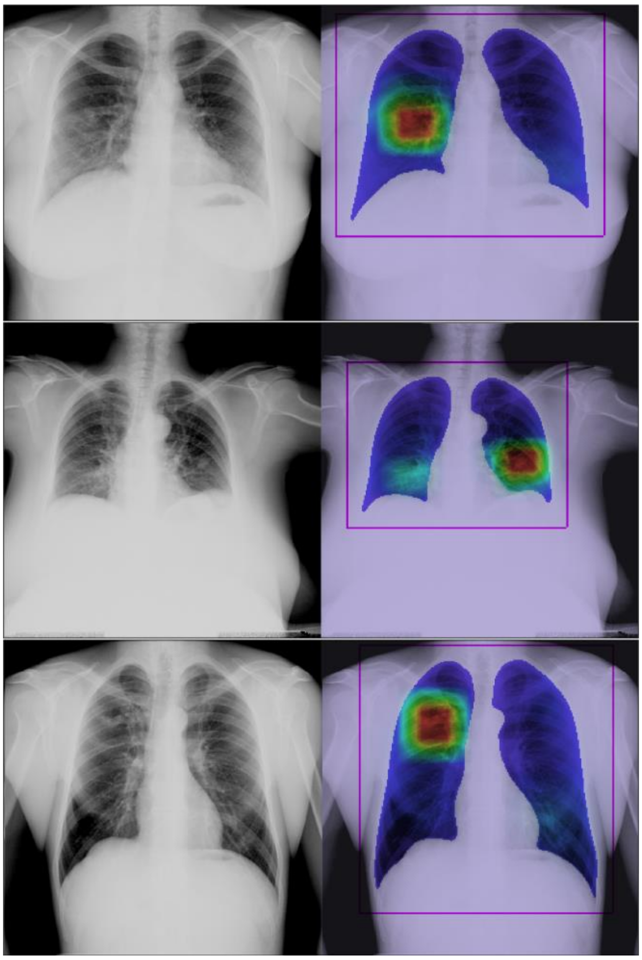


Fig. 4. Examples of X-ray images and the corresponding AI processing results which have visual signs of abnormalities but were marked as normal in the screening database

Fig. 5 shows a case which demonstrates the potential use of the developed X-ray CAD for early detection of signs of disease. Here, the top row shows an X-ray scan of a patient from October 2007, the

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

textual description reports no visual signs of abnormalities. The bottom row shows another scan of the same patient from October 2008, the description reports a "medium-intensity linear shadow in lower lung field on the right side". On the other hand, the X-ray CAD applied to these images highlighted the suspicious region with shadow in both cases.

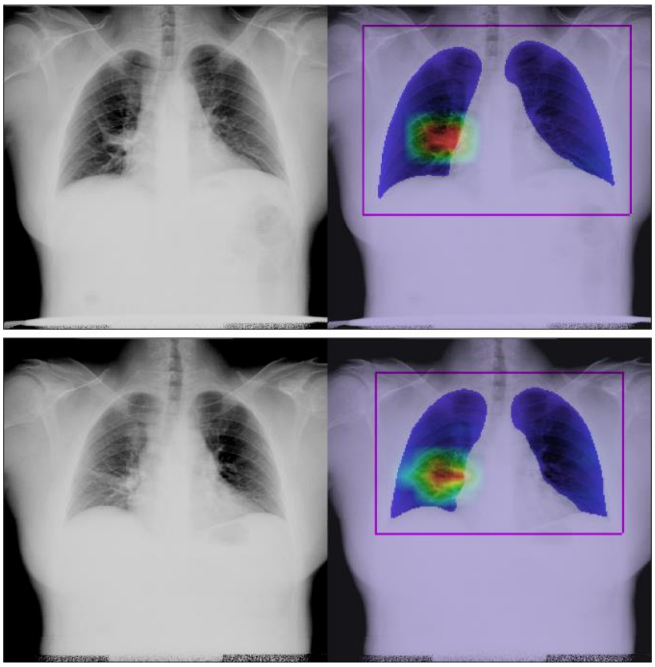


Fig. 5. Two subsequent X-ray scans of a single patient with ~1 year difference; the first image was diagnosed as normal, whilst the second one is reported to have an abnormality in the right lung; automatic X-ray CAD correctly localized the abnormality in both cases

IV. CONCLUSIONS

The experimental results presented above suggest drawing the following conclusions.

1) The X-ray CAD system employed with this study is capable of alarming the potential miss-classifications of X-ray images.

2) The employed X-ray CAD system is helpful in detection of early signs of lung diseases.

It should be noticed that in this preliminary study only a small portion of the potentially misclassified X-ray images was visually verified. Out of only 500 cases with the highest abnormality confidence scores we found 54 images which were indeed annotated with an error. This is roughly 10% of the suspicious cases examined. Presumably, a large-scale analysis of tens of thousands of suspicious X-ray images could reveal hundreds or even thousands of misclassified cases.

In general, the results of this retrospective study suggest that using a high-precision X-ray CAD system in routine population screening can increase the diagnostic sensitivity and, in some cases, detect early signs of lung diseases.

ACKNOWLEDGMENT

This study was partly supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, U.S. Department of Health and Human Services, USA through the CRDF project DAA9-19-65987-1 "Year 8: Belarus TB Database and TB Portal".

REFERENCES

[1] V. H. Buch, I. Ahmed, and M. Maruthappu, "Artificial intelligence in medicine: current trends and future possibilities", The British journal of general practice : the journal of the Royal College of General Practitioners, Vol. 68, pp. 143–144. March 2018.

[2] A. S. Ahuja, "The impact of artificial intelligence in medicine on the future role of the physician.", PeerJ, Vol. 7, pp. e7702, October 2019.

[3] "Detection of Lung Diseases in Chest X-Rays": https://image.org.by/xray_lung – last visited August 7th, 2021.

[4] S. Jaeger, S. Candemir, S. Antani, Y. X. Wáng, P. X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases", Quantitative imaging in medicine and surgery, Vol. 4, pp. 475–477, December 2014.

[5] "Chest X-ray.com": http://www.chestx-ray.com/ - last visited August 7th, 2021.

[6] V. Kovalev, V. Liauchuk, D. Voynov, and A. Tuzikov, "Biomedical Image Recognition in Pulmonology and Oncology with the Use of Deep Learning", Pattern Recognition and Image Analysis, Vol. 31, no. 1, pp. 133–151, April 2021.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Analysis of Motion of Dynamic Scenes in Microscopy Images: Formalization, Criteria and Results

Olga Nedzved
Belarusian State University
Minsk, Belarus
onedzved@bsu.by

Vera Yashina
Federal Research Center
“Computer Science and
Control” of RAS
Moscow, Russia
werayashina@gmail.com

Sergey Ablameyko
Belarusian State University
Minsk, Belarus
ablameyko@yandex.by

Igor Gurevich
Federal Research Center
“Computer Science and
Control” of RAS
Moscow, Russia
igourevi@ccas.ru

Ren Tiaojuan
Zhejiang Shuren University
Hangzhou, China
rentj@zjsru.edu.cn

Ye Fangfang
Zhejiang Shuren University
Hangzhou, China
cliney@zju.edu.cn

Abstract. **In this paper, we formalize the problem of the motion analysis of dynamic objects and scenes based on the algorithms and methods developed by the authors for analysis of the cell population behavior. Cell population is considered as a system of dynamic objects and motion is analyzed by using concept of an integral optical flow. On the base of the main types of motion, the key points of cell movement in the population are identified and stages of cell development and interaction are described. The formalization of operations on dynamic objects has been completed.**

Keywords: **dynamic object, scene, motion**

I. INTRODUCTION

In the study of living cells, time-lapse microscopy is important. This is a sequential recording of microscopic images in long-term monitoring systems for the observation and analysis of the cell population in vitro, which allows to study the cell dynamics in detail.

Video-microscopy can be considered as one of the types of time-lapse microscopy. Its advantages include high temporal resolution and the ability to shoot continuously over long periods of time. Video-microscopy allows you to obtain frame-by-frame recording of changes in the shape and mobility of living cells, as well as the brightness of their images. Cell motion can be described based on dynamic objects [1].

The currently existing technologies for video sequences analysis are focused mainly on the motion of individual objects rather their moving aggregates, which combine the motion of the entire system with the motion of its components [2].

One of the ways to track the motion of cells is the method of tracking dynamic objects, which is a continuous determination of the position of the object [3, 4]. In this paper, we analyze the motion of dynamic objects and scenes, formalize this process and demonstrate the results.

II. DYNAMIC OBJECTS AND SCENES

An elementary dynamic object is a small movable localized object with physical parameters such as volume, density or mass. Their motion can be rotational, rectilinear, accelerated, or even barely noticeable. The complexity of detection and tracking is determined by their size, change in shape, and the nature of motion.

Dynamic objects in microscopic images can be divided into the following classes:

– individual small objects that are a movable component of the background and can be removed when sorting by size,

– large background components with small displacement due to stochastic motions in a specimen,

– fragments of the environment around moving objects that are changed due to their optical characteristics.

Cells in microscopic specimens are mobile objects, three types of cell motility can be distinguished:

– real cell motion,

– displacement of intracellular structures,

– changes in cell shape.

Dynamic objects form scenes. The scene can be defined as static, with the added time variable. 3D motion is estimated by modeling forward and backward flow of objects as dense three-dimensional vector fields.

A continuous scene can be represented as a 5D vector function, where the input determines a three-

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

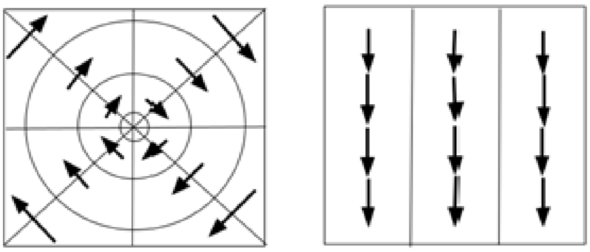
dimensional location $x = (x, y, z)$ and the viewing direction 2D (θ, φ). The output corresponds to the color of the pixels and can be defined as $c = (r, g, b)$ and volume density σ. In practice, the direction is expressed as a three-dimensional vector in Cartesian space d. This continuous 5D representation of the scene can be approximated by a network MLP $F\,\Theta : (x, d) \rightarrow (c, \sigma)$, where Θ means the weight to match each 5D input coordinate with the corresponding volume density and corresponding color. Dynamic objects can be described in the same way.

The key problem of monitoring a dynamic scene is to separate the background and objects that can be static or change over time. In this case, the direction of motion plays an important role, it can be divided into the following levels:

- background motion generated by camera motion;
- background motion formed by a change in the surrounding space;
- object motion;
- motion inside objects;
- motion of groups of objects.

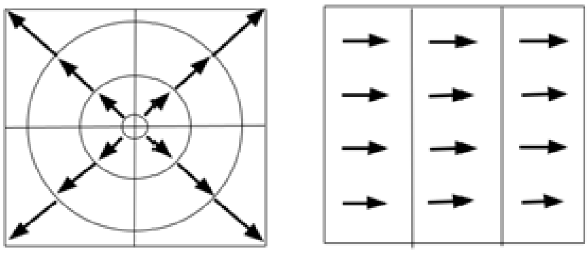
Thus, in the first step, it is necessary to divide the image field into different types of motion. For this purpose, motion maps based on optical flow calculations can be used [5].

Obviously, for a uniform motion of the optical system, the background is formed in the form of a constant flow, which has a unique image. Depending on the motion of the camera, the image has its own unique characteristics as in Fig. 1.



Camera rotation

Downward movement of the camera



Forward movement of the camera

Movement of the camera to the right

Fig. 1. Direction of the optical flow field when the camera moves

The behavior of system of objects is determined by group motion. In this case, motion maps are used, they determine general trends and can indicate individual events in motion [6]. Motion mapping based on the summation of vectors in a local area of the image. But simple summation can lead to the same values for different motion patterns, as in Fig. 2. Thus, the result of the movement must be determined using several different maps.

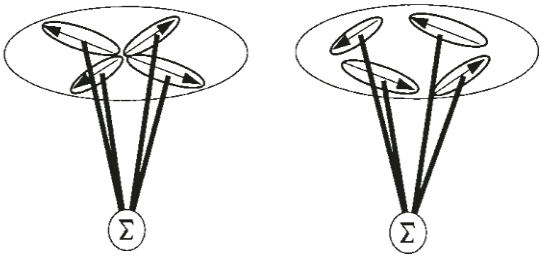


Fig. 2. Two cases when the adder gives the same result for different patterns of motion

It is assumed that these directions of movement are spatially integrated by local cells in the image space in the same way as provided in the template model in Fig. 3.

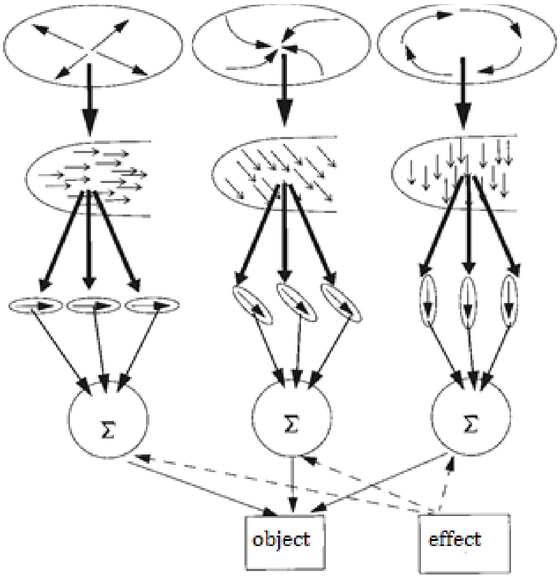


Fig. 3. Scheme of the model of building motion maps. The optical flow field is converted to a vector map. The model cells form the setting of the Gaussian direction of movement in this coordinate system. Adders determine the directions of movement based on the direction of the cells

The fields in the flow analysis model combine vectors that encode similar directions of motion in the locally generalized space. The model assumes the selectivity of traffic flows, which is based on the properties of the local field, rather than on complex and specialized interactions. Model local cells calculate the main directions of optical flow. In particular, at each position the cell has a generalized field corresponding

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

to the main direction of movement. This field makes each cell less sensitive to deviations from its main direction of motion. The local field model summarizes data with the same preferred direction of movement in a spatial area around their center, as in Fig. 3. In addition, the model forms the probability of the direction of motion. Thus, based on the summation of vectors, it is possible to generate motion maps that allow to determine the contours of a dynamic object and the types of its motion.

III. MONITORING OF DYNAMIC OBJECT MOTION

The general scheme for analyzing images of dynamic objects includes five stages (Fig. 4):

– image capture and preprocessing;

– segmentation of the image scene and selection of cells areas;

– characteristics measurement;

– definition of laws for the formation of a general description of cell motion;

– cell state classification.

The most important characteristics for describing dynamic objects in microscopic images are: position characteristics (coordinates, speed of motion, direction change, trajectory), morphological characteristics (size, shape, degree of object shape change), optical characteristics (brightness, color, optical density), and also the duration of the observation.

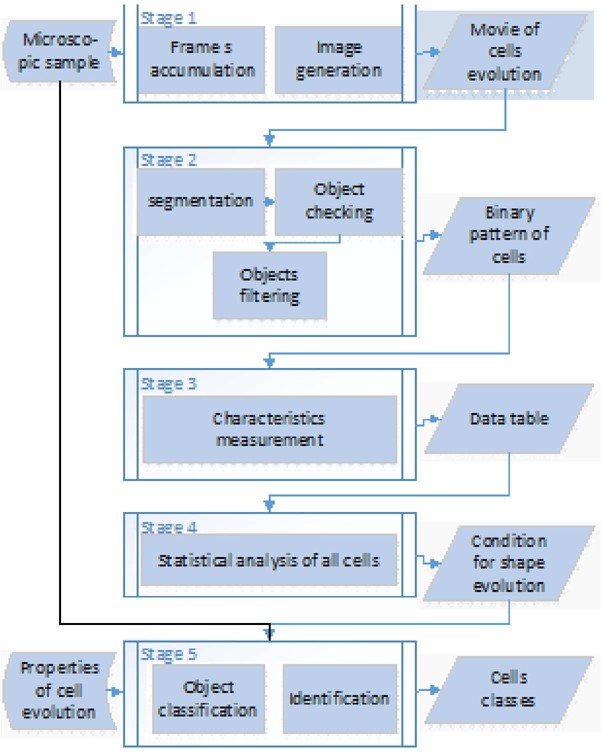


Fig. 4. General scheme of video processing of movable cells

IV. FORMALIZATION OF OPERATIONS WITH DYNAMIC OBJECTS

The formalized concept of the structural model of a dynamic object is an ordered set of patterns $A=\langle A_1, A_2, \ldots, A_n\rangle \subseteq \Omega_n$, which corresponds to the total structural description A, if object pattern $A \in \Omega$ can be completely uniquely reconstructed by combining elements from A, which formally can change over time:

$$A=\delta(A)=A_1 \oplus A_2 \oplus \ldots \oplus A_n, \qquad (1)$$

where δ – operation of structural reconstruction of the pattern according to the structural description; ⊕ – operation of combining patterns from Ω, on which, in the general case, no additional conditions are imposed, except that Ω closed with respect to ⊕. It should be noted that the effect of blocking some objects by others, typical for images in a video sequence, in contrast to many other areas of application of structural analysis, makes the order of combining elements of the visible scene fundamentally important. Therefore, the operation ⊕ in the general case cannot be either symmetric or associative, although sometimes such a restriction is nevertheless imposed.

Thus, the definition of a model of dynamic objects is performed through a changing preimage L with time-constant properties, which consists of n components:

$$L=\delta(L)=L1 \oplus L2 \oplus \cdots \oplus L_n. \qquad (2)$$

In this case, the types of dynamic elements are known, which are specified by the characteristic predicates like elements Mi(Li)∈{0,1}, i=1,…,n. In addition, given m conditions or connections predicates Mk(L)∈{0,1}, k=1,…,m. Then the preimage model takes the form:

$$M(L)=M_1(L_1)\cdot \ldots \cdot M_n(L_n)\cdot M_1(L)\cdot \ldots \cdot M_m(L). \qquad (3)$$

The variables n and m are considered to be time-variable parameters that are also subject to optimization, and predicates are considered as probabilistic or fuzzy, taking values by [0,1]. In this case the problem of structural segmentation of dynamic objects will correspond to the most general case of structural image analysis.

The motion description of objects system can be performed not only at the pixel level, but also at the region level. The characteristics for describing the movement at the region level are the direction of movement, the speed of movement and the intensity of movement in the region determined on its basis.. The determination of motion at the region level and the corresponding motion maps are described in [5, 6].

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The motion maps are used to determine the type and directions of motion in a video sequence and to determine the number of pixels moving in the selected directions. In this case, the motion is described for all nodes (hypothetical centers of motion) through which the moving pixels pass. Thus, the dynamic object model is defined as:

$$A = \delta(A) = OP \oplus OCM \oplus ICM \oplus OQ \oplus IQ. \qquad (4)$$

Here OP, OCM, ICM, OQ, IQ are corresponding motion maps. Thus, the simple motion of a dynamic object can be described using the concepts of an algebraic field and a ring.

There are three types of dynamic objects motion: directional motion, aggregation (moving towards a common center), and scattering (moving away from the center). The signs of aggregation are: a) several objects move to one image region from other areas; b) the speed of motion of these objects is greater than the speed of chaotic motion; c) at least two predominated directions of movement can be distinguished. The signs of scattering include: a) several objects move in the direction from their common center to other regions of the image; b) the speed of their motion exceeds the speed of chaotic movement; c) at least two predominated directions of movement can be distinguished.

For the last two types of motion, it is possible to define the algebraic concepts of a group and a ring based on the presence of a hypothetical center of motion. In this case, by analogy, we can define the additive operation for the group $\mathrm{C}(q^i, \vec{v}, \vec{a})$, with the displacement operation specified on it $\{\circledast\}$: $\mathrm{C} + \mathrm{C} \to \mathrm{C}$. Thus, it is possible to define an algebraic element as a motion towards the center, a neutral element as the absence of motion and an inverse element as a motion from the center. The associativity property is present in this case.

Thus, relative to motion, algebraic groups of events can be distinguished as simple motion and motion relative to a hypothetical center. It is allows to consider the problem of motion from an algebraic point of view with the vector as the reference concept. Pixel displacement vectors δq^i can be obtained based on the calculation of optical flows. The vectors of the integral optical flow are additionally used $\tilde{q}^i$. Therefore, the displacement vector obtained on the basis of the integral optical flux can be expressed as:

$$\tilde{q}^i = \sum_j \delta q^i{}_j,$$

where j is frame change index in the video sequence.

V. PRACTICAL RESULTS

As experiments, the movement of vesicles containing the GLUT4 protein onto the cell membrane was considered. GLUT4, an insulin-regulated glucose transporter protein, is predominantly found in the cytoplasm of adipose and muscle cells in the absence of insulin. Understanding the effect of insulin on the spatio-temporal regulation of intracellular GLUT4 transport is important for elucidating the pathogenesis of type 2 diabetes in humans.

The GLUT4 intracellular transport movement dataset was derived from a total internal reflection fluorescence microscopy (TIRF) image sequence. TIRF microscopy is an optical technique used to observe the fluorescence of individual molecules, based on the phenomenon of total internal reflection. Parameters such as the speed of movement of vesicles in the near-membrane region, their number and density were determined for the vesicles.

Registration of the appearance of vesicles is carried out on binary images obtained because of threshold segmentation, which is the result of calculating the difference between the brightness of the images of the vesicles and the background. Segmentation consists in the fact that the brightness value of each pixel is compared with a threshold value, which is determined based on the correspondence of its local environment to the Gaussian distribution and a predetermined user correction to determine the deviation from the Gaussian distribution.

The selected dynamic objects are controlled by means of motion maps, which make it possible to form a description and conduct monitoring at the level of the graph structure.

The construction of a set of protein transport graphs in cells corresponding to a set of all vesicles isolated at the segmentation stage consists of the following steps:

1) Selection of the initial branches of the graphs in accordance with the number of vesicles.

2) Splitting the video sequence into minimum time intervals is performed for the entire video sequence. The time for each event is determined, and then the minimum time interval according to which the timeline is split.

3) Construction a set of graphs (Fig. 5). For each node, the state of the cell is registered, according to which the number of inputs-outputs for each block is determined. The simple motion block has one input and one output, the intersection block has one input and two outputs, the membrane delivery block is the terminator of the graph branch. Thus, each branch of the graph represents a family tree for an individual vesicle.

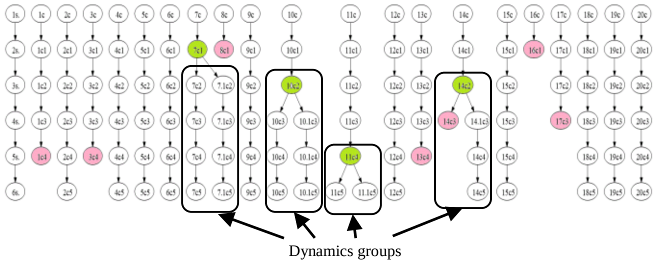


Fig. 5. A set of graphs describing the movement of individual vesicles

The use of graphs allows tracking the stages of GLUT-4 delivery in the cell. It also reduces the analysis time based on the transformation of dynamic objects into dynamic groups on the frame of the video sequence as in Fig. 6.

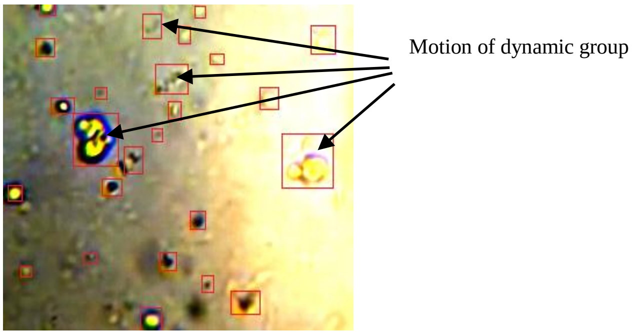


Fig. 6. Combining vesicle tracks into dynamic groups on a video sequence frame

VI. CONCLUSION

The paper formalizes the problem of the motion analysis of dynamic objects and scenes based on the algorithms and methods developed by the authors for analyzing the behavior of the cell population as a system of dynamic objects, which are based on the use of the concept of an integral optical flow. The main types of movement have been determined, which make it possible to distinguish the key moments of the movement of cells in a population and describe the stages of development and interaction of cells. The formalization of operations on dynamic objects has been completed.

ACKNOWLEDGMENT

This research was supported by RFBR and BRFBR, the project numbers are 20-57-00025/BRFFI F20R-134 and BRFFI F20KIGT-006.

REFERENCES

[1] S. Huh, et al. Automated Mitosis Detection of Stem Cell Populations in Phase-Contrast Microscopy Images, IEEE Transaction on Medical Imaging, 2011, vol. 30, iss. 3, pp. 586 –596.

[2] F. Ascione, Investigation of cell dynamics in vitro by time lapse microscopy and image analysis, Chemical Engineering Transactions, 2014, vol. 38, pp. 517–522.

[3] L. Kang, et al., Cell population tracking and lineage construction with spatiotemporal context, Medical Image Analysis, 2008, vol. 12, pp. 546–566.

[4] P. Perner, Tracking Living Cells in Microscopic Images and Description of the Kinetics of the Cells, Procedia Computer Science, 2015, vol. 60, pp. 352–361.

[5] Ch. Chen, Integral optical flow and its application for monitoring dynamic objects from a video sequence, J. of Applied Spectroscopy, 2017, vol. 84, iss. I, pp. 120–128.

[6] H. Chen, Image motion maps and their applications for dynamic object monitoring, Pattern Recognition and Image Analysis, 2019, vol. 29, no.1, pp. 131–143.

Neural Network Approach for Estimating the Level and Volume of Liquid in Transparent Containers

Vladimir Golovko
Intelligent Information
Technologies Department
Brest State Technical University
Brest, Belarus
gva@bstu.by

Egor Mikhno
Intelligent Information
Technologies Department
Brest State Technical University
Brest, Belarus
dzinushi.kun@gmail.com

Aleksandryna Mamyha
Intelligent Information
Technologies Department
Brest State Technical University
Brest, Belarus
mamygas@yahoo.comcom

***Abstract.* The main purpose of this paper is to represent and investigate a neural network approach for determining the volume of the container and the volume of an opaque liquid ina transparent bounded container.**

To achieve the purpose, we apply two models of neural networks, namely AlexNet and eXnet. We have prepared and created a training dataset.

The results of experiments on determining the level and volume of liquid in transparent bounded containers are presented.

Keywords: training dataset, deep neural network, container volume, liquid level, liquid volume

I. INTRODUCTION AND THE PROBLEM DEFINITION

Advances in computer vision are leading to new practical and research solutions [3, 4]. Currently, an urgent problem inthe field of computer vision is the determination of the volume of opaque liquid in transparent bounded containers. To solve this problem, it is proposed to use a neural network approach [4].

In this paper, we investigate various models for determining the volume of opaque liquid in transparent bounded containers without trademarks. The results obtained allow speaking about the degree of applicability of the used models to the proposed recognition problem.

Most of the research in the field of computer vision is focused on the problems of search (detection [7, 8]) and recognition (classification [9, 10]) of solid physical objects. However, little attention is paid to working with liquid objects[5]. Let us consider the problems that were solved in their article:

1. Calculation of the total volume of the container. It is necessary to calculate the volume of a certain bounded container (50 ml, 200 ml, etc.).
2. Percentage-based estimation of the fullness of the container. It consists in getting information about how full the container is (empty, 10%, 50%, etc.).
3. Comparative estimation of the volume. It is necessary to answer the question of whether it is possible to pour the contents of one container into another one. The planned answer options are: yes, no or I cannot say (because there are opaque containers in the dataset).
4. Estimation based on the placement of the container. It consists in determining the amount of liquid after changing the placement of the container (for example, if it is tilted).

To solve these problems, we used a Container with liquid contents (COQE) dataset with a large number of images, which is reasonable for solving all 4 problems. COQE contains more than 10,000 images of various categories of containers: bottles, glasses, jugs, teapots, etc.

The dataset was formed on an open platform, on which everyone could upload and mark up the necessary images. Random people took photographs of various containers with and without liquid using cameras or mobile phones; the volume of containers was measured using a measuring cup or any other convenient way.

In addition to the usual images, 34 CAD models were also loaded from a three-dimensional warehouse, indicating the degree of correspondence between the CAD models and the containers from the images.

The first three problems are mainly related to estimating the geometry of the container and its contents. The last problem considers the assessment of

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

the behavior of the liquid inside the container, for example, at different tilting angles.

By the time the platform was closed, more than 10,000 marked-up images had been collected, 6386 of which were used for training, 3000 – for testing and 1000 – for checking control.

As the tested models, a modification of the well-known model of the CNN family – ResNet-18 – was considered. We used a variation of the configuration of the model, that was pre-trained on the ImageNet dataset. The context tensor was combined with the conv4 1 layer, which accepts data of size 28 x 28 x 128 as the input one. As a result, the size of the input layer was increased to 28 x 28 x 209. The size of the group sample was 96 images. The resulting architecture was called Context Containers ResNet (CRC).

The classical cross-entropy [6] was used as the error function. The process of training the network was carried out in two stages: the estimation of the volume and contents was conducted separately. We will talk about the results obtained in the following sections.

II. PROPOSED APPROACH

Our problem is to estimate the percentage of the fullness of the container (similar to point 2 discussed above). Models of the CNN family – AlexNet and eXnet – were also used as tested models.

A. Creating datasets

When solving the problem, two datasets were used: one to determine the volume of the container and one to determine the volume of liquid in containers. Data collection was carried out using professional cameras “Canon EOS m50 kit18-150mm” and “Canon PowerShot S5 IS”.

Fig. 1. Examples from the Container Volume Dataset

To determine the volume of transparent bounded container, 900 images were collected, 809 of which were allocated for training and 91 – for testing. To determine the volume of liquid in similar conditions, 599 images were collected, 539 of which were allocated for training and 60 – for testing.

Fig. 2. Examples from the Liquid Volume and Level Dataset

We used 3 types of bounded containers: transparent plastic bottles with volumes of 500 ml, 1000 ml and 1500 ml, respectively (excluding brands and subtypes of containers themselves). The investigated volumes of containers as well as the levels and volumes of liquid in containers will be taken into account when marking up the data along with other additional information. The photographs of the bottles were taken with account of different lighting, at different angles, with different sharpness and different backgrounds. The creation of such a context is a necessary criterion for assessing the efficiency of the system under the given conditions.

The first dataset consists of transparent plastic bottles with- out liquid and is only needed for the problem of determining the volume of the container itself. For each type of volume, 3 bottles of different producers were chosen. As a result, 9 bottles were taken. All bottles have different shapes except of two: one 1500 ml bottle and one 500 ml bottle have the same shape.

TABLE I. EXAMPLES OF CONDITIONS OF THE CONTAINER VOLUME DATASET

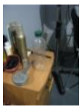
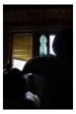
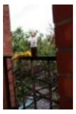
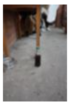
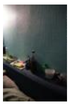
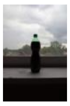
	light	***sharpness***	***corner***	***background***	***room***
500 ml					
1000 ml					
1500 ml					

TABLE II. EXAMPLES OF CONDITIONS OF THE LIQUID VOLUME AND LEVEL DATASET

	light	*sharpness*	*corner*	*background*	*room*
0%					
25%					
50%					
75%					
100%					

The second dataset consists of transparent plastic bottles with liquid and is necessary for the problem of determining the volume of liquid in a container as well as for the problem of determining the level of this liquid. 2 bottles of each volume were selected (6 in total), into which an opaque liquid was poured at 4 pre-selected levels: 25%, 50%, 75%, 100% and 0% for empty bottles.

The same containers were used for two datasets.

B. Dataset preprocessing

Before the direct processing the data by the model, a number of transformations of the input data were carried out:

- randomly change the brightness, contrast, intensity and hue of an image;
- resize the input image to the given size;
- rotate the image to an angle;
- crop the given image anywhere;
- horizontally flip the given image randomly with a given probability;
- perform a random perspective transformation of the given image with the given probability.

C. Description of the AlexNet and eXnet models

AlexNet [1] is a universal architecture that can provide high accuracy for complex and large datasets (Fig. 3). AlexNet is one of the most well-known architectures for object detection problems in the field of computer vision. The AlexNet model has about 60 million parameters; the size of the input image – 224x224. In the experiments, we used a pre-trained version of the model on the ImageNet dataset.

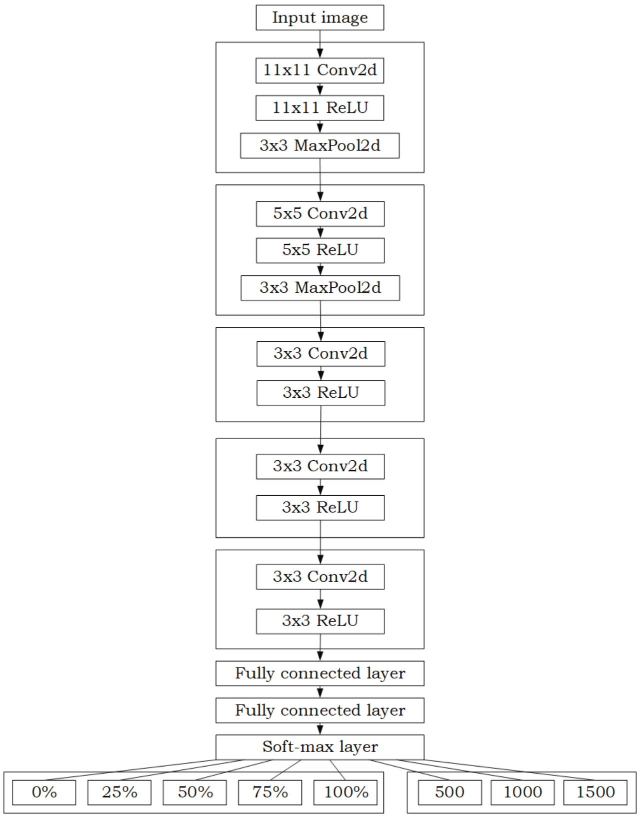


Fig. 3. Architecture of AlexNet

The architecture of the eXnet [2] network (Expression Net) is based on the parallel extraction of objects borrowed from the Inceptions model range but that contains a much smaller number of parameters (eXnet – 4.57 million, AlexNet – 62.3 million, InceptionV4 – over 40 million). Models of this kind provide high performance in systems with limited hardware, and a smaller number of tunable parameters allow high generalization within the framework of the Occam's razor rule (a simpler explanation of the essence is preferable from two competing theories; this rule states that entities should not multiply unnecessarily). An input image size: 48x48. In the experiments, we used a self-trained version of the model based on standard initialization of weights from Pytorch.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

An example of architecture is shown in Fig. 4:

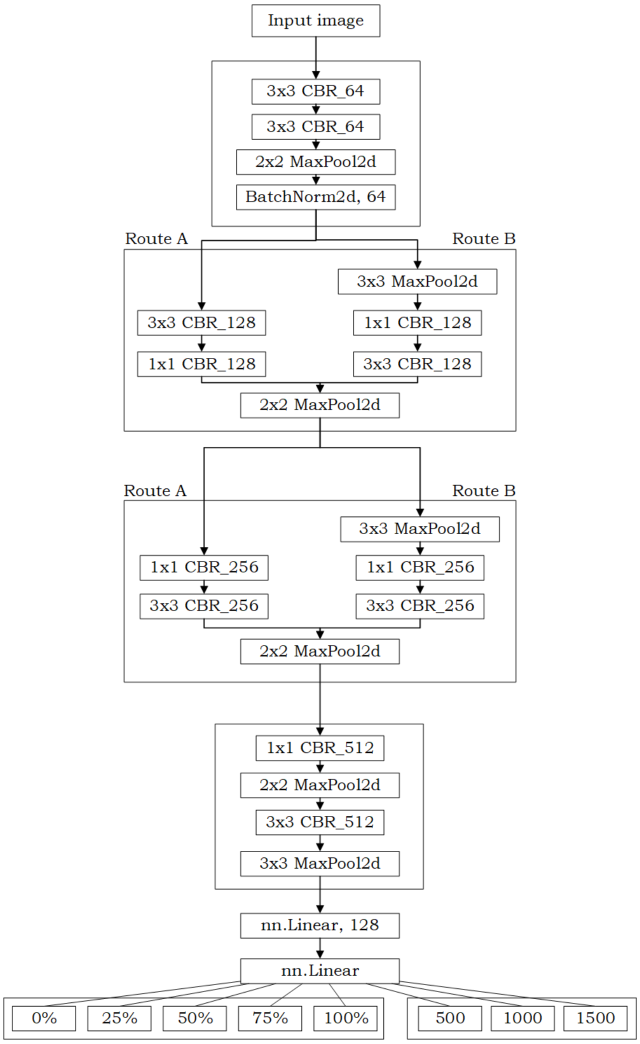


Fig. 4. Architecture of eXnet

III. RESULTS

When comparing the results, validation accuracy was used, at which the values were rounded either to the True Positive side or to the False Negative side.

Six independent experiments were carried out: three for each model at a different number of epochs. The best value for the specified period was taken as the final value. As a result, the AlexNet model showed worse results than the lightweight eXnet model, despite the lack of any pre-training in the case of the last one. In particular, its advantage was observed when assessing the volume of the liquid itself (0.86 versus 0.69).

Based on the test data obtained, we can conclude that the neural network built on the eXnet architecture was well trained, in particular, on a dataset with volumes of containers. The accuracy rate on the second dataset is lower because it contains fewer images and is a more difficult problem for training a neural network. The first dataset has 3 classes,in which different volumes of containers are used. There is no liquid in this dataset. In the second dataset, there are 5 different classes with different volumes of containers (data about containers is used for training in the first dataset) and they also contain an opaque liquid in different volumes.

TABLE III. VOLUME CONTAINER DATASET RESULTS

Epochs	AlexNet	eXnet
30	0.92129	0.93259
50	0.92148	0.92209
100	0.91088	0.94371

TABLE IV. LIQUID VOLUME IN CONTAINER DATASET

Epochs	AlexNet	eXnet
30	0.67794	0.72714
50	0.69378	0.79546
100	0.69432	0.86124

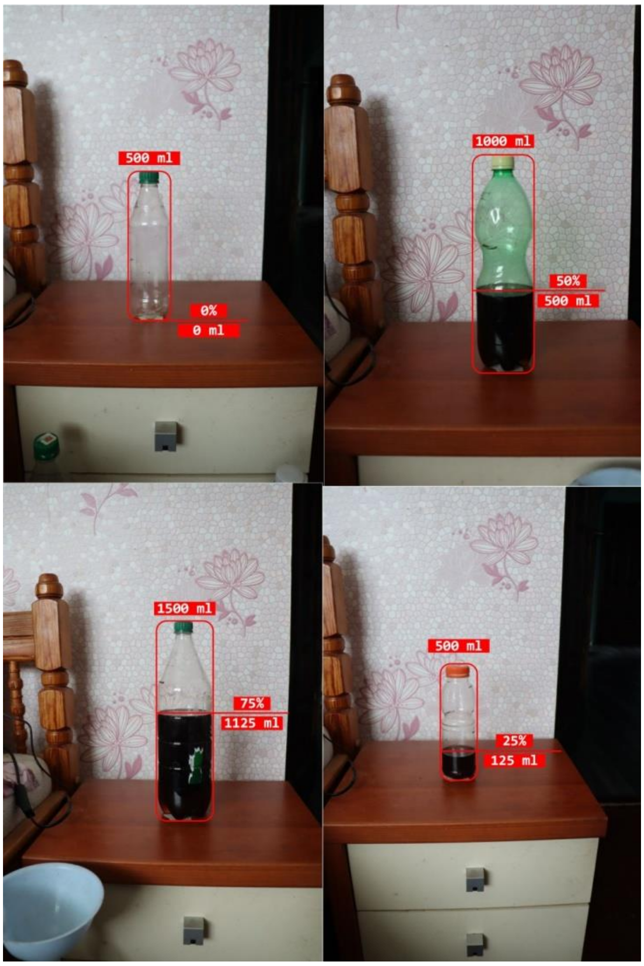


Fig. 5. Result of the system activity

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The conducted tests allow us to conclude that the data selection and the choice of models were reasonable. To increase theefficiency of the system, it is necessary to increase the amount of data for the conditions described above or to increase the flexibility of the system by increasing the number of conditions themselves.

IV. CONCLUSION

As a result of the conducted studies, a solution was obtained for estimating the volume of the container as well as the level and volume of the liquid in a transparent bounded container with a validation accuracy of 0.94 and 0.86, respectively.

It is generally assumed that the average person copes with the problem by 0.95. The experimental results obtained are still inferior to a human (0.94 and 0.86), so we may continue experiments on this topic in the future.

REFERENCES

[1] Md Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, Mst S. Nasrin, B. C. Van Essen, A. A. S. Awwal, V. K. Asari: “The History Began from AlexNet: A Comprehensive Survey on Deep Learning Ap- proaches.“ https://arxiv.org/ftp/arxiv/papers/1803/1803.01164.pdf. Ac-cessed 6 Aug 2021

[2] M. N. Riaz: eXnet: “An Efficient Approach for Emotion Recognition in the Wild.“ https://www.researchgate.net/publication/339324794. Accessed 6 Aug 2021.

[3] V. Golovko, A. Kroshchanka, M. Kovalev, V. Taberko, and D. Ivaniuk “Neuro-Symbolic Artificial Intelligence: Application for Control theQuality of Product Labeling.“ 2020 International Conference on Open Semantic Technologies for Intelligent Systems, 2020, pp. 81-101.

[4] V. Golovko, A. Kroshchanka, E. Mikhno: “Brands and Caps Labeling Recognition in Images Using Deep Learning.“ Pattern Recognition and Information Processing, 2019, pp. 35-51.

[5] R. Mottaghi, C. Schenck, D. Fox, A. Farhadi: “See the Glass Half Full: Reasoning about Liquid Containers, their Volume and Content.“ https://arxiv.org/pdf/1701.02718.pdf. Accessed 6 Aug 2021

[6] J. Brownlee: “A Gentle Introduction to Cross-Entropy for Machine Learning.“ https://machinelearningmastery.com/cross-entropy-for-machine-learning/. Accessed 6 Aug 2021

[7] V. Golovko, A. Kroshchanka, E. Mikhno, M. Komar, A. Sachenko: “Deep convolutional neural network for detection of solar panels“. Data- Centric Business and Applications, 2021, pp. 371-389.

[8] V Golovko, A. Kroshchanka, E. Mikhno, S. Bezobrazov, M. Komar, A. Sachenko, I. Shylinska: “Deep convolutional neural network forrecognizing the images of text documents“. CEUR Workhop Proceeding,2019, 297-306.

[9] V. Golovko, E. Mikhno, A. Brich, A. Sachenko: “A shallow convo- lutional neural network for accurate handwritten digits classification“. International Conference on Pattern Recognition and Information Processing, 2016, pp. 77-85.

[10] A. A. Elngar, M. Arafa, A. Fathy, B. Moustafa, O. Mahmoudm, M. Shaban, N. Fawzy : “Image Classification Based On CNN: A Survey.“ https://www.researchgate.net/publication/352152485. Accessed 6 Aug2021

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Lightweight Deep Neural Networks for Dense Crowd Counting Estimation

Stanislav Sholtanyuk
Belarusian State University
Minsk, Belarus
ssholtanyuk@bsu.by
ORCID 0000-0003-0266-7135

Aliaksandr Leunikau
Belarusian State University
Minsk, Belarus
Alex.levnikov@gmail.com

Abstract. **In this paper, productiveness problems of deep neural networks for dense crowd counting prediction have been explored. Deep neural network CSRNet has been considered, and its shallow modifications (named CSRShNet-1 and CSRShNet-2) have been designed and researched. It has been shown that for relatively small crowds (up to 500 people) it is possible to reduce training time by using shallow networks with keeping an appropriate prediction accuracy.**

Keywords: **crowd counting, deep neural networks, convolutional neural networks, supervised learning, neural network performance, neural network accuracy**

I. INTRODUCTION

Nowadays crowd size estimation and prediction of various crowd characteristics are important tasks for such activities as industry, traffic organization, social services, security systems and many others. By knowing crowd characteristics, it is possible to make immediate decisions for safety and preventing emergencies.

Using deep convolutional neural networks is the most common approach for crowd counting. Most of them use supervised learning algorithms. State-of-the-art methods include many implementations, e.g., CSRNet [1], D-ConvNet [2], MCNN [6], MRCNet [3], SANet [5], SPANet [4].

For our research, the CSRNet is chosen as the basic model because it has best results for counting a highly congested crowd, and it has simple realisation. This neural network demonstrates accurate results, but the training process is characterized by long time. We constructed and investigated shallow networks based on CSRNet which predict crowd counting with a decent accuracy and not consume too much time.

II. DESCRIPTION OF MODELS AND THEIR FEATURES

A. *CSRNet and shallow networks*

CSRNet has sequential architecture and consists of convolutional and pooling layers, as pictured on Fig. 1. The input for this neural network is an image with arbitrary resolution, and the output is the estimation of crowd density map represented with a matrix with 1/8 size of the initial image. The neural network consists of two parts: pre-trained layers sequence from VGG-16 network [7] and successive trainable layers. All convolutional layers use kernels with size 3x3, but trainable layers use dilated kernels (Fig. 2).

Two shallow modifications, denoted as CSRShNet-1 and CSRShNet-2, have been designed for reducing the training time. They also consist of some pre-trained VGG-16 layers, and trainable layers use dilated kernels.

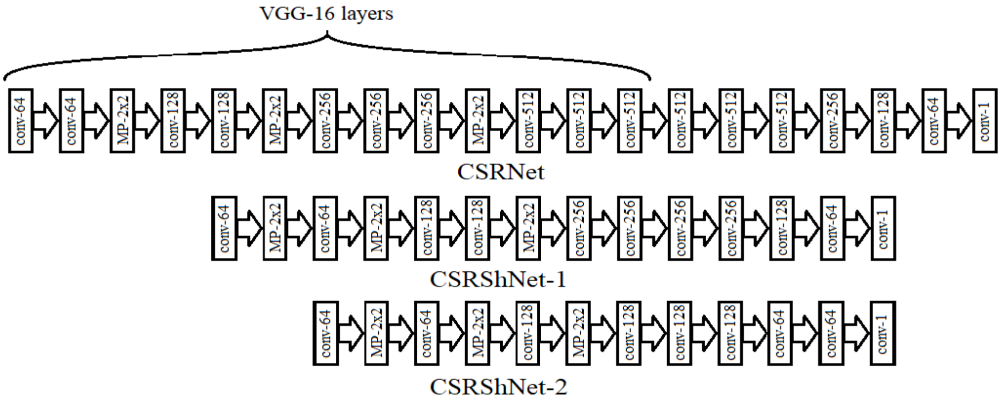


Fig. 1. Architectures of three considered neural networks, from top to bottom: CSRNet (the original one), CSRShNet-1, CSRShNet-2. Convolutional layers are denoted as conv-X, where X stands for number of output channels, and max pooling layers (which use pooling kernels with size of 2x2) are denoted as MP-2x2. In each neural network, there are pre-trained layers from VGG-16 network. Convolutional layers from VGG-16 have dilated rate 1, and other have dilated rate 2

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

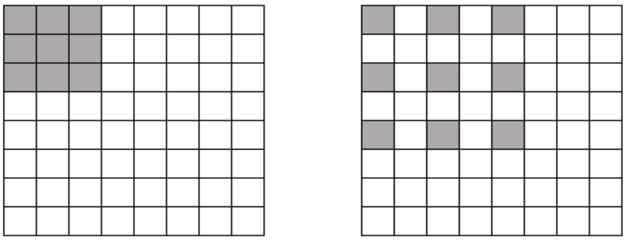


Fig. 2. 3x3 convolutional kernels with dilation rates 1 (left) and 2 (right)

B. ShanghaiTech dataset

In this research, ShanghaiTech dataset [3] of crowd images with different resolution was used for training and testing the neural networks. This dataset is divided into two parts, A and B. In the A part, there are as many images as 300 for training and 182 for testing. In the B part, there are 400 and 316 images in training and testing parts, respectively. Statistical characteristics of both samples are represented in Table I. As shown in this table, images from the part A represent larger crowds.

All neural network was trained on both parts with supervised learning. Ground truth was calculated with specific MAT-files that contain information about counting of crowd. Those files are included in ShanghaiTech dataset. It contains label positions for all people's heads for all images.

C. Software implementation

For realizations networks we use Google Colab with GPU. PyTorch framework on Python was used for neural network implementations [8].

III. METHODOLOGY

Program implementation includes the following steps:

Step 1: *Construction of ground truth density maps for each image*. For this, from the corresponding MAT-file, matrix with zeros and ones is constructed (one stands for a label, corresponding to an individual, and zero means no label). Then, gaussian blurring is applied to this matrix. Thus, the density map is created (Fig. 3b).

TABLE I. NUMERICAL CHARACTERISTICS OF CROWD SIZES ON IMAGES FROM SHANGHAITECH DATASET

Characteristic	**Part A**	**Part B**
Minimal value	33	9
1st quartile	217	54
Median	359	95.5
3rd quartile	600	165.5
Maximal value	3138	576

Step 2: *Training*. For training the neural networks on a given training set (from either part A or part B) four copies of each image are taken, and the quadrupled sample is shuffled. Validation sample is composed of 1/5 randomly picked images from the initial dataset. Mean squared error (MSE) has been used as a loss function:

$$MSE = \frac{1}{n}\sum_{i=1}^{N}\left(\hat{n}_i - n_i\right)^2,$$

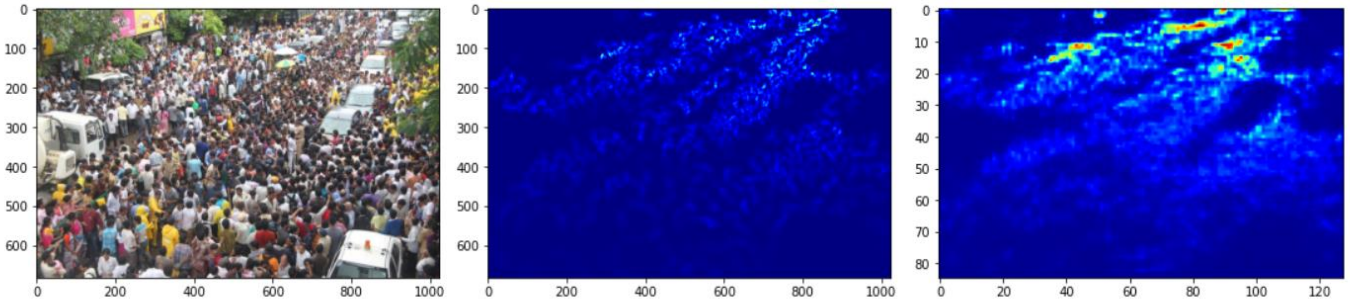
where N is sample size, $\hat{n}_i$ is estimation of the crowd size on i-th image, and n_i is the ground truth for the crowd counting. Stochastic gradient descent is used as a network optimizer. Neural network is saved after each epoch. Besides, saving the best iteration is taking place.

Step 3: *Testing*. After training, neural networks have been tested on both part A and B. Mean average error (MAE) and mean average percentage error (MAPE) was used as testing results:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{n}_i - n_i\right|,$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\hat{n}_i - n_i\right|}{n_i}.$$

An example of results is shown on Fig. 3.



a) b) c)

Fig. 3. An initial image (a). Ground truth density map (b). Density map estimated with CSRShNet-2 network (c)

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

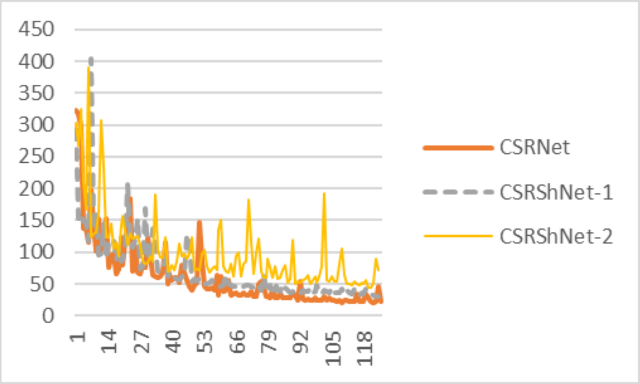


Fig. 4. Time performance results when training on part A of ShanghaiTech dataset. Horizontal axis is number of epochs, and vertical axis is MAE for the validating dataset

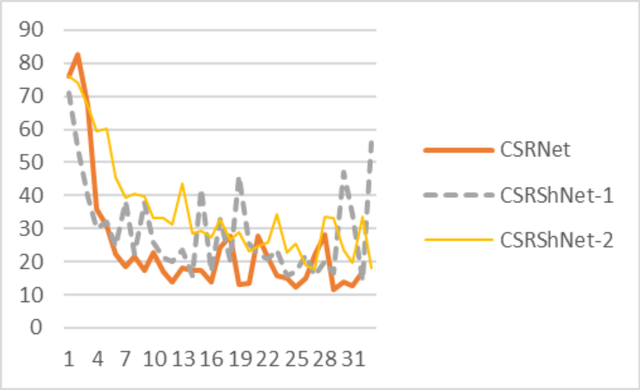


Fig. 5. Time performance results when training on part B of ShanghaiTech dataset. Horizontal axis is number of epochs, and vertical axis is MAE for the validating dataset

TABLE II. TIME AND ACCURACY RESULTS AFTER TRAINING

Name	**Training Dataset**[a]	**Training Time, sec**[b]	**MAE for part A**	**MRE for part A**	**MAE for part B**	**MRE for part B**
CSRNet	A	1413.93	71.64	0.187	22.80	0.208
CSRShNet-1	A	316.98	88.42	0.253	22.51	0.282
CSRShNet-2	A	199.61	97.69	0.301	31.14	0.426
CSRNet	B	719.75	126.55	0.299	13.87	0.129
CSRShNet-1	B	173.01	142.16	0.336	17.95	0.193
CSRShNet-2	B	99.30	171.12	0.440	21.17	0.204

[a] A stands for part A of ShanghaiTech dataset, and B stands for part B of the same dataset.

[b] Time for only epoch themselves is shown, all other activities have not been considered.

IV. RESULTS

Results for accuracy during training the neural networks on part A for 124 epochs and on part B for 33 epochs are represented on Fig. –Fig. .

From the chart on Fig. , it can be concluded that neural networks accuracy improved significantly after 15th epoch, and approximately after 55th epoch, CSRNet and CSRShNet-1 demonstrate stable results, and results of CSRShNet-2 are non-stable. According to chart on Fig. , CSRShNet-2 has greater MAEs than CSRNet, and CSRShNet-1 on different epochs has accuracy comparable with both other neural nets, so an instability is taking place.

Results of neural networks training are shown in Table II. For shallow networks training it was faster by 4-7 times than for the original network. MAE became 25-40% greater when training on the part A, and 30-60% greater when training on the part B. In terms of MAPE, performance of the shallow networks isn't very good when training on the part A since they demonstrate 25–30% error for the same sample, but there are decent results after training on part B which are about 20%. However, when testing the shallow networks on another sample rather than one they have been trained on, unsatisfactory results are obtained (25-45% when training on part A and testing on part B, and 30-45% when training on part B and testing on part A).

V. CONCLUSIONS

Proposed shallow neural networks have been designed which can be used in predicting crowd characteristics for dense crowds up to 500 people. Such network architectures for convolutional neural networks don't require too much time to train and have good performance in crowd counting.

The number of epochs for training CSRShNet-1 and CSRShNet-2 can be reduced to 30-50 without significant productiveness loss (MAPE is up to 20% for CSRNet and both shallow modifications which are trained on part B of ShanghaiTech dataset), therefore the training time can also be saved.

ACKNOWLEDGMENT

The reported study was funded by RFBR and BRFBR, project number 20-57-00025 / BRFFI F20R-134.

REFERENCES

[1] Y. Li, X. Zhang, D. Chen, "CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 1091-1100.

[2] Z. Shi, L. Zhang, Y. Liu, X. Cao, Y. Ye, M.-M. Cheng, G. Zheng, "Crowd Counting with Deep Negative Correlation Learning," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5382-5390.

[3] Y. Zhang, D. Zhou, S. Chen, S. Gao, Y. Ma, "Single-Image Crowd Counting via Multi-Column Convolutional Neural Network," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 589-597.

[4] R. Bahmanyar, E. Vig, P. Reinartz, "MRCNet: Crowd Counting and Density Map Estimation in Aerial and Ground

Imagery," British Machine Vision Conference; Workshop on Object Detection and Recognition for Security Screenin (BMVC-ODRSS), 2019, pp. 1-12.

[5] X. Cao, Z. Wang, Y. Zhao, F. Su, "Scale Aggregation Network for Accurate and Efficient Crowd Counting," European Conference on Computer Vision (ECCV), 2018, pp. 734-750.

[6] Z.-Q. Cheng, J.-X. Li, Q. Dai, X. Wu, A. G. Hauptmann, "Learning Spatial Awareness to Improve Crowd Counting," IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 6152-6161.

[7] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition" arXiv preprint arXiv:1409.1556, 2014.

[8] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," Advances in Neural Information Processing Systems 32, 2019, pp. 8024-8035.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Towards Creation of SmileID Obtained from Face Biometrics Binded to Concantenated Error-Correcting Codes

Boris Assanovich
Yanka Kupala State University of Grodno
Grodno, Belarus,
bas@grsu.by

***Abstract.* A new biometric system based on smile imprint obtained from face video frames with the use of stacked autoencoder and fuzzy commitment scheme exploiting the concatenated Reed-Solomon and linear error-correcting codes has been proposed. The performance of system verified with smiles from UvA-NEMO database has shown the achievement of FRR=0.5-1% for data size of 31-63 elements.**

***Keywords*: autoencoder, biometric features, smile imprint, fuzzy commitment, concatenated Reed-Solomon codes**

I. INTRODUCTION

A human smile that is a key factor in determining person’s psychological state, can also be used as a behavioral biometric identification element. Recently, several applications have appeared that implement the concept of SmileID [1] related to the face biometrics where identity is verified remotely. Last years, facial biometry has become one of the most preferred biometric methods both in video surveillance and in banking [2] due to the fact that it does not require precision equipment and uses a non-contact data processing method.

Few years ago Murat Taskiran et all [3] applied dynamic face features extraction from videos and used them for face recognition. Authors performed the analysis of face videos and extracted the statistical properties of facial distances during several phases of spontaneous and posed smiles on the UvA-NEMO smile database, that have been created before for biometric applications by Dibeklioğlu et all [4].

Despite the fact that there is a number of techniques that use facial dynamics to identify a person with extracted from face various parameters, the deep learning methods are increasingly being exploited for recognition tasks. Classical linear methods of image processing and feature extraction based on principal components analysis (PCA) are replaced by non-linear transformations. Compared to PCA, the use of an autoencoder increases significantly the classification accuracy, especially with a large number of items [5].

In this paper, we will consider the use of an autoencoder to extract soft biometric information from a person’s smile and apply them for authentication to provide access to digital services. In addition, we will propose an error correction method for creating biometric templates with the use of Juels and Wattenberg (JW) fuzzy commitment scheme [6] that can be revoked if compromised.

The article demonstrates the creation of a digital face *smile imprint* to provide a secure biometric interface for organizing access to digital services. It continues in Section 2 by reviewing the related structure of an autoencoder and the error correcting code parameters. The methodology used for the research work is presented in Section 3. Section 4 describes the results of this study. The paper concludes in Section 5 and also provides some direction for future work.

II. AUTOENCODERS AND ERROR-CORRECTING CODES

A. Autoencoders

There are several types of autoencoders. Sparse Autoencoder has a dimension of the hidden layer that is greater than the input. It consists of two parts: coder (encoder) G and decoder F [7]. The encoder translates the input signal into its representation (code): $y = G(x)$, and the decoder restores the signal by its code: $x' = F(y)$. Moreover, the transformation functions F and G contain activation function, weights and biases of trained artificial neural networks. By changing the mappings F, G, autoencoder tends to find out the identity function $x = F(G(x))$, minimizing the kind of error based on some functional $L = (x, F(G(x')))$. Let consider that a vector $x \in \mathbf{R}$ connected to the input of an autoencoder. Then the encoder maps the vector x to another vector $y \in \mathbf{R}$ as follows $y = h^i(W^i x + b^i)$ where the superscript i indicates the i-th layer. Then h^i is a transfer function for the encoder, $W^i \in \mathbf{R}$ is a weight matrix, and $b^i \in \mathrm{R}$

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

is a bias vector. Hence, the decoder maps the encoded representation *y* back into an estimate of the original input vector *x*, as follows: $x' = h^{i+1}(W^{i+1}x + b^{i+1})$ where the superscript *i*+1 represents the $i+1$ layer. Then a transfer function h^{i+1} for the decoder has a factor $W^{i+1} \in \mathbf{R}$ that is a weight matrix, and $b^{i+1} \in \mathbf{R}$ is a bias vector correspondently. If the encoder has only two layers then the expression for the transfer function can be represented as $x' = h^2(W^2 y + b^2)$. In our setup we applied the so-called stacked autoencoder (SAE) that is a neural network including only 2 layers where output of each hidden layer is connected to the input of the successive one. The effectiveness of user comparisons depends on similarity rates, which are often determined by the distribution of root mean square (RMS) distances of their characteristics.

The more the two distributions are separated and the smaller the standard deviation for each distribution, the better is the separation of the classified classes. This property of distributions is estimated by such a parameter as decidability index

$$DI = \frac{|\mu_g - \mu_i|}{\sqrt{\sigma_g^2 + \sigma_i^2 / 2}}, \qquad (1)$$

where μ_g, μ_i and σ_g, σ_i are the means and standard deviations of genuine and imposter distributions.

In addition to the decidability index, an equal error rate (ERR), which is the rate at which a false accept rate (FAR) is equal to a false rejection rate (FRR), is normally used as a measure of biometric system (BS) verification accuracy. In biometrics FAR is the rate at which an imposter print is incorrectly accepted as genuine and FRR is the rate at which a genuine print is incorrectly rejected as imposter.

The use of autoencoders will make it possible to present compactly biometric features and then apply the scheme to organize Human Computer Interface (HCI), where instead of tokens and passwords, a biometric key can be exploited for authentication. To handle the variability inherent in biometric verification, it is necessary to create and store a template for each user. To create it, we will use the fuzzy commitment scheme with application of error-correcting codes (ECC).

B. *Error-Correcting Codes*

In this paper, in contrast to the generally accepted application of binary ECC in Biometric System, we consider the use of non-binary Reed-Solomon codes (RS). However, the parameters we put for evaluating the effectiveness of biometric systems can easily be used for binary ECC.

More formally, a non-binary ECC have codewords *C* of a certain length *n*, consisting of symbols belonging to the set *C**∈{0, *q*-1}, where *q* is some integer. The use of the code is aimed at encoding a message of length *k* with the addition of redundancy *r*=*n*-*k*, so that if a certain number of symbols is corrupted, it is still possible to get the correct codeword *C* and message *R*. The robustness of an ECC depends upon some distance between codewords. Let the RS code be defined over the Galois Field $\mathrm{GF}(2^m)$ with a redundancy of *n*-*k* symbols and let the Symbol Error Rate (SER) caused by fuzziness of biometric data or so called “biometric noise” be *p*. The important performance parameters to consider when using EEC in BS are still FAR and FRR. The FRR, which depends on SER of RS code, is actually can be upper bounded by the probability that more than $(n-k)/2$ errors occur, i.e. [8]:

$$FRR \le \sum_{i=\lfloor (n-k)/2 \rfloor + 1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \approx (np)^{\lfloor (n-k)/2 \rfloor + 1}, \qquad (2)$$

where *n*, *k* are the ECC parameters defined above, and *t* is its error correction capacity, i.e. the ability to correct any set of up to *t* symbol errors.

Considering that an imposter produces a random syndrome during decoding for uncorrectable error patterns, and it is accepted by BS, FAR will be the probability that it is valid [8], i.e.

$$FAR = \sum_{i=0}^{\lfloor (n-k)/2 \rfloor} \binom{n}{i} (n+1)^{-(n-k)} \approx q^{-(n-k)/2} \qquad (3)$$

From (2), (3), we see that FRR, FAR can be reduced by increasing (*n*-*k*) and *t*.

The dimension of the $\mathrm{GF}(2^m)$ field and the redundancy of the RS code significantly affect the length of the cryptographic key *R* in the biometric JW scheme. The use of non-binary RS codes has the advantage that an increase in the symbol dimension leads to an increase in the length of their bit representation. On the other hand, in order to obtain high error correction capacity *t* of ECC, it is necessary to increase the redundancy, which reduces the code rate *k/n* and leads to the need to use several ECC codewords. For example, to obtain a user's secret key *R* with a length of more than 128 bits, when using RS (31,9) code over $\mathrm{GF}(2^5)$ with a code rate *k/n* of about 0.3, only 45 bits can be placed in one codeword. Hence for the key length of more than 128 bits, three such codewords are required. Whereas when using the RS (63,15) code with the rate 0.24 over $\mathrm{GF}(2^6)$, the key length of 180 bits can be distributed between two codewords of this RS code.

To evaluate the effectiveness of these codes, the corresponding FRR values for these codes have been calculated at different symbol error probabilities and placed in Table I.

TABLE I. EVALUATION OF FRR FOR RS CODES (63,15), (31,9)

SER	FRR: RS (63,15)	FRR: RS (31,9)
0.0050	2.8692e-13	1.9230e-10
0.0100	9.6275e-06	7.8766e-07
0.0150	0.2431	1.0220e-04

It follows from Table 1 that with the increase of *p*, FRR grows exponentially and, taking into account (2), to reach FRR=1.0e-04 for RS code (63,15), it is necessary to reduce SER by 1.36 times. Thus, in order to achieve the required performance of the biometric system, it is relevant to introduce a significant redundancy by ECC applied. In this case one of the ways to increase the efficiency of error-correcting coding is the transition to concatenated ECC. It is a class of error-correcting codes derived by combining an inner code and an outer code that can be tuned in a given way and show better performance than ECC of a certain type. According to our estimates, to reduce SER by several times, it is sufficient to use a class of linear codes with a suitable length.

III. PROPOSED SYSTEM

In this paper, we have used an autoencoder to obtain the biometric data on a person's smile and bind it to a secure user key. Due to the fact that biometric data has instability, error-correctiing codes should be adopted to ensure that fuzziness of biometric data can be alleviated.

In this study we propose to apply the concatenated ECC based on non-binary RS codes and binary linear codes with the use of hard-decision decoding technique. Initially, the user key will be encoded with a non-binary RS code, and then the bit representations of the symbols will be additionally encoded with a linear binary code. Then, after eliminating "biometric noise" in the received code vectors, the key is extracted by decoding the code constructions in reverse order.

The model of the proposed system is depicted in Fig. 1. Wel introduce the term *Smile Imprint* of biometric data obtained from SAE output layer and concatenated $Y = y_1 \cup y_2 ... \cup y_M$ to form a supervector from several vectors y_i, where *M* is a number of processed frames from "smile signal". It should be noted that the *Preprocessor* block also performs such operations as smile detection and smile frames selection related to its main three phases (onset, apex, offset) as it is also described in [9].

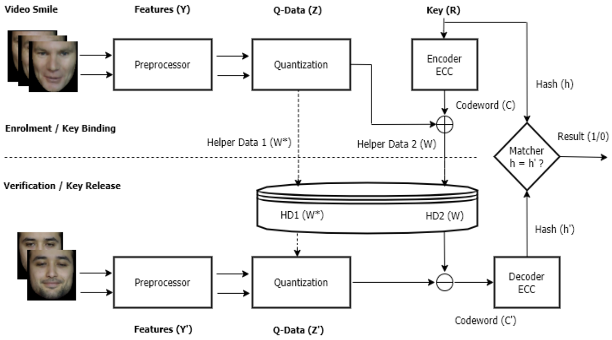


Fig. 1. A System model

At the stage of *Qantization*, the real values *Y* converted into their quantized versions *Z* producing also deviations *W** from the centers of the quantization intervals used as Helper Data 1 (HD1). In the *Encoder* block, the user's password or Key *R* is encoded with one or more ECC codewords, depending on the required password strength and required FRR, which will be discussed below. Further, for a biometric authentication purposes the bit representation of the resulting codeword is added modulo 2 to the quantized and encoded version *C* of smile imprint *Z*. The result is a sequence *W*, which serves as Helper Data 2 (HD2). Further, we will consider the authentication scenario, although this BS can also implement the identification process.

The proposed system works in two operating modes (see Figure 2). In the first mode, the user is registered and his smile imprint is obtained from SAE and with the use of HD1 and HD2 linked to the secret Key *R*, the *h*-hash of which is calculated and stored in the Biometric Database (BD). During verification, the reverse process of decoupling the "auxiliary" data HD1, HD2, decoding *C'* and comparing two hashes *h* and *h'* is performed by a *Matcher*.

Thus, the user smile imprint captured from his video can serve as the biometric key to organize the access to different external digital services. At the same time, the proposed system implements the JW scheme with the use of two helper data HD1, HD2 sets that can be stored in the public domain. To represent information in HD1 and HD2, various methods can be used up to encryption, which is determined by the complexity and required speed of the implemented biometric system. System parameters such as the length of the cryptographic key, the type and characteristics of the ECC must be determined by the quality of biometric data, the number of users, the resistance to external attacks, etc.

IV. EXPERIMENTAL RESULTS

A series of experiments were performed with SAE to get good compact biometric features. To reduce time spent, in these experiments the subsets of 40 subjects randomly selected from the entire UvA-NEMO Database were used, reproducing a posed smile. Then normalized grayscale images from corresponding video of 112x112 pixels in size, scaled to 50%, creating a vector length of the input layer of 6272 elements were used for unsupervised learning of SAE. Image scaling made it possible to reduce the time spent on data processing while maintaining sufficient differentiation of users' smiles when using the selected autoencoder structure. Combinations of the second and third SAE layer dimensions had values 255/63, 127/63, 127/31, and 63/31. The selected values were determined by the length of the applied RS codes, as well as the chosen dimension of equidistant quantization.

To perform quantization and encoding, the following structures were chosen 127/63 and 63/31. The selected values were determined by the length of the applied RS codes, and the chosen dimension of equidistant quantization.

To evaluate the quality of training, on the basis of latent layer data *y*, such values as FRR, FAR, GAR, ERR and DI were calculated, the values of MSE for controlling the intra-class and inter-class distance distribution and the ROC-characteristic were monitored. The results after unsupervised learning and then supervised tuning of SAE with parameters 127/63 for 40 users, in the form of histograms, are shown in Fig. 2. It can be seen from the figure that due to the fine-tuning procedure, the interclass distributions expanded significantly relative to each other.

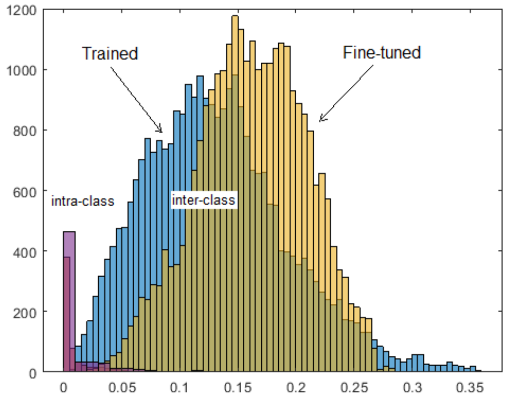


Fig. 2. Learning results of the SAE with parameters 127/63

Then several experiments were carried out to train SAE for 5 different groups of 40 users each, randomly selected from the UvA-NEMO database. For this compiled dataset the encoding-decoding procedures were modeled using the above-mentioned ECC. For processed 200 subjects from the dataset, the real-valued data with lengths of 63 and 31 elements, obtained from SAE and equidistantly quantized, were encoded with non-binary RS codes concatenated with linear codes.

Schemes of ECC used in modeling included: 1). RS code (63,15) concatenated with linear code (6,3,1) and separately with repetition (REP) codes (3,1,1); 2). RS code (31,9) concatenated with REP codes (3,1,1) and (5,2,2) separately. The conversion of symbols to bits was done on the basis of Gray's code representation [10]. For example, using the RS code RS (63,15) with Linear (6,3,1), we can obtain the FRR=1.0% and get the user key *R* with the length of 135 bits from the SAE data extracted from 3 video frames. And if you apply the RS (31,9) code together with REP (3,1,1), then the FRR will decrease to 0.5% for a key *R* with the length of 120 bits, processing 8 frames of user video. Thus, there is a trade-off between a decrease in FRR and an increase in the number of processed video frames. The experimental results showed that the efficiency of the proposed BS based on a neural network and non-binary codes is much higher compared to [11].

The hard-decision decoding has been applied for both type of ECC. An attempt to use the error-reducing technique [12] based on the properties of HD1 together with ECC failed to achieve adequate improvement due to the rather large variance of quantized data. However, in subsequent studies we are considering the possibility of using all types of HD and applying soft decoding of RS codes.

V. CONCLUSIONS

In this article, we examined the principles of implementing the Biometric System based on the use of facial smile imprints. Video frames from the main smile phases (onset, apex and offset) were used to obtain biometric data, on the basis of SAE, which consisted of two inner layers, and was trained with the use of smile videos of 400 subjects, taken randomly from UvA-NEMO database. The real data of the output layer was quantized and encoded by the concatenated ECC based on RS codes (63,15) and (31,9) with a redundancy of more than 70%, which affects the entropy loss or leakage rate [12]. The simulation experiments to assess the system performance showed the possibility of achieving FRR values of no more than 0.5% and 1% for crypto keys

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

of size 120-135 bits for biometric feature data dimensions of 31, 63 elements.

In our setup, we have used rather simple neural network (NN) structures and concatenated ECC based on non-binary RS codes which made it possible for us to obtain the FRR of less than 1% for the proposed biometric cryptosystem.

The direction of further work can be both the study of other NN structures to obtain deep features of a smile facial imprint, and the improvement of the technique of using ECC based on soft-decision decoding and side information from quantized data.

ACKNOWLEDGMENTS

This material is based upon work partially supported by the COST Action CA16101. The authors would like to thank Dr. Hamdi Dibeklioğlu for providing access to the data base UvA-NEMO.

REFERENCES

[1] Smileid. New Standard for Face Biometrics. Available from: https://www.electronicid.eu/en/solutions/smileid, last accessed 2021/05/05.

[2] S. Cook, Selfie banking: is it a reality? Biometric Technology Today, 9–11, 2017.

[3] M. Taskiran, et al. Face Recognition Using Dynamic Features Extracted from Smile Videos. In: IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA), Sofia, 3-5 July 2019, pp. 1–6.

[4] H. Dibeklioglu, A. A. Salah, T. Gevers, Are you really smiling at me? Spontaneous versus posed enjoyment smiles. In: European Conference Computer Vision, Florence, 7-13 Oct. 2012, pp. 525–538.

[5] K. Siwek, S. Osowski, Autoencoder Versus PCA in Face Recognition. 18th Intern. Conf. on Computational Problems of Electrical Engineering (CPEE). Kutna Hora, 11-13 September 2017, pp. 1–4.

[6] A. Juels, M. Wattenberg, A fuzzy commitment scheme. ACM Conf. on Computer and Communications Security, Singapore, November 1999, pp. 28–36.

[7] B. Assanovich, AutoEncoders for Denoising and Classication Applications. In: Open Semantic Tech. for Intel. Sys. (OSTIS), 4, 2020, pp. 309-312.

[8] A. J. H. Vinck, Coding Concepts and Reed-Solomon Codes. Institute for Experimental Mathematics, Essen. 2013.

[9] B. Assanovich et all. Recognition of Genuine Smile as a Factor of Happiness and its Application to Measure the Quality of Customer Retail Services. Proc. 14th Intern. Conf. Pattern Recognition and Inf. Proc. (PRIP'2019), Minsk, 21-23 May 2019, pp. 84-89.

[10] K. D. Rao, Channel Coding Techniques for Wireless Communications, Springer India, 2015, 394 p.

[11] B. Assanovich, Yu. Veretilo. Biometric Database Based on HOG Structures and BCH Codes. Proc. Information Techn. and Syst. 2017 (ITS2017). Minsk, 2017, pp. 286-287.

[12] V. Immler, K. Uppund. New Insights to Key Derivation for Tamper-Evident Physical Unclonable Functions. IACR Trans. Cryptogr. Hardw. Embed. Syst. 3, 2019, pp. 30–65.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Formalization of People and Crowd Detection and Tracking in Video

Rykhard Bohush
Computer systems and
Networks Department
Polotsk State University
Novopolotsk, Belarus
ORCID 0000-0002-6609-5810

Huafeng Chen
College of Information Science
and Technology
Zhejiang Shuren University
Hangzhou, China
ORCID 0000-0003-4229-4505

Sergey Ablameyko
United Institute of Informatics
Problems of NAS of Belarus
Belarusian State University
Minsk, Belarus
ORCID 0000-0001-9404-1206

***Abstract.* One of the promising areas of development and implementation of artificial intelligence is the automatic detection and tracking of moving objects in video sequence. The paper presents a formalization of the detection and tracking of people and crowd in video. The approach for tracking multiple people on video sequences for indoor and outdoor is described. The results of experiments for video sequences obtained using a stationary and moving video camera are presented.**

***Keywords:* video surveillance, moving object, convolutional neural network, tracking by detection, motion trajectory**

I. INTRODUCTION

Detection and tracking of objects on video sequences are one of the main tasks in computer vision, which currently have a different number of technical applications and will be increasingly used for: analyzing the environment in automated systems of driving vehicles; assessing the movement of people in medicine and sports; tracking objects in industrial vision systems; recognizing the type of human activity in monitoring and security systems [1]. Unlike images, video sequences contain a much larger amount of information, which changes both in space and time. Therefore, processing and analyzing them allows to identify not only static, but also dynamic features of objects, which leads to an increase in the effectiveness of automated operation of video surveillance systems as a whole.

There are many object detection and tracking algorithms have been developed. In [2], an algorithm for tracking people on video based on the Monte Carlo method for Markov chains is proposed, in [3] – an algorithm for detecting and tracking people in intelligent emergency detectors based on the support vector machine method. Currently, algorithms based on convolutional neural networks, which are resistant to changes in illumination, dynamic background, and allow detection even in the case of significant overlaps of objects, are widely developed and used for object detection [4, 5].

Tracking a group of people is one of the most urgent tasks for video surveillance systems, but at present it is not fully solved. There are a number of approaches to solve this problem, however, due to these problems, the effectiveness of their work is insufficient. The stage of forming an effective set of features that will be used to detect and track objects in the video sequence is one of the most difficult, since there are restrictions for it: use features that can be obtained in advance to describe objects; determine a limited set of features that will allow to get the maximum effectiveness, i.e. it is necessary to exclude uninformative features; it is possible to apply algorithms that meet the computational requirements of the developed systems. We have developed a number of solutions that are designed to detect and track dynamic objects in video sequences [6], people [7], smoke and flames [8]. Accordingly, we can say that the set of features used is largely related to the detection and tracking algorithms used to solve the tasks set. Therefore, in order to develop effective methods, techniques, and algorithms for detecting and tracking objects on video sequences, it is necessary to clearly formalize these tasks. It is necessary to determine the objects that will be detected and tracked, to determine the main stages of this process, as well as criteria that allow to evaluate the quality of processing and show how this is implemented in practice.

In this paper, we propose a formalization of the tasks of detecting and tracking objects on video sequences. On the basis of the considered generalization, an algorithm for tracking a set of people and an algorithm for tracking crowd are developed. The results of experiments on the basis of the considered criteria, allowing to evaluate the quality of the algorithms, are presented.

II. FORMALIZATION OF PERSON DETECTION PROBLEM

A video sequence or video stream is a sequence of digital images (frames) $V = \{F_k\}$, k - the number of the image in the sequence. The object in the image (Ob) is a local area that differs from the surrounding

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

background and displays some of the features of the real-world object.

On each frame of the sequence obtained from a stationary video camera, as a rule, there are many objects: $OB_{F_k} = \{Ob_q^{F_k}\}, q = 1, \dots, Q$. According to the criterion of movement, each of them can be assigned to two main classes:

A stationary (stationary) object in a sequence of images is described by a set of features $(Ft_{Ob_q}^S)$ and its coordinates $\left(x_{Ob_q}, y_{Ob_q}\right)$, which do not change during a time interval (t). Such an object can be represented by a formal model: $Ob_q^S = \left(Ft_{Ob_q}^S, x_{Ob_q}, y_{Ob_q}, Ns_{Ob_q}^{F_k}\right)$, where $\left(Ft_{Ob_q}, x_{Ob_q}, y_{Ob_q}\right) = const \forall F_k, k \in t$, $Ns_{Ob_q}^{F_k}$ – the set of possible noise effects on the object.

A moving object in a sequence of images is characterized by a change in one or more basic parameters: shape, size, and coordinates over a time interval (t). The transformation of the shape and / or size of an object leads to a change in its features in the frames $\left(ft_{Ob_q}^{F_k}\right)$. Such an object can be represented by a formal model: $Ob_q^D = \left(ft_{Ob_q}^{F_k}, x_{Ob_q}^{F_k}, y_{Ob_q}^{F_k}, Ns_{Ob_q}^{F_k}\right)$, where $x_{Ob_q}^{F_k}, y_{Ob_q}^{F_k}$ - object coordinates; $Ft_{Ob_q}^D$ – a set of features of moving object, $Ft_{Ob_q}^D \supseteq ft_{Ob_q}^{F_k}, \forall k \in t$. Then $ft_{Ob_q}^{F_k} \cap$ $\cap\, ft_{Ob_q}^{F_{k+i}}$, that is, for the same moving object on a sequence of frames, a change in its features is characteristic.

Object detection is the determination of the location of a given object Ob^e in the image F, while its size is smaller than the size of the image, and the number of objects in the image is obviously unknown. In general, the object Ob^e detection process is implemented by comparing the features of the reference (Ft^e) and all possible fragments on the image plane using the rule-based method:

$$S\left(Ft^e, Ft_{Ob_q}^F\right) \overset{M,Z}{-} \max,$$

where S - detection accuracy; Z - a set of restrictions.

When detecting a stationary object on a sequence of images, it is necessary to consider the variability of a dynamic scene, since in addition to static ones, there are moving objects on it, and their number can change. Objects (Ob_q^D) can overlap a stationary object, which will cause its features to change on the sequence of frames, i.e.: $Ft_{Ob_q}^{S,F_k} \neq Ft_{Ob_q}^{S,F_{k+1}}$. Therefore, to detect a stationary object on a sequence of images displaying a dynamic scene, it is necessary to use a method M_{STV}, that takes into account the change in features over time:

$$S\left(Ft^e, Ft_{Ob_q}^{S,F_k}\right) \overset{M_{STV},Z_{STV}}{-} \max,$$

where Z_{STV} - a set of constraints when detecting a stationary object on a sequence of images.

Detection of moving object means to determine a location of the object Ob^D on the current frame F_k of the video sequence based on the specified F_e:

$$SD\left(Ob_{F_k}^D, Ob_{F_e}^D\right) \overset{MD,ZD}{-} \max,$$

where SD - the accuracy of detecting a moving object; MD - the method used; ZD - a set of restrictions.

III. FORMALIZATION OF PERSON AND CROWD TRACKING PROBLEM

Object tracking can be divided into three types:

- tracking of a single object (Visual object tracking, VOT);

- tracking of multiple objects (Multiple object tracking, MOT);

- tracking of crowds.

The first case of tracking is characterized by the fact that the object is detected and localized in the first frame, other objects are not detected.

A. Single person tracking

Tracking a moving object - determining the location of the same object on each frame of the video sequence during a time interval. This makes it possible to plot the trajectory of an object, determine its speed and acceleration. When solving practical problems, in some cases, an analysis of the trajectory of movement is required.

To perform maintenance, detection and localization procedures are required. Different ways of describing the objects of observation are used:

- a single point that characterizes the center of mass of the object or the center of the minimum possible rectangle described around the object;

- a set of key points by which the object can be uniquely identified in subsequent frames;

- a geometric primitive described around the object (most often a rectangle, less often an ellipse);

- the external contour of the object;

- a set of areas that are as stable as possible when moving, or the entire area of the object;

- invariant characteristics of the object (for example, texture, color scheme, etc.).

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The trajectory of the object is a sequential display of the movement of this object on the video sequence: $Tr(Ob^D) = \left(Ob^D_{F_k}\right), \forall k \in t$.

In physics, the trajectory of motion is called the line that a particle describes when it moves. It is obvious that for the problem to be solved, in the end, it is necessary to determine a line on the required frame of the video sequence, which will show how the observed object moved over a certain period of time.

There may be different ways to determine the coordinates of an object on the frame, but the most commonly used approach involves finding its center (one pixel per frame) with coordinates $\left(x^{F_k}_{Ob_q}, y^{F_k}_{Ob_q}\right)$. As a rule, the movement is considered in the frame coordinate system. Then the trajectory of the object on the video sequence is described by a sequence in the form of a set of coordinates of the center of the object on each frame:

$$Tr(Ob^D) = \left(Ob^D_{F_k}\right) = \left(\left(x^{F_1}_{Ob_q}, y^{F_1}_{Ob_q}\right), \cdots, \left(x^{F_n}_{Ob_q}, y^{F_n}_{Ob_q}\right)\right).$$

The trajectory Tr' of its movement can be found using method MTS and constraints ZTS:

$$STS(Tr') \xrightarrow{MTS,ZTS} \max,$$

where STS - the accuracy of tracking a moving single object.

Despite the development of numerous algorithms over the past decade, due to the possibility of significant visual changes in the object and illumination, background noise, occlusions, the task of tracking a single object (VOT) is not fully completed.

B. *Multiple person tracking*

When tracking multiple objects, one need to determine the set of trajectories $TR' = \{Tr'_q\}$ of the objects in the frames, and then compare them to each other to determine the movement of all objects between frames:

$$STM(TR, TR') \xrightarrow{MTM,ZTM} \max.$$

Several objects can be present in the frame at the same time. Moreover, objects can have almost identical visual features. Thus, it is possible to lose an object that occurs due to its intersection with a similar one, or overlap with a background element. Tracking of multiple objects is performed at long time intervals, and it is possible to predict the location on subsequent frames.

C. *Crowd tracking*

Crowd is a large group of people with severe occlusions. Crowd tracking is to monitor state changing of crowds. Usually abnormally sudden change indicates emergency, for example, a crow splits into small groups could mean people run away from danger, two large crowds merge into one along with intensive motion could mean clash.

There may exist several crowds in one frame: $CR_{F_k} = \{Cr_q^{F_k}\}, q = 1, \dots, Q,$ where $Cr_q^{F_k} = \{Ob_{q_i}^{F_k}\}, i = 1, \dots, n_q$, n_q is the number of people that compose crowd $Cr_q^{F_k}$. One thing deserves to be mentioned is that crowd does not keep its composition through time, it can split, or join together with other crowds.

Because of severe occlusions, single person $Ob_{q_i}^{F_k}$ in the crowd can hardly be detected or tracked. One common way is to treat crowd as a single entity and consider imaginary particles occupy the crowd area. Along with particles moving, crowd will reshape or regroup. It is possible to track for one crowd in a certain frame where its sub-groups go in next frames ($Cr_q^{F_k} = \{Sb_{q_i}^{F_{k+t}}\}, q_i = 1, \dots, l_q$, where $Sb_{q_i}^{F_{k+t}}$ is a separated sub-group of $Cr_q^{F_k}$ in F_{k+t}), or where its sub-groups came from ($Cr_q^{F_k} = \left\{Sb_{q_j}^{F_{k-t}}\right\}, q_j = 1, \dots, m_q$, where $Sb_{q_j}^{F_{k-t}}$ is a separated sub-group of $Cr_q^{F_k}$ in F_{k-t}).

Once sub-groups of a crowd is located in a previous frame or a posterior frame, further analysis of the crowd can be performed to determine whether certain crowd behavior happens. Three types of behavior are considered:

- crowd directional movement: many people move in the same direction;

- crowd aggregation: many people move toward a certain region from different directions;

- crowd dispersion: many people move away from a certain region in different directions.

One of the main problems for the practical use of the tracking algorithm is to ensure high accuracy with limited hardware resources and input data. In general, accurate tracking can be achieved by solving the global optimization problem, which requires the entire sequence of frames at once, which is impossible in real video surveillance systems. In the existing methods, the object tracking problem is often formulated as an optimization problem using graph algorithms [9]. Each detected object is represented as a vertex, and the transition from one vertex to another is determined by the similarity function used. Establishing an association on graphs can be solved by the method of finding the path with the minimum cost, which is most effectively solved by global optimization. Tracking algorithms based on the selection and analysis of special points require the presence of corners in the image contours. With a small number of them, the effectiveness of tracking will be low. The use of colour characteristics of objects for tracking is considered in the method from [10]. Probabilistic approaches use the statement that a moving object has a certain state, which is measured on each frame, and to estimate its position on the next one, it is necessary to generalize the values from the previous ones. For this purpose, methods based on the Kalman filter [11] or the particle filter [12] are used. However, objects can have a pronounced nonlinear trajectory of movement, and in this case, the assessment of the new state based on the previous ones will be determined with a high error. Therefore, different approaches are used to solve different applied problems.

IV. EXPERIMENTAL RESULTS

The proposed mathematical background have been tested in many applied tasks that have been described in our papers [6-8]. Tracking by detection method is effective for people tracking. In this case, the detection stage is one of the key ones. The quality of its work largely determines the accuracy of people tracking in video. Therefore, for what follows, we will use a more accurate CNN YOLOv4, the advantages of which are indicated in [13]. After person detection in the frame, features of the selected fragment in the spatial area in the frame and in time domain in video sequence are calculated. We use features such as CNN and histogram features of H channel in HSV space when this person was last correctly detected in frame, centre coordinates for selected area of a person in frame, displacement in the current frame relative to the previous one, width and height of the area in previous frame, motion trajectory, motion time. The values of the similarity function are calculated for all accompanied and detected people in the current frame. Based on these values, a correspondence is established between the detected and tracked objects using Hungarian algorithm. The trajectory is created when a person is first detected. The trajectory is deleted if this person is not detected for a certain number of consecutive frames and there is no comparison for him with previous frames. thus, in this case, we consider that he left the scene, which is recorded by a video camera.

An important task is to determine the effectiveness of tracking, taking into account the joint work of the detection and tracking stages to assess the possibility of practical use of the algorithm. Therefore, testing of the tracking algorithm for indoor surveillance and a modified algorithm for outdoor, which uses the Kalman filter, was carried out taking into account the results of human detection by CNN YOLOv4. In this case, errors in the operation of this CNN lead to a deterioration in the criteria for tracking. However, experiments reflect the real effectiveness for tracking by detection algorithms, which is very important for making a decision about their application in practice.

For evaluation of proposed approach MOT16 metric [14] is used. Experiments for indoor video sequences were carried out on six videos from a stationary surveillance camera. The total number of frames is 11890. Video are characterized by change in lighting, a nonlinear trajectory for people, overlap by background objects or the intersection of people trajectories, similar characteristics. The experiments performed have confirmed that our proposed approach improves the tracking accuracy on test video sequences obtained indoors and outdoors. To implement the algorithm from [15], MOTA = 0.288306 is provided for all videos from [14], and for our proposed MOTA = 0.300860. An improvement is also provided for video sequences from a fixed indoor surveillance camera: for the algorithm from [15] MOTA = 0.8793, for the proposed algorithm MOTA = 0.9266.

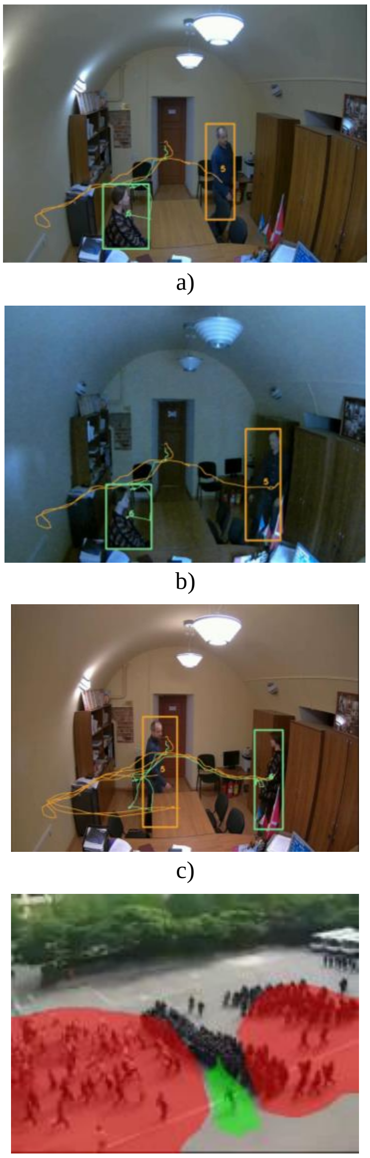

a)
b)
c)
d)

Fig. 1. Results of mutiple people tracking and crowd behavior detection

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

For crowd tracking, algorithm based on integral optical flow and motion maps [16] is used. The key step of this algorithm is to analyze geometric structure formed by movements of crowds and their sub-groups, which involves considering speed and incline of the movements and density of crowds. Based on geometric structure analysis, certain crowd behaviors can be detected.

Fig. 1 show examples of detecting and tracking people and crowd. Example for indoor video surveillance with a changing lighting level and building trajectories of people movement is shown in Fig. 1a-c. Fig. 1d shows crowd behavior detection during a riot control exercise. In Fig. 1d, both red areas show crowd directional movements are happening. Furthermore, these two moving crowds from the left and the right are about to meet in the middle, the core area of the collision zone is painted green.

V. CONCLUSION

The paper presents a formalization of the problem of detecting stationary and moving objects on video sequences, taking into account their features. Three cases of tracking are considered and described: single object, multiple objects and crowds on video sequences. The approach for tracking multiple people on video sequences for indoor and outdoor video surveillance is described. The first stage requires detecting person in the input frames by YOLOv4 convolutional neural network. For solving assignment problem of person, we store information about individual object in spatial domain of frames and in the time domain on a video sequence. For person description, the following feature set is used: neural network and histogram features, centre coordinates of a person in the frame, offset in the current frame relative to the previous one, person width and height in the previous frame, trajectory and time of movement.

This research was partly supported under BRFFR project F21UKR-001.

REFERENCES

[1] Cavallaro and E. Maggio, Video tracking: theory and practice. Wiley, 2011.

[2] D. Kuplyakov, E. Shalnov and A. Konushin, “Markov chain Monte Carlo based video tracking algorithm,” Programming and Computer Software, vol. 43, pp. 224-229, July 2017. doi: 10.1134/s0361768817040053.

[3] A. P. Kirpichnikov, S. A. Lyasheva and M. P. Shleymovich, “Detection and tracking of people in intelligent detectors ofemergency situations,” Kazan Technological University Bulletin, vol. 17, pp.351-356, 2014. (in Russian).

[4] Mohana and HV Ravish Aradhya, “Object Detection and Tracking using Deep Learning and Artificial Intelligence for Video Surveillance Applications,” Int. Journal of Advanced Computer Science and Applications, vol. 10, pp. 517-530, 2019. doi: 10.14569/IJACSA.2019.0101269.

[5] S. Mane and S. Mangale, “Moving Object Detection and Tracking Using Convolutional Neural Networks,” 2018 Second Int. Conf. on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 1809-1813, 2018. doi: 10.1109/ICCONS.2018.8662921.

[6] R. Bogush, S.Maltsev, A. Kastryuk, N. Brovko and D. Gluhau, “Motion Detection and Tracking Algorithms in Video Streams,” VNU Journal of Science, Mathematics - Physics, vol. 25, pp. 143-151, 2009.

[7] Bogush R. P., Zakharova I. Yu. and Ablameyko S. V., “Algorithm for person tracking on video sequences using face identification for indoor surveillance,” Herald of computer and information technologies, pp. 3-14, 2020. doi: 10.14489/vkit.2020.07.pp.003-014.

[8] S. Ye, B. Zhican, C. Chen, R. Bohush and S. Ablameyko, “An Effective Algorithm to Detect Both Smoke and Flame Using Color and Wavelet Analysis,” Pattern Recognition and Image Analysis, 2017, vol. 27, pp. 131-138. doi:10.1134/S1054661817010138.

[9] B.A. Zalesky and A.I. Kravchonok, “Tracking dynamical objects and their recognition by graph algorithms,” Informatics, vol. 2, pp. 7-26, 2006.

[10] B. A. Zalesky, “Object tracking algorithm by moving video camera,” Doklady of the National Academy of Sciences of Belarus, vol. 64, pp. 144-149, 2020. doi: 10.29235/1561-8323-2020-64-2-144-149.

[11] V.Yu. Agafonov, V.L. Rozaliev and A.V. Zaboleeva-Zotova, “Using the Kalman filter in object tracking tasks,” Intelligent systems. Theory and applications, vol. 20, pp. 13-17, 2016.

[12] F. Gustafsson, F. Gunnarsson and N. Bergman, “Particle Filters for Positioning, Navigation and Tracking,” IEEE Transactions on Signal Processing, vol. 50, pp. 425-437, Feb. 2002. doi: 10.1109/78.978396.

[13] A. Bochkovskiy, Ch.-Y. Wang and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” https://arxiv.org/abs/2004.10934. last accessed June 12, 2020.

[14] MOTChallenge: The Multiple Object Tracking Benchmark, https://motchallenge.net, last accessed April 14, 2021.

[15] Real-time Multi-person tracker using YOLO v3 and deep_sort with tensorflow, https://github.com/Qidian213/deep_sort_yolov3,last accessed Feb. 22, 2021.

[16] H. Chen, S. Ye, O. Nedzvedz, S. Ablameyko and Zhican Bai, “Motion Maps and Their Applications for Dynamic Object Monitoring,” Pattern Recognition and Image Analysis, vol. 29, pp. 131-143, April 2019.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

UNetX: Real-time Pedestrian Crosswalk Segmentation on Mobile Device

Eduard Adaska
Computer Systems Department
Belarusian State University
of Informatics and Radioelectronics
Minsk, Belarus

Anton Lechanka
Computer Systems Department
Belarusian State University
of Informatics and Radioelectronics
Minsk, Belarus
lechenko@ bsuir.by

Abstract. This paper presents a lightweight deep neural network that segments pedestrian crosswalks on an image in realtime. It is based on U-Net architecture with all its convolution layers substituted with depthwise separable convolution ones. This neural network was trained and tested against a set of manually segmented 3083 road images — with and without crosswalks. The resulting network has only 383K parameters and runs at 35 FPS on a mobile phone. The Jaccard index (IoU metric) on the validation set is 0.9138.

Keywords: **autonomous car, pedestrian crosswalk, image segmentation, deep neural network, U-Net, depthwise separable convolution**

I. INTRODUCTION

There have been considerable research interests in autonomous cars in the past few years, primarily concerning the safety of autonomous vehicles. Meanwhile, Every year, approximately 300,000 pedestrians die on the roads accounting for up to 26% of all deaths in road accidents [1]. That makes pedestrian crosswalk detection an essential task for the safety of autonomous cars.

The main three approaches for crosswalk detection are based only on camera images [2, 3], only on LIDAR point clouds [4] and on both images and point clouds [5, 6]. In this text, we present a convolutional neural network, UNetX, based entirely on camera image processing. The neural network can be used as a part of a complex autonomous vehicle system. Nevertheless, being a lightweight solution that can run in realtime on a mobile phone, it also can operate as a virtual offline road assistant for drivers.

This network solves the detection problem as image segmentation. It partitions an image into two segments: the one that belongs to a crosswalk and the one that does not. In this paper, we present UNetX, a five-time more compact modification of U-Net [7], which makes UNetX faster and more power-efficient.

II. METHODOLOGY

A. *Data set selection and annotation*

The data set of 35625 images for crosswalk detection [8] have been taken for that problem. That dataset effectively comprises three data subsets. The first data subset consists of images captured using a Bumblebee XB3 camera mounted on the IARA vehicle during the day. The second one was created using the same instrument at night. And the third subset was taken with GoPro HERO 3 camera in Full HD.

The set was annotated only with crosswalk presence property; thus, additional annotation was required. A portion of those images has been manually annotated using VGG Image Annotator [9]. JSON output from VGG Image Annotator was converted to a bit-mask representing whether the pixel belongs to a crosswalk. Several local streets images have also been added. Those images have been taken during both day and night and also during bad weather. The images have been evenly selected from all data subsets. In total, 3083 images have been annotated, 975 of which did not contain any crosswalk.

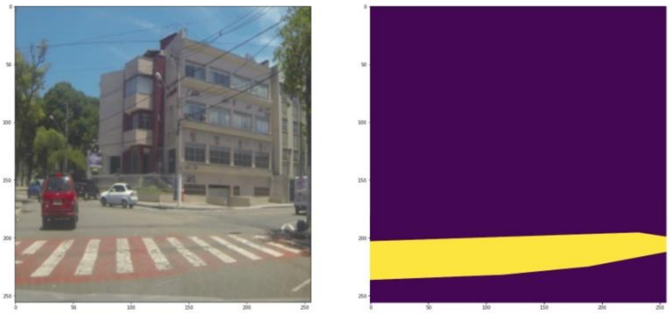


Fig. 1. Example of an image on the left and its annotation mask on the right

A benchmark application was developed to measure realtime performance on a mobile device. It is an iOS application that runs the CNN model using

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

CoreML framework [10]. It feeds a batch of images to a model, measures their processing time, and calculates frames per second (FPS). An iPhone 8 was used to measure real-time performance on an actual device.

B. *CNN architecture*

The U-Net was used as a base convolutional network with slight changes. Unlike the original paper, in this work, 256 × 256 image with 3 color channels is used as an input for UNet CNN. An encoder transforms an input image to a tensor 64×16×16, and a decoder with concatenation layers unpack this tensor to a single channel 256 × 256 mask. Each value of a mask is a probability that the corresponding pixel belongs to a pedestrian crosswalk on an input image.

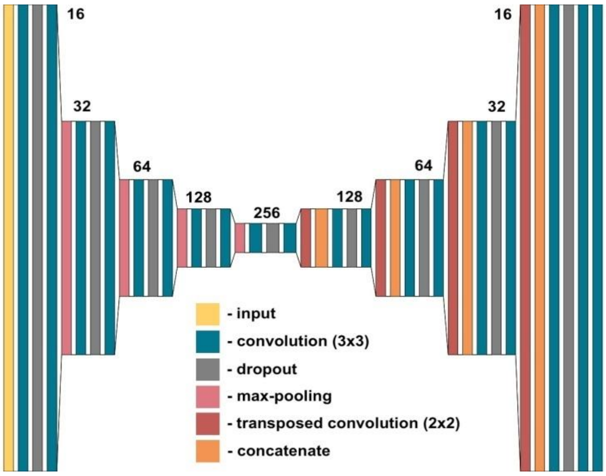


Fig. 2. U-Net CNN visualization

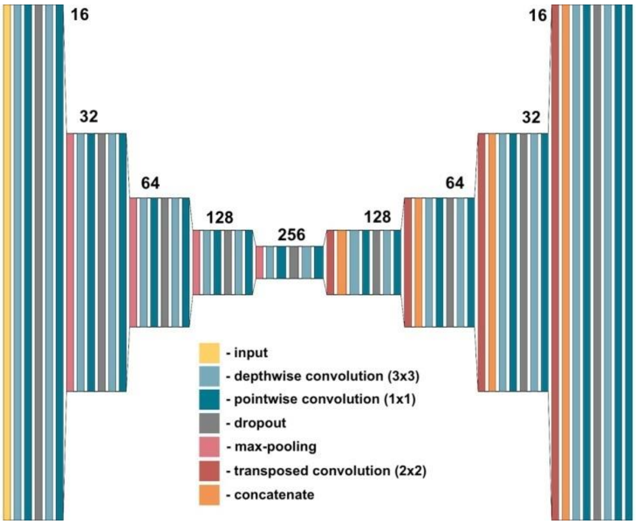


Fig. 3. UNetX CNN visualization

The U-Net network has only 1,941,105 parameters, but even a smaller network has been proposed. Convolution layers in U-Net have been substituted with depthwise separable convolution layers [11]. The other components of a CNN, like max-pooling and concatenation layers, are unchanged.

Depthwise separable convolution is essentially two convolutions. The first one is a depthwise convolution. It takes N×M×C tensor and performs C 2D convolutions with its filter on each N×M layer. The outputs of a depthwise convolution are stacked to a 3D tensor and fed to a point-wise convolution. This modification results in a five times smaller network with only 383K parameters.

C. *Figure of merit*

Intersection over Union (IoU) is used as an evaluation metric for the CNNs described. It is a commonly used performance metric for image segmentation problems [12].It measures the similarity between a predicted region and a ground truth region. The predicted region is defined by the output of a CNN with a probability higher than 0.5, and the ground truth is an annotation mask. In order to calculate IoU, one needs to find an intersection and a union of the predicted region and the ground truth. Thus, the IoU is a ratio of the area of intersection over the area of union.

Fig. 4. Intersection (yellow) over Union (red and yellow) visualization

III. TRAINING THE MODEL

Both UNet and UNet-X CNNs have been trained on NVIDIA Tesla V100 using Keras [13]. Each network was trained for 100 epochs; each epoch processed the whole training set divided into batches of 32. The optimization algorithm is Adam [14]. The learning rate and other training hyper-parameters were set to Keras defaults.

Binary cross-entropy [15] was used as a loss function during training, as in:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(p(y_i)) + (1-y_i)\log(1-p(y_i)) \quad (1)$$

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

N in (1) is the number of pixels, yiis the ground truth value whether a pixel number ibelongs to a pedestrian crosswalk segment, and p(yi) is a predicted probability that a given pixel belongs to the segment.

The data set has been randomly split into training and validation sets. The training set comprises 2159 images (70% of the annotated data set), and the validation set is 924 images. After each single training epoch is concluded, the trained network model is validated against the validation set. Fig. 5 shows model training loss and validation IoU after each epoch.

IV. RESULTS AND DISCUSSION

Substitution of convolution layers with depthwise convolutions insignificantly affects segmentation results. Values of the IoU metric given in a table I and the visual comparison of the segmentation masks, as shown in fig. 6, support this claim. At the same time, the number of trained network parameters decreased five times, from 1,941,105 to 386.543. This improved network performance on a mobile device by 33% up to 35 frames per second, which is enough to perform proper real-time segmentation.

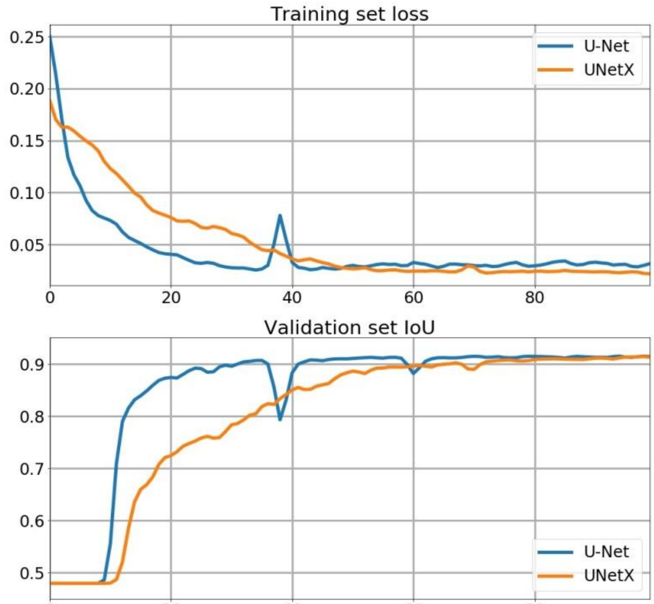


Fig. 5. Loss and IoU training progress

Despite the dramatic decrease in computation workload, the performance improvement is modest. This is due to the almost unchanged number of load/store operations per inference and the increased number of CNN layers. Performance results are summarized in Table I.

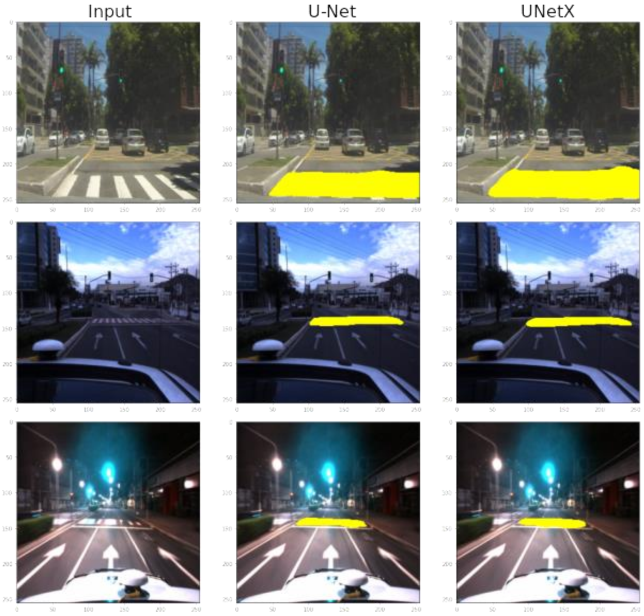


Fig. 6. Visual comparison

TABLE I. MODELS SUMMARY

Model	*# of parameters*	*IoU*	*FPS*
U-Net	1,941,105	0.9147	26.31
UNetX	386,543	0.9138	35.17

V. CONCLUSION

Encoder-Decoder CNN with depthwise separable convolutions can be used as a tool for real-time segmentation of pedestrian crosswalks. Moreover, according to the conducted experiment, switching to depthwise separable convolution has an insensible impact on segmentation quality. In conclusion, the proposed UNetX can be used as a part of a complex system of an autonomous car or as a virtual assistant for drivers, thanks to excellent real-time performance on mobile devices.

REFERENCES

[1] W. H. Organization et al., “Global status report on road safety 2018: summary,” World Health Organization, Tech. Rep., 2018.

[2] M. A. Malbog, “Mask r-cnn for pedestrian crosswalk detection and instance segmentation,” in 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS). IEEE, 2019, pp. 1–5.

[3] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool, “Pedestrian detection at 100 frames per second,” in 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012, pp. 2903–2910.

[4] H. Wang, B. Wang, B. Liu, X. Meng, and G. Yang, “Pedestrian recognition and tracking using 3d lidar for autonomous vehicle,” Robotics and Autonomous Systems, vol. 88, pp. 71–78, 2017.

[5] R. Guidolini, L. G. Scart, L. F. R. Jesus, V. B. Cardoso, C. Badue, and T. Oliveira-Santos, “Handling pedestrians in crosswalks using deep neural networks in the iara

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

autonomous car," in 2018 International Joint Conference on Neural Networks (IJCNN), 2018, pp. 1–8.

[6] J. Dou, J. Fang, T. Li, and J. Xue, "Boosting cnn-based pedestrian detection via 3d lidar fusion in autonomous driving," in International Conference on Image and Graphics. Springer, 2017, pp. 3–13.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[8] R. F. Berriel, F. S. Rossi, A. F. de Souza, and T. Oliveira-Santos, "Automatic large-scale data acquisition via crowdsourcing for crosswalk classification: A deep learning approach," Computers & Graphics, vol. 68, pp. 32–42, 2017.

[9] A. Dutta, A. Gupta, and A. Zissermann, "Vgg image annotator (via)," http://www. robots. ox. ac. uk/vgg/ software/via, 2016.

[10] M. Thakkar, Beginning machine learning in iOS: CoreML Framework. Apress, 2019.

[11] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251–1258.

[12] M. A. Rahman and Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation," in International symposium on visual computing. Springer, 2016, pp. 234–244.

[13] A. Gulli and S. Pal, Deep learning with Keras. Packt Publishing Ltd, 2017.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[15] S. Jadon, "A survey of loss functions for semantic segmentation," in 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE, 2020, pp. 1–7.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Developmental Milestones of Graphics Technologies

Dzmitry Mazouka
Department of Information Management Systems
Belarusian State University
Minsk, Belarus
mazovka@bk.ru

Viktor Krasnoproshin
Department of Information Management Systems
Belarusian State University
Minsk, Belarus
krasnoproshin@bsu.by

Abstract. **The complexity of modern graphics hardware and software has a long history. In this paper we review the historical milestones of computer graphics development. This analysis will help us to understand the common problems and trace a direction for the future improvements.**

Keywords: **computer graphics, graphics pipeline, directx**

I. INTRODUCTION

Computer graphics is a very distinct field in computer science. Unlike other areas, the primary concern of computer graphics lies in the presentation of information, rather than pure computation. This makes computer graphics universally applicable in a vast number of human activities, including science, engineering, manufacturing and entertainment. In a peculiar turn of events, computer graphics development and evolution has been primarily driven by the entertainment industry. Started late 20th century, video games have been at the forefront of computer graphics progress, which was later picked up and expanded by the movie industry. These days it is hard to find an artifact of the popular culture that was not produced using computers in one way or another.

Computer graphics technologies pursue two fundamental goals: photorealism and real-time imagery generation. Despite tremendous progress in the past decades, we are still far from being able to render 3D scenes of arbitrary complexity with perfect picture quality. Perhaps it is not an achievable goal after all, but the development will always be concerned with image quality and the speed of generation.

From general perspective, the technological stack of computer graphics consists of the following layers: graphics hardware, graphics Application Programming Interface (API), visualization system, and application.

Application layer defines a particular visualisation problem that needs to be solved with visualisation tools. It can be a video game, a visual effect in a movie, or a piece of engineering software. The requirements for application layer are defined externally and they often drastically differ from one problem, or product, to another. This is where the need for quality and speed arises, to propagate, eventually, to the lower layers. Due to the nature of application problems, little can be done for formalisation of this layer.

Visualisation system layer can be optional, however complex applications almost never get built from scratch. Many years of experience in various fields of computer graphics provided software engineers with enough information to construct systems that would be applicable to a wide array of visualisation problems. The best example of that would be a graphics engine, as a part of a game engine. A graphics engine (for example: Unreal Engine, Source, Frostbite) is a visualisation system that is developed and tuned to a specific subset of applications – game genres. Common visualisation problems are solved once in an efficient way in a graphics engine and that facilitates higher production speed for subsequent projects that a game development company may take. Another good example of a visualisation system can be Computer-Aided Design (CAD) systems (for example: AutoCAD, SketchUp, Archicad). These also solve common problems, but in the areas of engineering and architecture. The major difference of visualisation system layer with application layer, is that visualisation system plays the role of middleware, and each specific visualisation application problem has to evaluate and choose whether any existing visualisation system can help with the solution or not. None of the visualisation systems are applicable to all possible applications.

Graphics API layer is the most standardised layer in the technological stack. The API is normally represented by operating system drivers and specialised graphics libraries. Of which, the most prominent are: DirectX, OpenGL and Vulkan. These libraries do the mediation work between software and hardware within an operating system, which includes translation of application intent into graphics hardware commands, and control and execution of the rendering process. The development of the graphics libraries is substantially slower than visualisation systems, as they heavily depend on architecture of graphics hardware, and when the hardware updates, the libraries have to change accordingly. Graphics API represents a common

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

language that graphics hardware and software talk to each other, that is why there is so few implementations of it.

Graphics hardware is the base layer in the technological stack of computer graphics. It consists of various specialised hardware components that perform rendering. Most notable form of graphics hardware is graphics cards (for example: NVIDIA GeForce, AMD Radeon). A graphics card contains a dedicated Graphics Processing Unit (GPU), which runs a special algorithm called graphics pipeline. In a similar manner with visualisation systems, graphics hardware accumulates common patterns and solutions to common visualisation tasks coming from applications. And the changes then propagate via graphics API up, towards visualisation systems and applications.

Despite all the successes that computer graphics has enjoyed so far, there is one problem that is growing with each new feature and every new improvement: the complexity of the technological stack. In this paper we will analyse the developmental milestones of the technological stack, highlight the changes and the reasoning behind them. This will help with understanding of the technology in its current state and possibly give ideas for further development.

II. METHOD OF ANALYSIS

In this paper we will use a model for historical analysis that is based on the technological stack described earlier. We will look at three separate actors: graphics hardware, graphics API, and applications which is a combination of application and visualisation system layers from the technological stack.

Graphics hardware, graphics API and applications relate to each other in the manner depicted in Fig. 1. The relationship between the actors is cyclical and flows in one direction, with arrows pointing at the actors the development of which is informed by the source of the arrow. Graphics hardware development is informed by the nature and requirements of the applications, graphics API reflects the structure and capabilities of the graphics hardware, and applications can only do as much as it is possible with a certain API.

Of course, there is also external input into this system, for instance, graphics hardware is not being developed in a complete isolation and depends on the current state of art in chip manufacturing and related improvements in technologies. Applications are also influenced by the ideas of business products, or advances in a scientific or engineering thought. Graphics API may be influenced by competing technologies or general evolution in software development techniques. But for our purpose, a simpler model will be enough.

With every turn of the relationship cycle, the components change and improve. We identified DirectX API [1] versioning as a representative timeline of the stepping stones that graphics technologies have taken on the way, and in the next section we will begin with the first version.

III. ANALYSIS

A. *DirectX 1 (GameSDK) – 1995*

First GPUs were pretty simple, and were mostly concerned with 2D image processing and displaying. This was already a good starting point, since it introduced a separation of duties within a single machine: general CPU did not have to bother with graphical tasks consisting mostly in copying large buffers of data from one memory location to another.

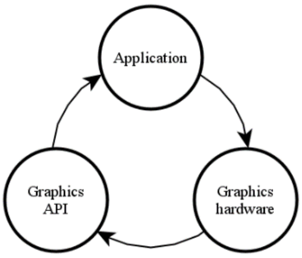


Fig. 1. Actors' relationship cycle

In a similar vein, DirectX 1 only contained a library called DirectDraw, dedicated specifically to the work with 2D graphics. This library unified and abstracted the work with video memory, so that the users would not need to bother about the kind of hardware their applications were running on. This technique was called Hardware Abstraction Layer (HAL).

If application required 3D capabilities, it was mostly on its own, all geometry preparation, including transformation and projection had to be performed by the CPU, and then passed to the graphics hardware via API. But even at this early stage, DirectDraw provided access to double buffering and Z-buffer support, which are essential features to this day.

B. *DirectX 2 – 1996*

Voodoo 3dfx graphics acceleration card [2] was a major step in development of what has become known as graphics pipeline. The card implemented the rasterisation algorithm and relied on the presence of another video card in the system for 2D output. Rasterisation was another essential step on the way to the true 3D rendering, but transformation and projection still had to be done on the CPU.

DirectX 2 had extensively improved capabilities of DirectDraw, and introduced a library to work with 3D graphics: Direct3D. Direct3D in DirectX 2 could be used in two separate modes with distinct APIs: Immediate

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Mode (IM) and Retained Mode (RM). Retained Mode was designed for high-level graphics programming, and included functionality and primitives for construction and management of 3D scenes containing hierarchies of objects. RM contained methods for camera manipulation and animations and was built on top of Immediate Mode. Immediate Mode on the other hand, was a low-level programming interface and required good understanding of the intricacies of graphics programming. This demonstrates that graphics development complexity was recognised even at these early stages.

Even though graphics hardware did not support lighting, transformation and projection at the time, DirectX implemented the missing parts of the rasterisation algorithm in software, and provided specialised components in the form of Transform, Lighting and Raster modules.

C. DirectX 3 – 1996

DirectX 3 did not have any improvements to DirectDraw or Direct3D, and only updated other components of the SDK, such as DirectSound, DirectInput and DirectSetup.

It should also be noted that DirectX version 4 was skipped and the next released version became DirectX 5.

D. DirectX 5 – 1997

In 1997 graphics cards manufacturers introduced a new data transfer interface in their products: Accelerated Graphics Port (AGP) [3]. This data transfer bus significantly increased the rate at which graphics data could be passed to graphics hardware. That, in turn, helped to increase the size of 3D scenes and quality of textures. In addition to this, graphics hardware added support of multitexturing – an ability to use multiple texture maps during single surface rasterisation. This feature had opened a way to many visualisation techniques, such as bump mapping, specular mapping, prebaked lighting and so on.

When it was introduced, Direct3D Immediate Mode required programmers to record the instructions that they wanted to pass to graphics hardware in a special data structure called Execute Buffer. Execute Buffers were pretty low level and required a fair amount of boilerplate code when operated. For this reason, DirectX 5 added a more convenient set of instructions in addition to Execute Buffers: Draw Primitive commands. Direct3D Retained Mode was expanded with a number of interfaces for animation and managing geometry with variable levels of details.

With the new improvements of the hardware and API, applications could render scenes with textured materials more efficiently and the selection of available visual effects have increased.

E. DirectX 6 – 1998

In DirectX 6, DirectDraw did not have any significant changes, but it increased the number of methods that simplified working with graphics hardware. Direct3D Immediate Mode improved its performance and added support for new hardware features: single-pass multiple texture blending, texture cache management, vertex buffers, and many others. Direct3D Retained Mode was incrementally improved without any noteworthy changes.

Graphics application were provided with a better selection of tools as the result of these changes; however, the more basic capabilities were still quite primitive, for instance, lighting and transformation was still performed by CPU, and this prompted the next big challenge for the hardware.

F. DirectX 7 – 1999

In 1999 graphics cards got a new module that extended the capabilities of the hardware graphics pipeline: a Transformation and Lighting module (T&L) [4]. This change allowed to remove from the CPU the need of performing geometry transformation (projection from 3D coordinate system onto 2D screen coordinates), and per-vertex lighting calculations. Now, graphics hardware started processing the true 3D data.

The new features of Direct3D included: T&L support, environment mapping with cubic textures, geometry blending, device state blocks. Additionally, Execute Buffers support was ceased, and draw primitive methods have become the only way of pushing the work to graphics hardware.

Direct3D Retained Mode was completely removed from DirectX SDK. The reason why it was done can be speculated that a decision was made to concentrate efforts on a single component of the library – Immediate Mode, – which would be better suited as a low-level, high-performance interface to graphics hardware. Retained Mode could not be made generic enough, and could not play the role of a general graphics engine. In lieu of removed Direct3D RM, DirectX introduced a special Direct3DX Utility Library (D3DX) that contained a wide selection of functions that helped with management of Direct3D interface objects, provided functions for loading graphical assets from files, and a range of 3D math functions. This made Direct3D a more complete package for graphics programming.

These improvements to hardware and API had increased performance of rendering applications. But there was another problem looming in the background: graphics pipeline as it was, was implemented in a particular manner, called Fixed Function Pipeline (FFP) [5]. Which meant, that the implementation of the pipeline algorithm was static, and the data passing

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

through was transformed on the way using fixed functions that could only be changed by special state values. FFP was a significant limiting factor in the generality of the pipeline: application developers were coming up with all sorts of possible visual effects that simply could not be implemented with the fixed functions.

G. DirectX 8 – 2000

The problem of FFP was solved with the introduction of programmable pipeline technology in graphics hardware. Programmable pipeline introduced two new stages into the process as a replacement for Transformation and Lighting and multitexturing modules – Vertex and Pixel Shaders [6]. Shader is a special small program that is executed by GPU at certain stages of data processing, and is applied to specific types of primitives: vertices for vertex shaders and fragments (or pixels, loosely) for pixel shaders. Shader programs were written at first in a version of assembly language that could be compiled in the runtime by application and uploaded to graphics hardware. Later, however, high level languages were developed to simplify shader programming, such as High-Level Shader Language (HLSL). HLSL later developed in a number of stages defined by Shader Model version, thus the first implementation was using Shader Model 1.

In DirectX 8, DirectDraw was completely merged into Direct3D since there was a little need in maintaining a separate library for display device management. The new combined component was named DirectX Graphics, and it included support of programmable pipeline and implemented a number of features, including: multisampling, point sprites, 3D volumetric textures, higher-order surface primitives, multiresolution geometry, vertex blending. Another useful addition was introduction of resources and the ability to manage where graphics data should be located in the memory, which gave applications better control over the data flow.

Direct3DX library was significantly expanded with functions that support working with meshes, geometry skinning (vertex blending), functions to assemble shaders, and a specialised Effect interface that encapsulated some of the common work of defining graphics pipeline using declarative syntax.

H. DirectX 9 – 2002

DirectX 9 had become a standard for Windows graphics development for many years to come. Even after the following versions were released, DirectX 9 was still in use. It can be said that this version encapsulated most of the requirements posed by applications, and some significant changes were needed on the application side in order to facilitate further development.

There were no radical changes in DirectX 9 compared to DirectX 8. All of the previous capabilities of the API were enhanced and improved. The API model underwent minor iterative version releases, which supported further extensive development in graphics hardware. HLSL was updated to Shader Model 2 and 3.

For the following years, graphics hardware, graphics API and applications were developing in an extensive manner, improving performance and increasing the number of supported resources.

I. DirectX 10 – 2006

In DirectX 10, graphics pipeline model was changed fundamentally. Legacy features of DirectX 9, like fixed function pipeline, were stripped off. And in general, the API had been upgraded and made cleaner. The following functional improvements were made: added a new programmable shader stage – geometry shaders (Shader Model 4); ability to output vertex data from the pipeline; pipeline state was organised into 5 immutable objects that significantly reduced loss of performance due to state switching; improved resource access; changed API architecture to have a layered runtime; and many others.

A drastic change like that meant that a lot of applications created using older versions of API could not be ported easily to use the new API. That throttled a widespread adoption of DirectX 10 for some time.

J. DirectX 11 – 2009

It had become clear at the time that programmable pipeline was the appropriate technology of choice for graphics hardware, since it combined great performance with a lot of flexibility necessary for applications. Thus, the main ways of graphics hardware development were to improve on the capabilities of shader stages and introduction of new ones.

DirectX 11 kept the architecture model of its predecessor. It expanded shaders to Shader Model 5 with addressable resources and resource types, subroutines, new types of shaders: compute, hull and domain. Two important improvements were introduced in DirectX 11: already mentioned compute shaders and multithreading. Compute shaders were a big change in graphics hardware world, they made it possible to execute general parallel algorithms very efficiently. Afterall, that was the whole purpose of graphics hardware from the beginning – to process large amounts of data in the most efficient manner. Now, graphics hardware had discovered a new use, and no longer was locked just to rendering. Multithreading support, on the other hand, was a big win for the rendering itself. Up until that moment, rendering processes were structured as a single conveyor belt with a single global state, which limited any attempts at parallel execution. And with GPUs hitting the limits of

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

single core improvements, it meant that some serious change was needed in order to unlock the next performance boost, and that change was multithreading.

From the applications perspective, there was a split: simpler applications that did not need cutting edge features were still using DirectX 9, but large players recognised and promoted DirectX 11 further. And the next version of DirectX was a conclusion of this endeavour, so far.

K. DirectX 12 – 2015

DirectX 12 had made the next large update to the architecture of the API. Compared with DirectX 11, the new version sported [7]: vastly reduced CPU overhead, up to 20% improvement in GPU efficiency and cross-platform development across Windows 10 devices. This came at a cost of the API being made lower-level, without attempting to abstract hardware capabilities any longer, it instead gave control over the hardware to graphics programmers.

The API had become thinner, and a lot of opportunities were opened for performance optimisation in relation to concrete visualisation applications. This, however, had significantly increased the entry level, and API documentation [7] explicitly said that DirectX 12 was designed for advanced graphics programmers. In a way, the API had become closer to the first version – DirectX 1. The API removed multiple ways it previously used to synchronise data and state between CPU and GPU processes, now all the work for resource management and command execution had to be performed by the application. Work submission was made truly parallel with introduction of a new model based on command lists. Those command lists may be reminiscent of Execute Buffers in the early versions, but in DirectX 12 they represented a completely independent executable workload for hardware that did not share a state with the rest of the computation, so it could be constructed and submitted in parallel in a stateless manner.

IV. CONCLUSION

The development arc of the graphics hardware, graphics API and applications has been dramatic in the past 25 years. The API started from low-level, attempted to add high-level abstractions, but eventually gave up and now it mirrors the hardware architecture. The hardware tried implementing rasterisation algorithm in a static manner, but it was proven not flexible enough to support the variety of applications, so the programmable pipeline was developed and enhanced. Applications started small in numbers, but then the number of them exploded, and after introduction of newer architectural models, a split appeared between simpler and more complex systems. Nowadays, DirectX 9 has become outdated, and it is harder for graphics programmers to base new applications purely on graphics APIs. In game development, for instance graphics engines have become very prominent since their development teams had enough expertise to utilise the power of the new API fully.

In our previous works we discussed the problem of complexity of the graphics pipeline and API [8, 9, 10]. Our approach to extension of the pipeline with higher-level primitive processing can make graphics API better suited for use in complex graphics application going forward.

REFERENCES

[1] Microsoft, DirectX SDK Documentation for versions 1–12, 1995–2015.

[2] “Rise of 3dfx”, https://vintage3d.org/3dfx1.php

[3] Intel Corporation, “Accelarated Graphics Port Interface Specification”, 1996.

[4] NVIDIA Corporation, “Transform and Lighting”, Technical Brief, http://developer.download.nvidia.com/assets/gamedev/docs/TransformAndLighting.pdf

[5] David B. Kirk, Wen-mei W. Hwu, Programming Massively Parallel Processors, 2nd ed., Morgan Kaufmann, 2013, pp. 23–39.

[6] Ron Fosner, Real-Time Shader Programming, Morgan Kaufmann, 2003, pp. 88–111.

[7] Microsoft, DirectX 12 Programming Guide, https://docs.microsoft.com/en-us/windows/win32/direct3d12/what-is-directx-12.

[8] V. Krasnoproshin, D. Mazouka, “Frame Manipulation Techniques in Object-Based Rendering” Communications in Computer and Information Science, vol. 673: “Pattern Recognition and Information Processing”, Springer, 2017, pp. 97–105.

[9] V. Krasnoproshin and D. Mazouka, “Graphics Pipeline Evolution Based on Object Shaders” Pattern Recognit. Image Anal. 30, 2020, pp. 192–202, https://doi.org/10.1134/S105466182002008X

[10] V. Krasnoproshin, D. Mazouka, “Data-Driven Method for High Level Rendering Pipeline Construction”, Neural Networks and Artificial Intelligence. Communications in Computer and Information Science, vol. 440, 2014, pp. 191–200.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Extraction of Human Body Parts in Image Using Convolutional Neural Network and Attention Model

Viktoria Sorokina
Belarusian State University
Minsk, Belarus
viktoria.sorokina.96@gmail.com

Sergey Ablameyko
Belarusian State University
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
ablameyko@bsu.by

Abstract. **In computer vision, human body parts extraction is a challenging task for many applications. In this work, we propose the algorithm to extract human body parts in images using the OpenPose system and attention model. The novelty of the proposed work is that algorithm is based on a convolutional neural network that uses a nonparametric representation to associate body parts with people in an image in combination with a new attention model that learns to focus on specific regions of different input features. The algorithm is a part of Smart Cropping system developed by us which aim is to cut the images and prepare e-commerce catalog.**

Keywords: **human body parts extraction, attention model, convolutional neural network, smart cropping**

I. INTRODUCTION

Task of human body parts extraction is very important for many applications, particularly in e-commerce. It is well known that deep learning can be used to identify parts of a person's body [1]. Extraction of human body parts is the foundation of another task of computer vision – pose estimation, which can also be considered as the problem of determining the position and orientation of the camera relative to a given person or object.

Solving the problem of a human body part extraction, an object or person (or several people) could be tracked in real world at an incredibly detailed level. This powerful capability opens up a wide range of possible applications.

In this article, we will consider the application of the human body parts extraction in e-commerce tasks, namely, when creating an e-commerce catalog.

Creating a catalog for an e-commerce store includes preparing images and content for them [2]. When preparing images of clothes, a full-length photograph of a person presenting several items of clothing at the same time is usually used. Such an image is cut into pieces in accordance with certain rules. For example, for a skirt it is necessary to show the part from the waist to the feet, for a shirt - from the crown of the head to the hips. Currently, the slicing process is done manually. To automate the process, the Smart Cropping system has been developed, which allows, by solving the problem of human body parts detection, to cut images.

In this paper, we propose an algorithm for human body parts extraction using a neural network and attention model. Having received the key points of the human body, the positional relationship between them is calculated, after which it is used to crop the original image and create a set of images representing the goods. The algorithm is capable of preparing images of shoulder clothing (clothing resting on the upper supporting surface of the body, bounded from above by the articulation lines of the body with the neck and upper limbs, and from below by a line passing through the protruding points of the shoulder blades and chest), waist clothing (clothes resting on the lower supporting surface of the body, bounded at the top by the waist line, and at the bottom by the hip line), hats and shoes. The average computation time for one frame is 1.5 seconds.

Our main contributions are

- developing the new system of image cutting in ecommerce sphere based on the deep neural network modified by the techniques that help a “model-intraining” notice important things more effectively;
- extend the existing OpenPose system using Attention model that helped VGG architecture used in OpenPose to learn and detect more image features.

We achieve state-of-the-art results on standard benchmarks including the COCO dataset.

II. RELATED WORK

The classical approach to the problem of a human body parts extraction and pose estimation, presented in [3], includes representation of the object as a set of "parts" located in a deformable configuration (not rigid). Most of the newer pose estimation systems use

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

convolutional neural networks as the main building block, largely replacing hand-crafted functions and graphical models; this strategy has significantly improved standard approaches.

DeepPose [4] is the first deep convolutional neural network architecture applied to the problem of human pose estimation. It achieved the performance of advanced algorithms and outperformed existing models. In this approach, pose estimation is formulated as a convolutional network regression problem to determine the joints of the human body. The work also uses a cascade of such regressors to refine and obtain more accurate estimates of the pose. However, the disadvantage of the model is the complexity of training due to the specifics of regression, which weakens generalization and, therefore, does not work well in certain regions.

Newer techniques transform the pose estimation problem into a heatmap estimation problem, where each heatmap indicates the reliability of the location of the n-th key point of the human body. The work [5] is based on this approach.

The work [5] is based on the architecture that uses a convolutional neural network ConvNet [6] and a refinement model. In the method, heatmaps are created by parallel running an image rendered at different resolutions to capture objects at different scales at the same time. The disadvantage of this approach is the lack of structural modeling.

In this article, the OpenPose architecture [7] that is divided into feature selection block modified by the attention model [8] and maps generation block is used to identify parts of the human body.

OpenPose [7] is a deep feed forward neural network. This method can effectively detect 2D positions of human body parts in real-time RGB images. It does this by detecting and associating body parts using part similarity fields (PAFs) and confidence maps. A confidence map is a probability density function for a new image that assigns a probability to each pixel of a new image, which is the probability of a pixel belonging to a body part in an object in the previous image. The detection of body parts occurs in a sequential manner, performing bottom-up prediction using spatial context.

Attention Model (AM), first introduced in 2015 for Machine Translation [8] has now become a predominant concept in neural network literature. Attention has become enormously popular within the Artificial Intelligence (AI) community as an essential component of neural architectures for a remarkably large number of applications in Computer Vision [9].

The main purpose of the attention model is to use attention maps. An attention map is a scalar matrix representing the relative importance of activation layers at different 2D spatial positions with respect to the target. Attention model uses maps to define and use effective spatial support of visual information used by the convolutional network in decision making.

The constructed in this work model makes it possible not only to structure parts of the human body due to the fields of similarity of parts, but also to highlight parts of the human body in more detail due to stimuli that enhance significant and suppress insignificant objects in the image. This result is achieved due to the construction of a two-dimensional matrix of estimates for each heat map.

III. METHOD

We developed the method allowing to crop the image automatically without usage of the human resources. It calls Smart Cropping and includesthe following components:

- Feature selection module;
- Vector and heatmap building module;

□Position relations calculation module; □Cutting according to the specified rules module.

The architecture of the Smart Cropping method is shown in Fig. 1.

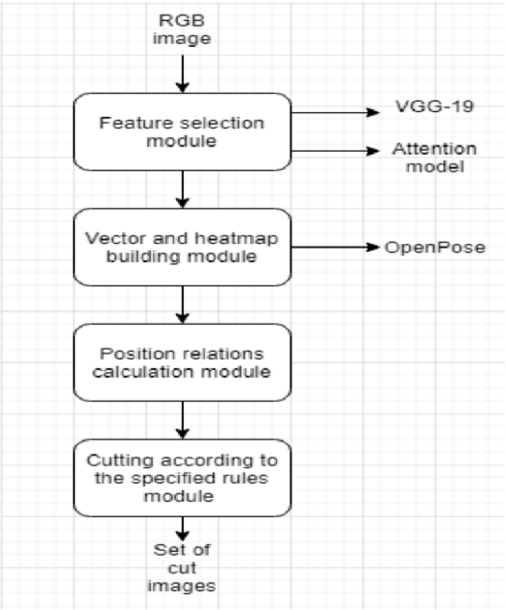


Fig. 1. Smart Cropping architecture

Due to theme of this article is extraction of human body parts in image using convolutional neural network and attention model, special attention will be for Feature selection module, other modules will be described briefly.

A. *Feature Selection Module*

Feature selection module is the first step of the Smart Cropping method and the main focus of this article. In this module the image is processed with the help of the deep convolutional neural network VGG-19 that is a part of the OpenPose architecture. The novelty of the algorithm proposed in this article lies in the modification of the network VGG-19 by the attention model – the learned attention maps neatly highlight the regions of interest while suppressing background clutter.

VGG-19 [10] is a type of the VGG (Visual Geometry Group) model, which consists of 19 layers (16 convolutional layers, 3 fully connected layers, 5 MaxPool layers and 1 SoftMax layer). The architecture is shown in Fig. 2. The main idea behind VGG is to show that classification/localization can be improved by increasing the convolutional block and using a 3×3 convolution kernel to help highlight the features of the image.

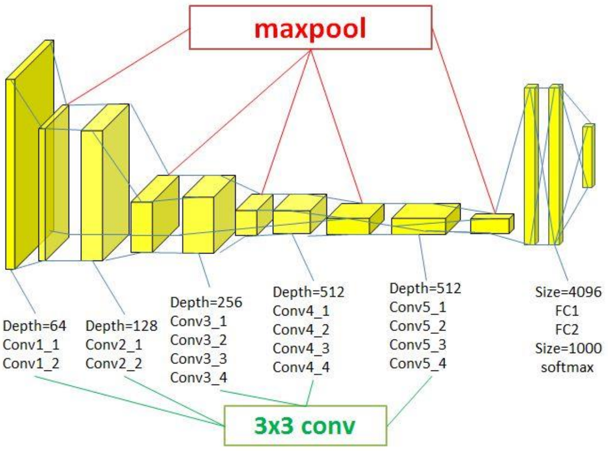


Fig. 2. VGG-19 architecture

One of the ideas of this work is to enforce the existing architecture by combining it with the attention model. The attention model engine learns during network training and should help the network focus on key image elements.

We developed the algorithm which is based on the hypothesis that there is an advantage in identifying significant areas of the image and enhancing their influence, while suppressing irrelevant and potentially misleading information in other areas. In particular, it is expected that providing more targeted and economical use of image information should help in generalizing changes in data distribution, as happens, for example, when training on one set and testing on another. In the standard convolutional network architecture, the global image descriptor $\mathcal{g}$ is obtained from the input image and traversed through the fully connected layer to obtain the prediction probabilities. The attention model expresses $\mathcal{g}$ through the mapping of input data into a multidimensional space in which observable visual concepts are presented in different dimensions to make classes linearly separable.

We included 2 key changes in VGG-19 architecture and the algorithm is the following (Fig. 3):

- after layers 7, 10, and 13 (highlighted in yellow in Fig. 3), attention estimators are inserted, on the basis of which a binary mask is calculated, where 0 is irrelevant information for the desired object, and 1 is important. The mask, represented by the matrix, is then multiplied by the original result of the layer for which it was calculated, for example, 7, thus overestimating attention;
- the last fully connected layer was replaced with a fully connected layer, the input of which is the results of 3 attention estimators.

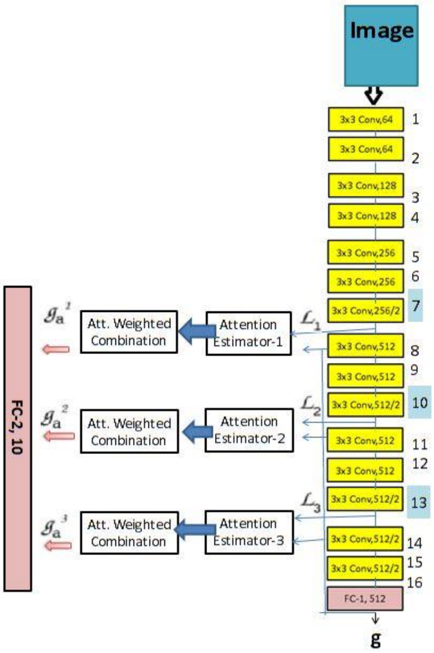


Fig. 3. Attention model architecture

B. *Vector and heatmap building module*

The second module in the Smart Cropping method is vector and heat maps building. Inspired by the work [7], we use an OpenPose system to produce the vector and heat maps to detect the position of human parts on the image. Since our work is an extension of their model, we will only present a very brief overview of the architecture.

The OpenPose input is an RGB image that forwards through the VGG-19. There are two branches at each stage: one for the heatmap detection and the other is for vector map detection. Obtaining heatmap and vector map, one could determine all the key points in the image. The OpenPose architecture is shown in Fig. 4.

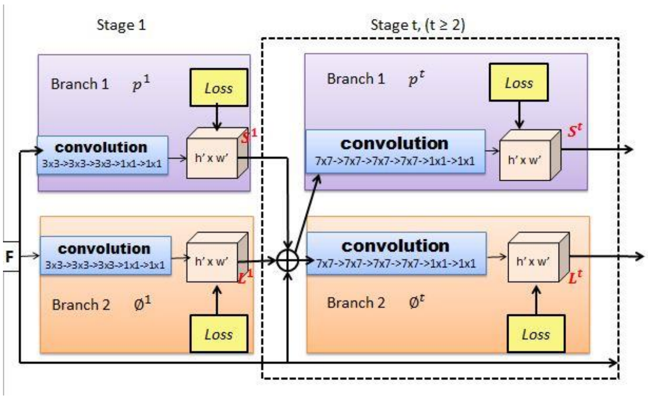


Fig. 4. OpenPose architecture

C. *Position relations calculating module*

To implement a cropping of the original image and create a set of images representing goods, it is needed to get the coordinates of the key points of the human body, and then calculate the positional relationships between them through the coordinates of each key point. The coordinates of three points should be known in order to calculate the angle formed by the three points. Then using the range of values of these angles the pose of the person is etimated and correct cropping is made. The formulas for calculating these angles are shown below.

Let 3 points $A(x_1, y_1)$, $B(x_2, y_2)$, $C(x_3, y_3)$ are known. The corresponding vectors are

$$\overrightarrow{AB}: (x_2 - x_1, y_2 - y_1),$$
$$\overrightarrow{AC}: (x_3 - x_1, y_3 - y_1)$$
$$,$$
$$\overrightarrow{BC}: (x_3 - x_2, y_3 - y_2).$$

Then

$$|AB| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$
$$|AC| = \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2},$$
$$cos\angle A = \frac{(x_2 - x_1)(x_3 - x_1) + (y_2 - y_1)(y_3 - y_1)}{|AB||AC|}.$$

System could detect 23 key points of the human body (e.g. right elbow, left hip, etc.).

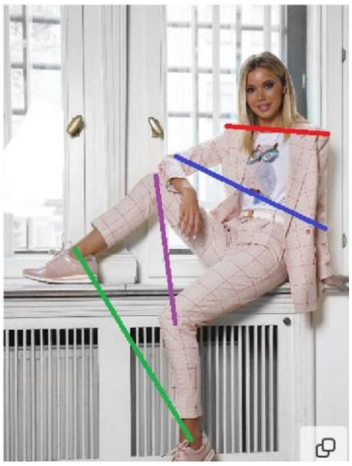


Fig. 5. Example of the staight lines construction

23 key points construct the straight lines that form the person's pose, for example, left and right shoulders, left and right hips, etc. An example of building straight lines is shown in Fig. 5.

IV. EXPERIMENTS AND RESULTS

In e-commerce, a product catalog is an illustrated list of goods or services. The catalog is compiled for the needs of customers, buyers or other interested parties. The hierarchical structure of the catalog consists of categories and subcategories, which contain the actual information about the goods. An electronic catalog is a type of the product catalog where all the information is presented in electronic form.

Such catalogs are the most important, and often they are the only communication channel between the manufacturer or supplier of products or goods and the buyer. The main goal of the electronic catalog is to present information in such a way that the buyer has the ability to effectively search for the necessary information without any difficulties with its understanding and use.

During the creation of an electronic catalog for clothes, a person is usually photographed in full-length, then unnecessary clothes are cut off: for example, when generating a page for a jacket, tit should be cut from the head / neck of the model to the hips. This cropping is done manually and takes a long time.

The network for determining key points was trained on an NVIDIA T4 GPU using VGG-19 modified by the attention model for feature extraction, batch_size = 6, the number of iterations was 800,000.

The resulting accuracy is 86% for the dataset including images of e-commerce products.

Fig. 6. Example of the image cropping

Our algorithm is capable of cropping shoulder clothes (from eyes / nose / shoulders to wrists / hips), waist clothes (from wrists / elbows to toes / knees / ankles), hats (from top of image to shoulders), and shoes (from knees / ankles to toes / bottom of the image), as well as their combinations.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The time needed to prepare one catalog containing 10 products is 5 minutes.

Example of the image cropping is presented in the Fig. 6.

V. DISCUSSION AND CONCLUSION

In the course of the research, an algorithm based on the OpenPose architecture using VGG-19 modified by the attention model was developed. The algorithm showed improvement in the accuracy by 8%, and it is capable of recognizing 23 key points of the human body.

To identify 23 key points of the human body the COCO dataset [11] was used. However, for e-commerce tasks, it is also necessary to define points such as the chest, crown of the head, abdomen, which are not represented in COCO. Thus, the system can be improved by training on an extended dataset.

The resulting network became the foundation of the Smart Cropping system, which allows cropping images based on positional relationships between key points of the human body to create a catalog of e-commerce products. This allows preparing images that form the product catalog for the classes of shoulder, waist and outerwear clothes, as well as shoes and hats.

The accuracy of the trained model is 86% for the dataset represented by images of e-commerce products, due to the specifics of the area. Accuracy can be improved by introducing a block of generative adversarial networks that can predict the presence of a key point that is not explicitly present in the image (for example, a floor-length dress that covers the knees).

The developed algorithm can be improved by expanding the recognized key points, and also used to crop images of other product categories.

ACKNOWLEDGMENT

This research was partly supported under the project BRFFI F20KIGT-006.

REFERENCES

[1] Y. Chen, Y. Tian, M. He, Monocular human pose estimation: A survey of deep learning-based methods, Computer Vision and Image Understanding, 2020, vol. 192.

[2] eCommerce Product Image Guide [Electronic resource]. Mode of access: https://www.threekit.com/blog/ecommerce-product-imageguide-2020. – Date of access: 25.03.2021.

[3] Y. Yang, D. Ramanan, Articulated Human Detection with Flexible Mixtures of Parts, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, vol. 35, no. 12, pp. 2878–2890.

[4] DeepPose: Human Pose Estimation via Deep Neural Networks / A. Toshev and C. Szegedy // IEEE Conference on Computer Vision and Pattern Recognition, June 24-27, 2014, Columbus, OH, USA, 2014, pp. 1653–1660.

[5] J. Tompson, A. Jain, Y. LeCun, C. Bregler, Join training of a convolutional network and a graphical model for human pose estimation, Advances in Neural Information Processing Systems, 2014, pp. 1799–1807.

[6] J. Tompson [et al.] Efficient object localization using Convolutional Networks, 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, 2015, pp. 648-656.

[7] Zhe Cao [et al.] Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, 21-26 July 2017, Honolulu, 2017, pp. 1302–1310.

[8] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In 3rd International Conference on Learning Representations.

[9] Feng Wang and David MJ Tax. 2016. Survey on the attention based RNN model and its applications in computer vision. arXiv preprint arXiv:1601.06823 (2016).

[10] Karen Simonyan and Andrew Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, International Conference on Learning Representations, San Diego, May 7-9, 2015, San Diego, 2015, pp. 1137–1149.

[11] COCO dataset // COCO 2018 Keypoint Detection Task [Electronic resource]. – Mode of access: http://cocodataset.org/#overview – Date of access: 05.04.2019.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Reindeer Recognition and Counting System Based on Aerial Images and Convolutional Neural Networks

Vladimir Mikhailov
St. Petersburg Federal Research Centre
Russian Academy of Sciences
Saint Petersburg, Russia
mwwcari@gmail. com

Vladislav Sobolevskii
St. Petersburg Federal Research Centre
Russian Academy of Sciences
Saint Petersburg, Russia
arguzd@yandex. ru

Abstract. **Described animal recognition and counting system based on convolutional neural networks with MRCNN architecture. Initial training of the network is performed using a basic array of MS COCO images, and additional training is performed using an array of aerial photographs of reindeer herds. A web-interface of the system has been developed. The error of counting reindeer in the image from the verification sample is about 13%.**

Key words: **reindeer, recognition, convolutional neural networks, aerial imagery**

I. INTRODUCTION

The development of an automatic system for recognition and counting the number of reindeers was driven by the following reasons. The currently used methodology for counting the number of wild reindeer in tundra populations (Taimyr, Yakutia, Chukotka reindeer, and migrating herds of reindeer from Canada and Alaska) is based on the ecological characteristics of the species, consist in the fact that in hot weather, during the flight of blood-sucking insects, reindeer gather in herds of many thousands in a limited area in the northern part of their summer range (subzones of arctic and typical tundra) [1, 2]. The herds in the aggregations are photographed and the number of animals in them is counted directly, "by head".

The advantage of this approach over purely approximation-based population estimation methods [3] is the significantly greater accuracy of the results, as the vast majority of animals in the population (up to 90%) are counted directly from herd photographs and only a small number are estimated by area-based approximation. However, manual processing of the survey results for large populations takes about three months, whereas for the ecologically based management of population dynamics, the non-depletion of biological resources of the species and the determination of the norms of commercial reindeer harvesting, it is desirable to have the population data in the second half of August, that is, 10-15 days after the end of the aerial count.

The task was therefore to automate the processing of aerial photographs in order to reduce the time it takes to obtain aerial survey results.

II. SELECTING A CALCULATION MODEL

The technology of convolutional neural networks (CNN) is adopted as the intellectual basis of the recognition system. This class of architectures is a highly specialized tool, suitable primarily for images and other data that can be represented in matrix form. As images store all information as two-dimensional matrices (i.e., as pixels), it is necessary to consider not only values from the neurons themselves but also values from a group of nearest neurons when working with images. To this end, besides neurons there is another type of elements in convolutional layers of CNN that apply certain linear operations to all input data of each neuron of the layer - the convolution core. The convolutional kernel is a grid that "slides" across the image (or previous layer convolutional layer) and looks for patterns and patterns in the data. If it finds a part of the image that matches a kernel pattern, it passes a large positive value to the current layer's computational neuron. If there is no match, the kernel will pass a small value or zero.

Because the convolution kernel is applied to every position in the image, the convolution layer of CNN is extremely effective in image processing tasks because features or patterns in the images can appear anywhere in these images. That is, CNN is able to analyze context-dependent data.

The Mask Regions with Convolution Neural Networks (MRCNN) architecture [4] was chosen for this task. This architecture is a subset of the classical CNN. Due to the complexity of the architecture, it more successfully copes with the tasks of semantic and object segmentation of images. It is a modification of the existing architecture Fast Region-based Convolution Neural Network (FRCNN), in which was added a module responsible for recognition and generation of object masks.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

FRCNN is an CNN that searches the image for objects and then additionally classifies the found object. The output of this model is the bounding rectangles localizing each object in the image and the class label of the found object with a confidence score.

In terms of operating logic, MRCNN, the first stage of operation is the same as that of a conventional FRCNN. It consists of simultaneously running two enabled artificial neural networks: a mainline network (ResNet, VGG, Inception or similar) and a regional positioning network. These networks process each image received at the MRCNN input and provide an output of a three-dimensional array - an array of suggested regions. This array contains the coordinates of regions in the input image that contain an object.

In the second stage of the work, conventional FRCNN predicts the bounding rectangle coordinates and feature classes for each of the proposed regions obtained in the first stage. Each proposed region can have a different size, but since convolutional layers in CNN always require a fixed size vector for prediction, this step also scales the regions found. The size of the regions is scaled using either the RoI algorithm or the RoIAlign method.

MRCNN is in turn an extended version of FRCNN, augmented with a branch to predict segmentation masks for each area of interest. The second phase of MRCNN already uses only RoIAlign, which helps to preserve the original spatial coordinates that are offset when RoI is used. This is then necessary so that the RoIAlign output can be combined with the data from the first phase, and a mask can be generated for each RoIAlign response using the Mask Head module (which in turn is also implemented using convolutional layers). Such masks are a two-dimensional matrix which, for each pixel within a region's boundaries, determines whether or not that pixel belongs to the object in question.

This approach allows the boundaries of the object being searched to be defined more precisely. Ideally, the MRCNN can accurately calculate all the pixels in the image that represent the object being searched for.

III. THE MODEL CREATED

Since there is insufficient data for training in the chosen application domain, it was decided to perform the main training on a dataset that includes images of other animals. This was necessary for the MRCNN to learn to recognize animals as a class of objects. And for the specific task, this MRCNN was already further trained on a specific dataset, including aerial images of reindeer herds. The specifics of the task are that the herds are photographed from different distances, in different landscapes, under different light conditions, the animals on the pictures have different colors and can be at different angles to the camera, can overlap each other. These peculiarities of aerial photos create additional difficulties in solving the task of reindeer identification. Therefore, the presented task is non-trivial and the application of CNNs trained on common datasets is not possible.

The main training dataset for MRCNN was the MS COCO (Microsoft Common Objects in Context) image array [5]. This array is one of the largest datasets used to date for training machine learning models to solve detection and segmentation problems. The dataset consists of 328 thousand images. All images are marked up and formed into training samples. Therefore, using this dataset for basic training of MRCNN allows all the basic concepts of different object classes, including animals, to be specified for CNN. However, images of reindeer are not part of MS COCO and MRCNN by default is not able to distinguish them from a number of other animals (sheep, gazelles, cows, horses). Therefore, in order to recognize reindeer, the MRCNN needs to be further trained using aerial photo arrays with images of these animals.

An input dataset containing a training sample of 100 aerial images of herds with all animals tagged and a test sample of 30 original herd images was prepared for training the CNN. The model was trained on 20 epochs, with 60 training steps per epoch, with a training rate of 0.0058 and a detection miss threshold of 0.7. On the test data set, the trained model recognized an average of 82% of the deer correctly. The deer recognition accuracy can be improved by retraining the CNN on the extended training dataset.

So far, a software package with a web interface https://regionview.ru/ai/ has been created for the developed network, and the program itself has been deploy for limited use.

To use the system, users must upload a JPEG (.jpg) or GIF (.gif) image to their computer.

The system interface contains a set of window forms that provide:

- downloading of aerial images from the user's computer for processing,
- running a program to recognition and count the reindeer in the pictures,
- presentation of the results of the program work, with an image on the screen showing the reindeer images recognized by the system and the total number of animals counted,
- downloading the results to the user's computer.

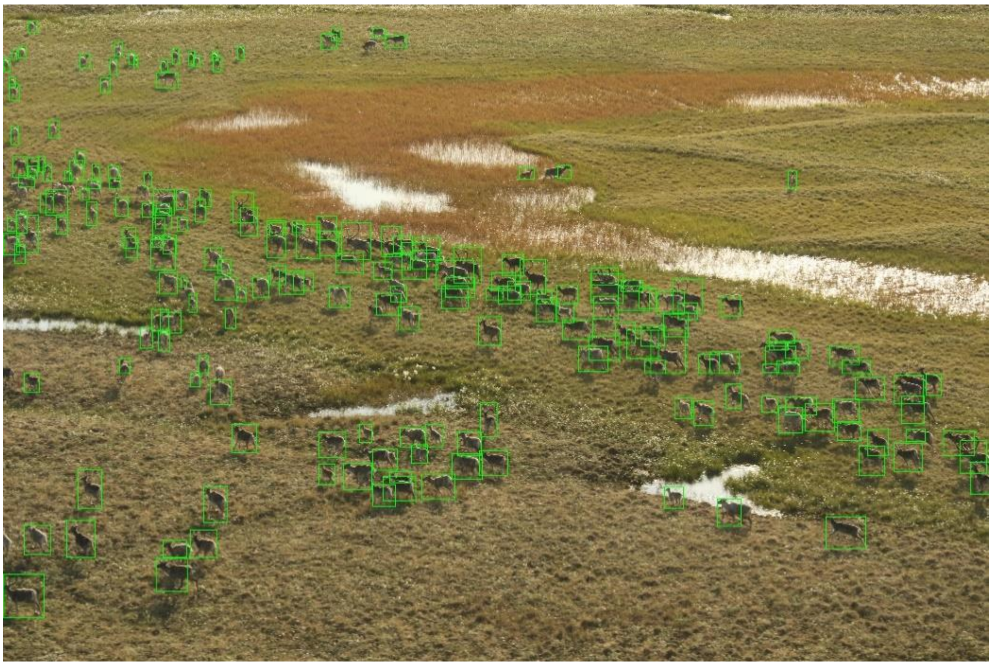


Fig. 1. Program-marked aerial photo of a herd of wild reindeer

After viewing the marked image, if the user is not satisfied with the accuracy of the software package, they can continue to process the image manually in any graphics editor that supports the. jpg file extension.

As an example, Figure 1 shows the result of the software, a raw image with automatically recognized and tagged animals. The figure shows that MRCNN can work with images that are noisy with background objects - puddles, lakes, bumps, etc. None of the background objects were mistaken for a reindeer. Also, it is noticeable that the MRCNN well with herds in which the reindeer gather in very dense groups and partially overlap each other. The recognition error was about 10%.

IV. CONCLUSION

In general, verification of the system on an independent sample of aerial images showed that MRCNN can work with images that are noisy with background objects - puddles, lakes, hillocks, polygons, etc. The network successfully distinguishes individual animals in dense groups, animals at different angles and at different distances from the camera. The recognition error of the deer in the image was about 17%.

ACKNOWLEDGEMENTS

The work was carried out at the St. Petersburg Federal Research Centre of the Russian Academy of Sciences (SPb FRC RAS) with the support of the World-Wide Fund for Nature - Contract №BBF003107, RFBR grant №19-37-90112 and Budget Subject №0073- 2019-0004.

REFERENCES

[1] V. A. Zyryanov, B. M. Pavlov, G. D. Yakushkin, Ecological basis of counting the number of game animals in the tundra zone of Taimyr. Problems of hunting economy of Krasnoyarsk Region, Krasnoyarsk, 1971, P. 70-72.

[2] L. A. Kolpashchikov, B. M. Pavlov, V. V. Mikhailov, Methodology of aerial counting and determination of polling rates of Taimyr wild reindeer population: methodological recommendations, Saint-Petersburg, 1999, 25 p.

[3] N. G. Chelintsev, Mathematical basis of animal recording, Moskva, 2000, 431 p.

[4] K. He Mask R-CNN, G. Gkioxari, P. Dollar, R. Mask Girshick, Computer Vision and Patter Recognition, Cornell University, 2017.

[5] Common Objects in Context, https://cocodataset.org/.

[6] Automated Animal Identification Using Deep Learning Techniques, Proc. of the Nat. Acad. of Sciences of the USA, 2018.

[7] P. Ganesh, K. Volle, T. F. Burks, S. S. Mehta, Deep Orange: Mask R-CNN-based Orange Detection and Segmentation, IFAC-PapersOnLine, 2019, vol. 52, issue 30, pp. 70-75.

[8] G. Zhao, J. Hu, W. Xiao, J. Zou A mask R-CNN based method for inspecting cable brackets in aircraft, Chinese Journal of Aeronautics, 2020.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

A Simple Indoor Fall Control System for the Elderly Based on the Analysis of Object Bounding Box Parameters

Katsiaryna Kosarava
Dept. of System Programming and Computer Security
Yanka Kupala State University of Grodno
Grodno, Belarus
koluzaeva@gmail.com
ORCID 0000-0001-7326-5307

Boris Assanovich
Dept. of Information Systems and Technologies
Yanka Kupala State University of Grodno
Grodno, Belarus
bas@grsu.by

Abstract. **In this paper a human fall detection problem using video recordings features such as bounding box height to width ratio and speed of the bounding box movement is considered. We examine two datasets, the first of which, in addition to video recordings, also contains accelerometer readings. Support vector machine, decision tree and random forest were used as classification models. These models were built on secondary features generated with the use of the tsfresh library. The experimental results showed that the analized datasets are almost linearly separable according to some newly generated features. Thus, the use of the tsfresh library allows to apply simpler models and at the same time to provide a higher classification accuracy up to 1.0 in comparison with more complex LSTM models.**

Keywords: **human fall detection, machine learning, classification models, tsfresh, LSTM**

I. INTRODUCTION

In recent years, much attention has been paid to monitoring the activity of the elderly. One of the biggest problems of people over 70 consists in losing their balance and falling. Falls are especially dangerous for people who live alone, which reduces the ability to provide quick assistance, Giannakouris et al. [1]. This problem has led to an increase in the number of studies on remote detection of falls, allowing fast and qualified assistance to the elderly. Such systems often use sensor devices to register falls based on the collected data. Approaches, based on wearable sensors or mobile phones, require wearing a monitoring device. The problem here is that old people often forget to wear such devices. Alternative systems based on computer vision methods do not require the participation of the elderly [2]. In addition, today video cameras are becoming cheaper and do not require complex manipulations for their installation.

Kwolek and Kepski [3] collected data from a 3D accelerometer and image depth maps to detect falls. The measured acceleration value was exploited to control the defined threshold to extract a number of human characteristics and run the classifier to determine the presence or absence of a fall. Redmon et al. [4] studed a convolutional neural network (CNN) to analyze images, extracted from a video and get features using an optical flow method to detect motion between two consecutive video frames. Several cameras or additional devices can be exploited to detect movement, as was done for UP-Fall Detection dataset. The disadvantage in the considered cases is the use of a special camera and/or several cameras, as well as resource-intensive structures to obtain the required accuracy. Recently Lezzar et al [5] proposed a simple algorithm for determining fall, and common daily activities with the use of computer vision and deep learning, based on a 2D camera with occlusion recognition and control of the transition between states before and after a fall. By increasing the controlled states, the authors claim that they achieved an accuracy of 93.94% and recall 100% with SVM classifier.

In this study we describe a simple application of both linear (SVM, random forest, etc.) and nonlinear classifiers (LSTM) for detecting a person's fall, and also propose a transition to new features for time series analysis [6]. The article is organized as follows. The next section describes the algorithm for controlling a person's fall and the structure of the system. Further, experiments using various classifiers are presented and their effectiveness is compared. The last section contains the conclusion and prospects for the development of the methodology proposed.

II. FALL DETECTION MODELS BASED ON VIDEO AND ACCELEROMETER RECORDS

A. Dataset description

In this section, we compare the performance of fall detection models trained on a dataset [3]. This dataset contains 70 (30 falls + 40 activities of daily living (adl)) sequences. Fall events are recorded with 2

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Microsoft Kinect cameras and corresponding accelerometric data. ADL events are recorded with only one device (camera 0) and accelerometer. Sensor data was collected using PS Move (60Hz) and x-IMU (256Hz) devices.

Dataset containes synchronized video camera records: frame number, time in milliseconds since sequence start and interpolated accelerometric data, corresponding to image frame. The cameras are recorded independently, so they are not strictly synchronized (synchronization based on nearest timestamp value). Raw accelerometric data contains time in milliseconds since sequence start and accelerometer data: SV_{total}, A_x, A_y, A_z. All accelerometer data are in gravity units (g). Total sum vector is calculated as follows: $SV_{total} = \sqrt{A_x^2 + A_y^2 + A_z^2}$.

The dataset also contains files with extracted features from depth maps stored in CSV format. Each row is one sample of data corresponding to one depth image with following features: *sequence name* ('fall-01','adl-01' etc.), *frame number*, *label* - '-1' means person is not lying, '1' means person is lying on the ground; '0' is temporary pose, when person "is falling", *HeightWidthRatio* - bounding box height to width ratio, *MajorMinorRatio* - major to minor axis ratio, computed from BLOB of segmented person, *BoundingBoxOccupancy* - ratio of how bounding box is occupied by person's pixels, *MaxStdXZ* - standard deviation of pixels from the centroid for the abscissa (X axis) and the depth (Z axis), respectively, *HHmaxRatio* - human height in frame to human height while standing ratio, *H* - actual height (in mm), *D* - distance of person center to the floor (in mm), *P*40 - ratio of the number of the point clouds belonging to the cuboid of 40 cm height and placed on the floor to the number of the point clouds belonging to the cuboid of height equal to person's height.

B. Experimental setup

In [3], the authors came to the conclusion that the fall detectors using only inertial sensors generate too much false alarms. This means that some daily activities are erroneously signaled as fall, which in turn leads to frustration of the users. Therefore, the authors train the SVM model to classify the position of a person in the frame (person is lyings / is'n lying) based on the extracted features of the depth maps to confirm the SV_{total} accelerometer indicators.

In this paper, we explore the possibility of combining accelerometer records and features extracted from depth maps.

The first approach to training the model was based on the assumption: if a person in the frame is in a horizontal position (i.e., person is lying), it is necessary to determine whether this is a consequence of a fall or not. To do this, we select frames from the video records on which a person is lying according to the *label* attribute, and the corresponding values of the remaining attributes. From the accelerometer records we take the SV_{total} value corresponding to the numbers of the selected frames for each video recording. We took only *HeightWidthRatio* from all the features of depth maps, since it is this indicator that allows us to determine the position of a person in the frame (lies / does not lie). The described sequence of actions was applied separately to the datasets for fall and for adl. Thus, we got a labeled dataset consisting of the following fields: SV_{total}, *HeightWidthRatio*, *label* (-1: not fall / 1: fall). After balancing the data, we got 903 records for each class. Visual analysis showed that with this approach to data generation, the classes are strongly mixed with each other. This is also confirmed by the modeling results: the best trained SVM model with 'rbf' kernel showed an accuracy of only 75% on 10-fold cross-validation.

The second approach to analyzing the dataset was to consider each feature of depth maps and accelerometer readings in the sequence in which they were measured, i.e. like time series. And construct new significant features for these time series. For this purpose, the library **tsfresh** [7] was used. It allows to extract over 1200 features from time series. Tsfreash contains a feature selection method that evaluates the importance of the different extracted features. To do so, for every feature the influence on the target is evaluated by an univariate tests. Those tests generate p-values that are then evaluated by the Benjamini Hochberg procedure to decide which features to keep and which to delete [8]. It was visually determined that the *HhmaxRatio, H, D, P*40 signs are practically identical for the two classes, so they were removed. For the remaining features $Sv_{total,}$ *HeightWidthRatio, MajorMinorRatio, BoundingBoxOccupancy* and *MaxStdXZ*, 824 new significant features were generated using tsfresh. Thus, we got a dataset with a dimension of 70 by 824 (70 is the number of videos, 824 is the number of new significant features). On these data, models of SVM, decission tree and random forest were trained, with the selection of parameters according to 10-fold cross-validation, 1st Model Table I. The best result was shown by the SVM model with a 'linear' kernel (based on normalized data), the numbers of support vectors are 9 and 7 for two classes, respectively, 1st Model Table I. Moreover, the Decision Tree model almost accurately classified the data using only two features: the average, absolute value of consecutive changes of the series SV_{total} inside quantile corridor [0.0, 1.0] and the absolute value of zero coefficient of the one-dimensional discrete Fourier Transform for *MaxStdXZ* by fast fourier transformation algorithm [10].

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

If we leave only the SV_{total} and *HeightWidthRatio* features and generate new significant features with tsfresh, then we get data with dimensions 70 by 434. The best model was 'linear' SVM model, the numbers of support vectors are 8 and 7 for two classes, respectively, 2nd Model Table I.

TABLE I. EXPERIMENT RESULTS for FALL DATASET [3]

Trained Models	**Models Accuracy**			
	1st Model		**2nd Model**	
	Train acc.	Test acc.	Train acc.	Test acc.
SVM, ‘linear’ n_support	1.0 [9, 7]	1.0	1.0 [8, 7]	1.0
SVM, ‘rbf’ n_support	0.86 [17,16]	1.0	1.0 [16, 12]	1.0
Decision Tree	0.98	1.0	0.98	0.86
Random Forest	0.91	1.0	1.0	0.93

III. FALL DETECTION MODELS BASED ON VIDEO RECORDS

A. Dataset description and preprocessing

Obviously, using only video cameras is a more convenient and less expensive way of monitoring. As the second fall detection dataset we take one collected by the Imaging and Artificial Vision laboratory (ImViA) [9]. ImViA has been developing an architectures for specific vision and visualization systems and new applications in medical imaging unconventional imagery for robotics and scene reconstruction and analysis. ImViA fall detection dataset was collected in realistic video surveillance setting using a single camera. The frame rate is 25 frames/s and the resolution is 320×240 pixels. Video sequences contain variable illumination, and typical difficulties like occlusions or cluttered and textured background. The actors performed various normal daily activities and falls. The dataset contains 191 videos that were annotated, for evaluation purpose, with extra information representing the ground-truth of the fall position in the image sequence. Each frame of each video is annotated: the localization of the body is manually defined using bounding boxes. This annotation allows to evaluate the classification features independently from the automatic body detection.

Typically, they use several available datasets for fall detection, use the same location for detection and training. Thus, it does not allow the use of methods to change the position between the training data and the data for testing. To gauge this reliability, the dataset contains videos from different locations that allow multiple assessment protocols to be defined (Home, Coffee Shop, Office, and Lecture Hall).

It was visually determined that most of the videos are around 300 frames in length. Therefore, it was decided to select data with a length of no more than 350 frames and bring all records to a length of 300 values. Longer recordings were truncated to a size of 300, and shorter ones were supplemented with a repetition of the twentieth value from the end, to preserve the dynamics in the recording values.

It should be noted that out of 191 videos, only 112 were annotated and only 87 passed the validation test. Out of 87 records, 13 records are “without falling” records, the rest are records “with falling”. After data transformations, there were 11 records “without falling”. The number of records “with falling” and “without falling” had to be balanced, and after that there were 23 records left for training.

In annotation records, the first two lines indicate the presence or absence of a drop in the record. Thus, if the first number in the record is zero, then the record refers to records without falling. In other cases, the fall is present.

After determining the label of the record and bringing to a single size, all record values complement the object containing the values of all records. After performing labelling of all records, the object with all values is exported to a file for further work with data.

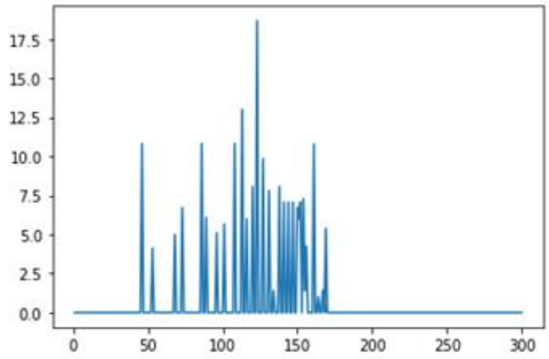


Fig. 1. The speed of the frame movement for each frame

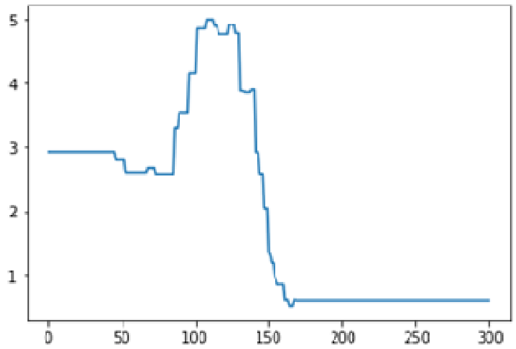


Fig. 2. *HeightWidthRatio* values interpolation

Further, for each record and each frame, two new features were calculated: *HeightWidthRatio* - the value of the aspect ratio of the bounding box,

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

converted from the lengths of the frame sides, and *BoundingBoxSpeed* - the speed of the frame movement for each frame, converted from the frame coordinates, Fig. 1. *HeightWidthRatio* values greater than 10 were replaced by 0 and later have been reconstructed using interpolation, Fig. 2.

B. Model setup

First, experiments were carried out using the LSTM neural network. The length of the LSTM input sequence was 300 elements, training was carried out on selected data (as described above) using the Adam optimization method, with control of the loss function based on binary cross-entropy. As a result, a trained neural network based on cross-validation data using batch normalization gave an average accuracy of 0.83.

Next, we used the second approach, described in the previous paragraph, to generate new significant features for the *HeightWidthRatio* and *BoundingBoxSpeed* time series using the tsfresh library. As a result, a dataset of 23 by 64 dimensions was obtained, where 23 is the number of records, 64 is the number of features, 1st Model Table II. Both 'linear' and 'rbf' SVM achieved train and test accuracy 1.0, but a 'linear' classifier is preferable, since the number of support vectors in both classes equals 1, which indicates a good generalizing ability of this model. The decision tree unmistakably classified the data according to only one feature X_0 - means the average, absolute value of consecutive changes of the series *HightWidthRatio* inside the quantile corridor [0.2, 0.8], Fig.3. Further, leaving only the X_0 feature, similar results were obtained, 2nd Model Table II.

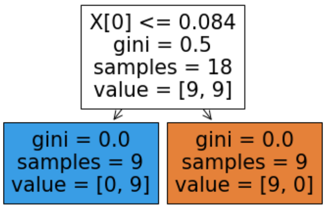


Fig. 3. Decision Tree classifier

TABLE II. EXPERIMENT RESULTS FOR FALL DATASET [9]

Trained Models	**Models Accuracy**			
	1st Model		**2nd Model**	
	Train acc.	Test acc.	Train acc.	Test acc.
SVM, 'linear' **n_support**	1.0 [2, 3]	1.0	1.0 [1, 1]	1.0
SVM, 'rbf' **n_support**	0.86 [5,3]	1.0	1.0 [3, 1]	1.0
Decision Tree	1.0	1.0	1.0	1.0
Random Forest	1.0	1.0	1.0	1.0

Analysis of more complex algorithms using the detection of key points of a person and training classifiers to fix the fall of the elderly [11] has not been carried out, since this technique requires rather high computational costs when processing video recordings.

IV. CONCLUSION

A method for controlling the fall of the elderly with the application of computer vision procedures has been proposed. The use of neural networks (LSTM structure) for the fall determination allowed to achieve 83% of the average classification accuracy.

The result obtained can be mainly explained by a rather small amount of the training data, remaining after the preparation phase. The approach to the fall control, using the full set of secondary parameters obtained with tsfresh library, allows to perform a linear separation of classes and achieve accuracy and precision close to 100%. The application of the tsfresh-based technology with only 2 secondary features associated with the ratio of bounding box parts, covering the person, and its movement speed allows to separate classes linearly with an accuracy of more than 93%. The efficiency of the tsfresh-based technology was confirmed on samples from 2 datasets: UP-Fall detection dataset and ImViA fall detection dataset. It is obvious that the method proposed has a lower computational complexity and achieves no less accuracy in comparison with known approaches. Further it can be improved by increasing the number of video cameras and/or additional sensors for the multimodal detection. It should also be noted that the analysis was performed on the abbreviated, normalized and balanced part of the ImVia dataset.

ACKNOWLEDGMENT

The work was partially supported by the COST Action CA16226 project. The authors thank Maxim Stupaktvich for participating in computational experiments.

REFERENCES

[1] K. Giannakouris, et all. "Ageing Characterises the Demographic Perspectives of the European Societies", EUROSTAT Statistics in focus, 2008. Available from: http://ec.europa.eu/eurostat.

[2] A. Ramachandran, R. Adarsh, P. Pahwa and K. Anupama, "Machine Learning-Based Fall Detection in Geriatric Healthcare Systems", 2018, IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Indore, India, 2018, pp. 1-6.

[3] B. Kwolek, M. Kepski, "Human Fall Detection on Embedded Platform Using Depth Maps and Wireless Accelerometer", Computer Methods and Programs in Biomedicine, 2014, 117(3), pp. 489-501.

[4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection",

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779-788.

[5] F. Lezzar, D. Benmerzoug, I. Kitouni, "Camera-Based Fall Detection System for the Elderly With Occlusion Recognition", Applied Medical Informatics, Vol. 42, 2020, pp. 169-179.

[6] H. Ponce, L. Martínez-Villaseñor, J. Brieva, and E. Moya-Albor, "Challenges and Trends in Multimodal Fall Detection for Healthcare": Springer, 2020, 259 p.

[7] tsfresh,https://tsfresh.readthedocs.io/en/latest/text/introduction.html, Accessed 21 Jun 2021.

[8] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing", Journal of the Royal Statistical Society: Series B, 57, pp. 289–300.

[9] F. Franchetti and M. Püschel, "FFT (Fast Fourier Transform)", In: D. Padua (eds) Encyclopedia of Parallel Computing: Springer, Boston, MA, 2011, pp. 658-671.

[10] "Fall detection dataset", ImViA, https://imvia.u-bourgogne.fr/en/database/fall-detection-dataset-2.html, Accessed 21 Jun 2021.

[11] O. Seredin, A. Kopylov and E. Surkov, "The study of skeleton description reduction in the human fall-detection task", Computer Optics, Vol. 44, 2020, pp. 951-958.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

GAN-SSL Classification for Identification Expertise in Chemistry

Aleksandra Maksimova
Theory Control System Department
Institute of Applied Mathematics and Mechanics
Donetsk, DPR
maximova.alexandra@mail.ru

***Abstract.* In this work we investigate the generative adversarial nets for classification problem of identification expertise in Chemistry. The identification expertise problem is challenging for classification because of complex structure of classes, outliers and cross-classes. The generative-adversarial nest for semi-supervised learning (GAN-SSL) is proposed for complex classification problem. The training samples are partially labeled for the semi-supervised tasks. Two groups of experiments were carried out. The first group of experiments for the model dataset that consist of classes of points normally distributed about vertices an eight-dimensional hypercube. The second groups of experiments for the petrol identification expertise dataset we get from laboratory of petrol quality. The experiments with good model examples get good quality more than 99%. The classification model for petrol identification expertise was created and has 93% quality but convergences training much worse. In this work we use GAN-SSL classification on petrol identification expertise example, but this classification model can be used for diesel fuel, household chemicals items, different oils and for various other objects.**

***Keywords:* GAN, classification, identification expertise, semi-supervised learning**

I. INTRODUCTION

Proposed in 2014 by Jan Goodfellow [1] generative adversarial nets in [2] were improved for classification problems using semi-supervised learning. In semi-supervised learning is used unlabeled data for end-to-end learning of classifiers. The training samples are partially labeled for the semi-supervised tasks.

The identification expertise problem in Chemistry is formally seems as a classification problem [3]. In our previous works was proposed fuzzy portrait method for identification expertise of petrol [4], [5]. The quality of received models heavily dependent on available data. The specific problem for identification expertise is a very few of items for some classes, for example 15 items. If we consider objects of each class as elements of a certain probability distribution, we don`t have enough items to get good classifier. The representation of samples set is not good enough. The main idea is to use semi-supervised learning to overcome our limitations with datasets using generative adversarial nets for semi-supervised learning (GAN-SSL). The work is aimed at research of novel state-of-the-art GAN-SSL classification method.

II. RELATED WORKS

GAN publications have increasingly focused on the use of class labels. The first multiclass inference strategy for GAN was developed in [6], where the number of outputs of the discriminant classifier is equal to the number of classes, and training is carried out both on unmarked and partially marked data. Such a network is called categorical generative adversarial network (CatGan).

The most interesting for identification expertise classification problem is proposed in [2], [7] classification model. The number of outputs of the discriminator corresponds to the number of real classes and one more for the fake class, produces by generator. This strategy is good working for semi-supervised learning using the GAN loss functions.

In [8] proposed novel approach to semi-supervised learning on graphs – GraphSgan. Generator and classifier play a novel competitive game, when generator generates fake samples. This idea can help in identification expertise to find counterfeit items.

There are some papers where classical GANs architectures, like DCGAN and PGGAN using for classification [9]. The generator is training to produce realistic chest X-ray images and lymph node histology images. These images add to training data sets for classical convolutional net.

III. MODEL FRAMEWORK

A. Problem Definition

We consider petrol identification expertise as classification problem. We should identify the petrol producer and mark by a sample of petrol with solving the classification problem.

We have in our consideration ten classes of objects grouped by producer and mark parameters. We use eight features: research octane, motor octane number, density, volume fraction of olefinic hydrocarbons, benzene, toluene and mass fraction of methanol and of oxygen.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The identification expertise problem belongs to sensitive to negative examples problem. The reason is the frequent cases of falsification of identification objects, when the composition of the product partially changes when using cheaper substitutes and additives.

B. *GAN-SSL Architecture*

We used GAN-SSL architecture proposed in [2]. GAN architecture includes two networks: generator and discriminator neural nets. The goal of model training is to train a generator $G(z)$ that produces samples from the data distribution $p_{data}(x)$ by transforming vector of noise z as $x = G(z)$. The discriminator is training to distinguish real data from the generator distribution $p_{data}(x)$. For GAN-SSL architecture discriminator is changed to standard K-classes classifier. We do semi-supervised learning with any standard classifier by adding samples from the GAN generator to data set and add $K+1$ class to classifier for these samples labelled like “generated fake”.

Both discriminator and generator in GAN-SSL network are multiple layer perceptrons. The generator takes noise z from uniform distribution on the interval [0,1) as input and outputs fake samples having the similar shape as x. Batch normalization is used in generator [10]. It is used weight normalization trick [11] before output layer in generator. There are three layers in the generator.

Discriminator consists of six linear weight norm layers based on weight normalization trick too. It is used additive Gaussian noise in every layer before output for smoothing purpose in the training mode only.

The discriminator takes in object feature vector x as input and outputs K-dimensional vector of logits $\{l_1, ..., l_K\}$ and one more input for “generated fake” class. Then we can use softmax activation function to get class probabilities $p_{model}(y=j|x)=\exp(l_j)/\mathrm{sum}(\exp(l_k))$. In practice, we only consider the first K outputs and assume the output for “generated fake” class is always 0 before softmax, because subtracting identical number from all units before softmax does not change the softmax results.

The number of neurons in hidden layers depends on identification expertise dataset and can be modified corresponding to velocity of problem. The discriminator must be more powerful than the generator and powerful enough for specific task. Denote the base number of neurons as N_{base} for experimental result section. It means, that there are N_{base} neurons in layers of generator and $2N_{base}$, $2N_{base}$, N_{base}, N_{base}, N_{base}, respectively in layers of discriminator.

The optimal discriminator in GAN-SSL is expected to be perfect on labeled and unlabeled data, but the generator will be always imperfect [12].

C. *Learning Algorithm*

There are two techniques to improve the training of GANs proposed in [2]. We use feature matching technique that addresses the instability of GANs by specifying a new objective for generator that avert it from overtraining on the current discriminator. We don`t use minibatch discrimination in our work because it further improves the generator examples that is not necessary for identification expertise problem.

The loss function for discriminator consists of two components: supervised $L_{supervised}$ and unsupervised $L_{unsupervised}$ loss functions [2]:

$$L_{supervised}=\mathbf{E}_{x,y\sim pdata(x,y)}\log[D(x)], \qquad (1)$$

$$\begin{aligned} L_{unsupervised}=-\mathbf{E}_{x\sim pdata(x,y)}\log[D(x)] - \\ -\mathbf{E}_{z\sim noise}\log[1-D(G(z))], \end{aligned} \qquad (2)$$

where $\mathbf{E}_{x,y\sim pdata(x,y)}$ is expectation of labeled data, $\mathbf{E}_{x\sim pdata(x,y)}$ is expectation of unlabeled data, $\mathbf{E}_{z\sim noise}$ is expectation of noise, $D(x)$ is output of discriminator applying softmax, $G(z)$ is output of generator.

The output of discriminator layer before softmax we denote as $\boldsymbol{f}(x)$ function that uses in new objective function for generator:

$$L_{gen}=||\mathbf{E}_{x\sim pdata(x,y)}\boldsymbol{f}(x)-\mathbf{E}_{z\sim noise}\boldsymbol{f}(G(z))||_{L2}, \qquad (3)$$

where $\mathbf{E}_{x\sim pdata(x,y)}$ and $\mathbf{E}_{z\sim noise}$ like in (2), $||f||_{L2}$ is L_2 norm.

We use minibatch stochastic **Algorithm 1** to train generator and discriminator iteratively minimizing their losses.

Algorithm 1: Minibatch stochastic gradient descent training of GAN-SSL for identification expertise.

```
Input:  $X^1_{unlabeled}$- identification objects dataset – shuffled unlabeled data set #1
        $X^2_{unlabeled}$- identification objects dataset – shuffled unlabeled data set #2
        $X_{labeled}$- identification objects dataset for supervised learning; $y$ – labels of classes correspond to pare (producer, mark of petrol)
Make $X^1_{unlabeled}, X^2_{unlabeled}, X_{labeled}$ equal length datasets by folding $X_{labeled}$
for number of epochs do
        Sample minibatch from $X^1_{unlabeled}$
        Sample minibatch from $X_{labeled}$
        Sample minibatch from noise prior $p_g(z)$
        Update the discriminator by descending gradients of losses:
                $L=L_{supervised}+L_{unsuprvised}$
        Sample minibatch from $X^2_{unlabeled}$
        Sample minibatch from noise prior $p_g(z)$ as $G(z)$
        for 2 steps do
```

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

```
            | Update the discriminator by descending
            | gradients of losses:
            | $L = L_{gen}$
        end for
    end for
```

The gradient-based updates can use any standard gradient-based learning rule. We use Adam, based on adaptive estimates of lower-order moments [13].

We train the model for different power of the generator and discriminator adjusting base neuron parameter N_{base}.

IV. EXPERIMANTAL RESULTS

We use GAN-SSL for two classification problems with the same numbers of features and classes to compare and analyze the speed and quality of training GAN-SSL. First one is model examples, but second is real petrol identification expertise problem with cross-classes and outliers.

The model data set consist of classes of points normally distributed about vertices of eight-dimensional hypercube. These data have the same objects in every class and easy for classification. The total number of objects is 5000, of which 100 labeled examples of an equal number from each class.

The visualization of training for experiments with model dataset is presented in Fig. 1. The x-axis shows the number of epochs. The model 1 dataset consists of 10 normally distributes classes, and the model 2 consist of classes with two normally distributed clusters for every class. The 1% of object is outliers for every models. The number of neurons N_{base} in the third example “weak net” is 15, in the rest – 25.

The petrol identification expertise data we get from laboratory of petrol quality. There are different number of objects in classes from 72 in smallest to 671 in the most popular. We use only 100 labeled examples for training. The total number of objects is 6710, of which 100 labeled examples of an equal number from each class.

We train three GAN-SSL networks with different N_{base} parameter for network architecture: 25, 35 and 45. The visualization of training is presented in Fig. 2. The x-axis shows the number of epoch.

The experiments with good model examples get good quality more than 99%. The training process for the model data was fairly stable, without sharp fluctuations. Expectedly, training was fastest on model 1 data. The GANs with N_{base} equal to 25 train better than for weak net, where the base number of neurons is equal to 15. The generator loss L_{gen} grows corresponding to imperfect quality of generator, which was substantiated in [12].

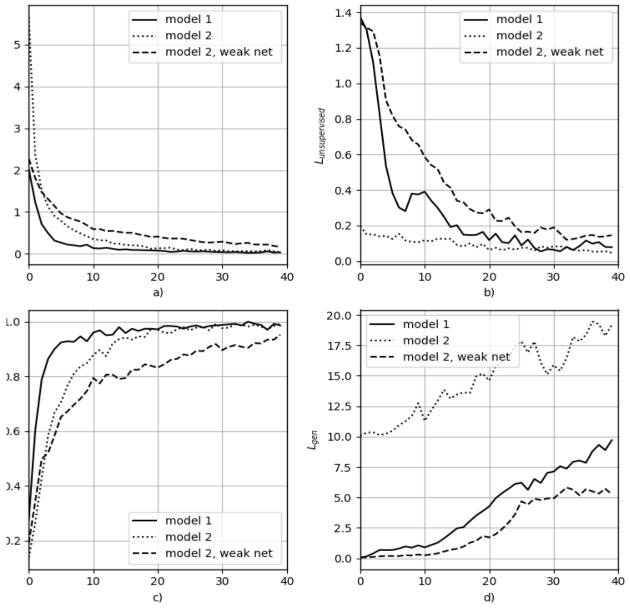


Fig. 1. Visualisation of training process for model datasets: simple model 1 with 10 normal distributed classes, model 2 has two normal distributes clusters in every class: a) supervised loss (1), b) unsupervised loss (2), c) validation of model with respect to testing data, d) generator loss (3)

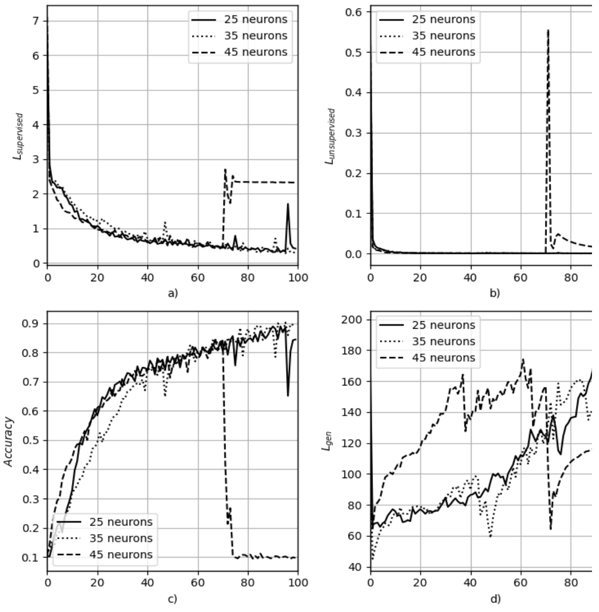


Fig. 2. Visualisation of training process for petrol dataset for different N_{base}:25, 35 and 45 a) supervised loss (1), b) unsupervised loss (2), c) validation of model with respect to testing data, d) generator loss (3)

But for heavy practical real-world examples training GAN-SSL is very difficult. This is due to one big problem for GANs: they convergence very unstable. We can`t get more than 93%.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The training was performed for a different base number of neurons N_{base}: 25, 35 and 45. The power of the generator and the discriminator must be selected to be the best for the problem of a given dimension, since it can be seen that an overly complex neural network can train unstable. The base neuron parameter for petrol identification problem is selected as 35 neurons.

The total experiments result presents in Table I.

TABLE I. EXPERIMENTAL RESULTS

Base number of neurons	Accuracy	Epochs
Model 1 dataset		
25	0.999	30
Model 2 dataset		
15	0.977	52
25	0.997	27
Petrol dataset		
25	0.883	200
35	**0.934**	**150**
45	0.906	300

The good situation if we have half of labeled data in whole quantity. But the part of labeled data in data set can be small enough if there are no good labeled objects in consideration.

V. CONCLUSIONS

In this research we investigate the GAN-SSL performance for identification expertise classification problem. This work has showed that GAN-SSL classifiers converge quickly and have good model quality for good normal distributed classes with the same number of examples. The classification model for petrol identification expertise was created and has 93% quality but convergences training was much worse. The number of neurons N_{base} needs to be adjusted.

The identification expertise problem is challenging for classification because of complex structure of classes, outliers and cross-classes. In future work we plan to use the “generated fake” examples to generate missing data to reconstruct certain probability distribution p_{data} of difficult for classification classes.

In this work we use GAN-SSL classification on petrol identification expertise example, but this classification model can be used for diesel fuel, household chemicals items, different oils and for various other objects.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, Sh. Ozair, A. Courville, Y. Bengio, “Generative adversarial nets,” Advances in Neural Information Processing Systems 27. Curran Associates, Inc., 2014. pp. 2672–2680.

[2] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, “Improved techniques for training GAN,” Advances in Neural Information Processing Systems, 2016, pp. 2234–2242.

[3] A. Maksimova, “Formal statement of the problem of identification expertise, ” Donetsk Readings 2017: Russian World as a Civilizational Basis for Scientific, Educational and Cultural Development of Donbass, Donetsk, pp. 69–70.

[4] A. Maksimova, “The approach to the construction of information automated systems of identification examination based on machine learning methods,” International scientific and technical congress Intellegent systems and information technologies”, Taganrog, 2017, Vol. 1, pp. 438–443.

[5] A. Maksimova, “Fuzzy approach to solve pattern recognition problem for automatisation system for identification expertise (example for petrol identification), ” International scientific conference Computer Science and Information Technology, Saratov, 2016, pp. 256–258.

[6] J Springenberg, “Unsupervised and semi-supervised learning with categorical generative adversarial networks, ” 2016, URL: https://arxiv.org/abs/1511.06390.

[7] D. Kingma, D. Rezende, S. Mohamed, M. Welling, “Semi-Supervised Learning with Deep Generative Models Proceedings of the International Conference on Machine Learning, ”, 2014, pp. 3581–3589.

[8] M. Ding, J. Tang, J. Zhang, “Semi-supervised learning on graph with generative adversarial nets, ” ACM int. conf. on information and knowledge management, 2018, pp. 913–922.

[9] V. Kovalev, S. Kazlouski, “Examining the capability of GANs to replace real biomedical images in classification models training, ”

[10] S. Ioffe, Ch. Shegedy, “Batch normalisation: accelerating deep network training by redusing interval covariance shift,” 2015, in ICML`15, pp. 448–456.

[11] T. Saliman, D. Kigma, “Weight normalisation: a simple reparametrization to accelerate training of deep neural networks,” in NIPS`16, pp. 901–909.

[12] X. Liu, X. Xiang, “How does GAN-based semi-supervised learning work?”, 2020, URL: https://arxiv.org/abs/1809.00130.

[13] D. Kigma, J. Ba, “Adam: a method for stochastic optimization,” 2015, Computer science, mathematics, n.pag.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Contextualizing of Architectural Security Patterns as a Knowledge Management Challenge

Andrei Brazhuk,
Yanka Kupala State
University of Grodno
Grodno, Belarus

Evgeny Olizarovich
Yanka Kupala State
University of Grodno
Grodno, Belarus

Abstract. **Security-by-design as adoption of security solutions for a system design is in focus of this work. This field is treated as requiring expert knowledge and heavy for automation. A perspective way to improve exiting security design methodologies is the use of security patterns as a mechanism of collecting secure design artifacts. To apply security patterns as a part of automation of secure design, it requires well-formed collections of security patterns and innovative method to support the design decisions.**

This work considers a contextualizing challenge as a way to define the necessity of a security pattern in a given case. Understanding of context includes two main questions: "Is the security pattern suitable for a system design?" and "Does the security pattern affect a particular security challenge?".

We approach a direct architectural contextualizing as a basic mechanism of automatic mapping of security artifacts (threats, security solutions) to components of a computer system during early design stages (requirements, design). Also, this work describes two use cases of the architectural contextualizing based on an ontological cloud threat pattern catalog: the use of a query language for finding relevant security patterns and analysis of graphical system representations based on an ontology driven threat modeling.

This work uses a strict ontological approach, implemented with Web Ontology Language (OWL) and automatic reasoning procedures.

***Keywords*: security pattern, ontology, contextualizing, threat modeling, OWL**

I. INTRODUCTION

Security methodologies solve various challenges of secure development by improving security attributes of computer systems [1]. They describe security as a set of processes (threat modeling, risk management, secure design, etc.) and operate different artifacts (threats, controls, mitigations, metrics etc.) of conceptual security models.

Security-by-design is in focus of this work, i.e., adoption of security solutions to a particular design. This is commonly considered as an informal field, required expert knowledge, and it is most challenged from the automation point of view. They need well-formed collections of artifacts, also methods and algorithms to take right decisions.

Security patterns are known as a way of representation of various security artifacts (especially holding architectural decisions in some form) and reusing them. They are important in improvement efficiency of security methodologies: making the process iterative, and integration of the threat modeling and the secure design subprocesses. The common approach is to use different artifacts at early lifecycle stages: use cases and abstract security patterns at the requirements stage, and threat taxonomy and concrete security patterns at the design stage. Note, dealing with threats can also be possible with a special kind of security patterns, called threat patterns.

A security pattern is considered as a class, and applying it to a design is called 'instantiation'. Having a description (texts, diagrams, artifacts) of a system with flaws and vulnerabilities, it requires to correct items of the description from security perspective, injecting adequate security patterns. Instantiation as a complicated process is out of scope of this work, as well as its possible supplementary processes like integration and verification [2].

All the things that define the necessity of the pattern in the design we call 'contextualizing' in this work. Understanding of context includes two main questions: "Is the security pattern suitable for the system design?" and "Does the security pattern affect a particular security challenge?".

Contextualizing can be done manually (semi-automatically) or automatically.

Manual (semi-automatic) use case is like: an architect works with a system architecture, depicted as text or as a graphical notation, like UML. To help the architect to choose a pattern (or an ordered set of patterns) out of several hundred existing, a security pattern catalog can be used. The catalog can contain different labels and some sort of a query language can be used to find relevant patterns.

Automatic use case is based on a formal scenario that describes a computer system in general, the

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

scenario is used to build the system automatically by an orchestration system software. To analyze security aspects of such applications it could be useful apply automation, which allows to correct the deployment scenario by automatic implementation of a relevant security pattern, or automatically check dependent components for meeting of requirements of current application's SLA. Advanced knowledge management techniques and the Artificial Intelligence (AI) technologies should be used to implement the automatic use case.

This work considers essential items of contextualizing, and contributes by a description of a direct architectural contextualizing as a basic mechanism of automatic mapping of security artifacts (threats, security solutions) to components of a computer system during early design stages. It is a part of an ontological schema of security patterns, based on a strict ontological approach and implemented with Web Ontology Language (OWL) and automatic reasoning procedures.

Also, this work describes two use cases of the architectural contextualizing based on an ontological cloud threat pattern catalog. The first use case shows the usage of query language for finding relevant threat patterns. The second one depicts the analysis of a simple graphical system representation with data flow diagrams (DFD) based on an ontology driven threat modeling.

The rest of this work is structured as follows. Section II shows related works in the threat modeling and security pattern fields. Section III describes the direct architectural contextualizing. Section IV illustrates its use cases. And a summary of the results and future research is discussed in Conclusions.

II. RELATED WORK

Traditionally, threat modeling is considered as a semi-automatic process, happening at early stages (requirements, design) of system lifecycle. A challenge is to increase its automation part, because this gives opportunities to enable automatic contextualizing. Also, moving of threat modeling to run-time with approaches like reflective threat modeling [3] is in research focus now, these will require advanced knowledge management techniques. That is another reason for importance of the contextualizing challenge in automatic threat modeling.

Existing efforts of the contextualizing of common security knowledge are primary based on rule-based languages, graphs, domain specific languages (DSL), logics, and ontologies.

Work [4] has proposed to use a catalog of security patterns to automatically detect vulnerabilities in a software architecture. Their patterns have been defined with a graph language, and context should be determined by appropriate query rules. Several works [5, 6] are known as a continuation of use of rule-based languages to automate threat modeling.

Works [7, 8] have described efforts to apply a meta language for creation of domain specific languages to depict security challenges and its enhancement for cyber-attack scenarios, in particular with probability distributions. They have used attack graphs as implementation.

Work [9] has applied a bit of logic-based approach with Prolog based rules to define context.

Also, there are works that have described an ontological approach of conceptual structures of security frameworks [10, 11], in particular focused on the threat modeling [12, 13].

From our perspective, ontologies based on strict formalization (e.g., OWL and automatic reasoning) most meet the automation challenges of the threat modeling. Work [14] has described a framework of ontology driven threat modeling that potentially supports contextualizing based on security labels (CIA, STRIDE) and architecture.

Despite collecting the security patterns for decade [15, 16], several challenges of their use exist, like lack of approaches to recognize necessity of security patterns for a computer system design. There are diverse researches aimed at formalizing security patterns and creation their catalogs, and most of patterns are represented by the UML diagrams with text descriptions in the POSA format (at least it requires to fill the context, problem and solution fields). So, it can be considered two directions of research, related to contextualizing: both transformation of the diagrams and ordering of the textual meta information into knowledge-like formats (graphs, ontologies, etc.).

Work [17] has approached a security pattern classification, based on transformation of the UML descriptions to the Attack Deference Trees (ADT). Work [18] has described a security pattern detection framework in order to put them in a security pattern graph database.

Work [19] has proposed a modeling language to define security patterns based on metamodeling techniques. And work [20] has described a conceptual approach of interconnecting various pattern languages.

There are several researches, aimed to classify metadata of security patterns [21], up to creation of ontology-driven tools [22].

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

And several catalogs [23, 24] and domain specific models [25, 26] of security patterns have been developed.

However, it can be argued that lack of efforts exists in order to unite the threat modeling technologies and security pattern approach, also apply different automatic techniques and methods into the secure development process. And resolving of the contextualizing challenge can be a step forward in this field.

III. DIRECT ARCHITECTURAL CONTEXTUALIZING

Direct architectural contextualizing is a part of the ontological schema of security patterns [22]; the schema also represents common features of patterns, like idea, author, type, their hierarchy and relationship, and a set of characteristics used by the scientific community.

The ontological schema, implemented with OWL, allows creation of well-formed catalogs of security patterns. From architectural point of view a catalog includes two items.

First item is a metamodel of given domain with a hierarchy of typical components, implemented by the class-subclass relationship (for example, our cloud specific catalog, described below, includes components like Cloud Infrastructure, Cloud Application, Remote User; and Cloud Applications can be divided to Virtual Machines, PaaS Applications, and SaaS Applications).

Second item includes strict descriptions of security patterns with the security and context labels.

Note, architectural context is independent from pattern type, so it can be possible to apply the same approach to the security patterns, misuse patterns and threat patterns.

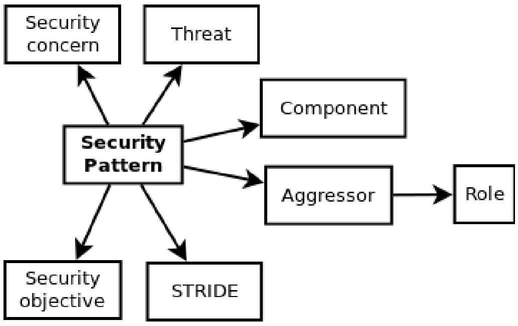


Fig. 1. Structure of direct architectural context

For example, it is considered interaction between remote user and cloud application (see Fig. 2) and it requires to put into context a security pattern that affects some destructive activity from remote user. In this case remote user is treated as an aggressor, and cloud application is an affected component. Additionally, it can be possible to apply role of the aggressor (client) in order to keep information about direction of the network connection.

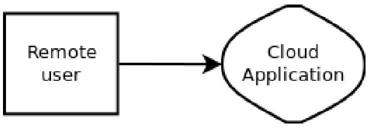


Fig. 2. Context example

And, in addition, it can be possible to apply security characteristics of the pattern. In general, this includes two points of view (Fig. 1): view of a security expert, represented by the threat concept, and view of an architect, represented by the security concern concept.

Depending of modeling goals different approaches can be used to define security labels. Threat taxonomy can be built from the CAPEC enumeration, what gives capability to map them with the ATT&CK, CWE, CVE enumerations [27, 28]. or use one of several original approaches [29, 30].

Security concerns are considered as security features that a security pattern holds in terms of software requirements. It can be useful to consider security control families from the NIST SP 800-53 publication as security concerns. In theory this enables mapping of security pattern catalogs with different security control catalogs [31, 32].

Also, the common security labels are used via the security objectives (CIA) and STRIDE concepts (Fig. 1).

Having a set of architectural (component, aggressor, role) and security (concern, threat) labels, it can be possible to make requests to a catalog in order to find relevant patterns. Also, strict contextualizing enables automatic procedures of comparing design templates, created from the architectural labels, and items of system description, made in some graphical or text notation, in order to define applicability of a security pattern.

IV. CLOUD COMPUTING USE CASE

Currently, creation of a security pattern catalog includes two stages. Firstly, an ontology, based on the schema [22], should be created to describe concepts and instances of a specific computing environment. Then a JSON-schema file from the ontology can be generated by a simple tool. Secondly, pattern descriptions can be created as JSON files with a JSON-schema based editor. The simple tool allows generating of an ontology of security patterns from the pieces of JSON.

We have been developing the Academic Cloud Computing Threat Patterns (ACCTP) catalog [33] (https://nets4geeks.github.io/acctp/) to research

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

feasibility of the ontological approach of management of security patterns. Threat pattern catalog is a kind of security pattern catalog, intended to collect architectural threats of a particular domain and used for analysis of security use cases and creation of a threat taxonomy. Our implementation of ACCTP has also included references to the common cloud security solutions to make it more close to the real security challenges.

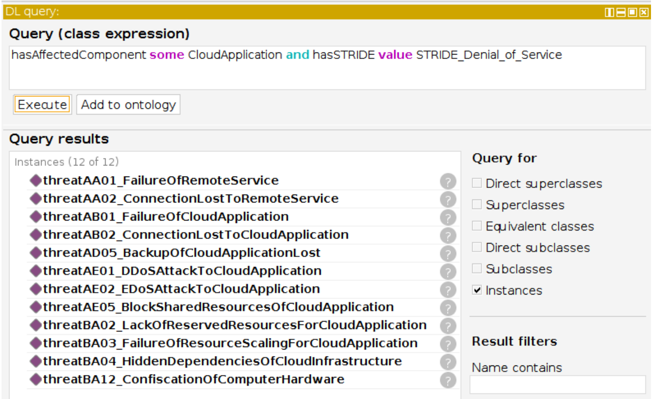
A. *Finding relevant threat patterns*

To illustrate this use case, we use the Protege ontology editor and the DL query language. For example, to show all the threats the Cloud Application concept can be affected, you can use a DL sentence like:

hasAffectedComponent some CloudApplication

To apply the STRIDE filter to the previous query, you can use request (its results are shown in Fig. 3):

hasAffectedComponent some CloudApplication and hasSTRIDE value STRIDE_Denial_of_Service



hasAffectedComponent some CloudApplication and hasSTRIDE value STRIDE_Denial_of_Service

Instances (12 of 12)

- threatAA01_FailureOfRemoteService
- threatAA02_ConnectionLostToRemoteService
- threatAB01_FailureOfCloudApplication
- threatAB02_ConnectionLostToCloudApplication
- threatAD05_BackupOfCloudApplicationLost
- threatAE01_DDoSAttackToCloudApplication
- threatAE02_EDoSAttackToCloudApplication
- threatAE05_BlockSharedResourcesOfCloudApplication
- threatBA02_LackOfReservedResourcesForCloudApplication
- threatBA03_FailureOfResourceScalingForCloudApplication
- threatBA04_HiddenDependenciesOfCloudInfrastructure
- threatBA12_ConfiscationOfComputerHardware

☐ Direct superclasses
☐ Superclasses
☐ Equivalent classes
☐ Direct subclasses
☐ Subclasses
☑ Instances

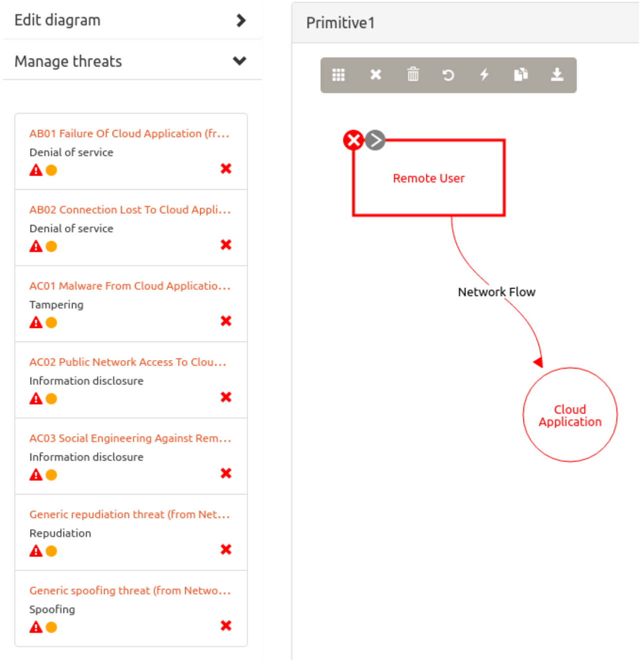
Fig. 3. Example of DL query

B. *Analysis of graphical system description*

Several graphical notations, like data flow diagrams (DFD) and process flows exist that be used to apply flow-based threat modeling.

In order to enable an ontological approach of DFD analysis, we have converted the ACCTP ontology to an appropriate ontological domain specific threat model [33], we have created a console modeling tool, and adopted third-party GUI threat modeling tool (OWASP Threat Dragon). The ontologies of a base threat model, the ACCTP domain specific threat model, and semantic interpretation of a diagram should be processed by automatic reasoning procedures to get a list of threats for a given system description.

Fig. 4 shows threats that touch remote cloud users, interacting with a cloud application. Threats are taken the catalog by the automatic reasoning procedures, and this 'automatic' decision is based on a sort of an ontological flow template that catches such kind of interactions.



AB01 Failure Of Cloud Application (fr…
Denial of service

AB02 Connection Lost To Cloud Appli…
Denial of service

AC01 Malware From Cloud Applicatio…
Tampering

AC02 Public Network Access To Clou…
Information disclosure

AC03 Social Engineering Against Rem…
Information disclosure

Generic repudiation threat (from Net…
Repudiation

Generic spoofing threat (from Netwo…
Spoofing

Fig. 4. Results of automatic threat modeling

V. CONCLUSION

This work contributes the direct architectural contextualizing as a basic mechanism of automatic mapping of security artifacts (threats, security solutions) to components of a computer system during the requirements and design stages. It can be possible strictly define a context of a security (threat) pattern with proposed properties. However, direct contextualizing is a naive approach that can be used as a proof of concept and for restricted use cases of semi-automatic secure design. To extend it, advanced knowledge management and decision support (AI-like) technologies are needed. Additionally, the challenge of automatic instantiation should be considered because strict contextualizing depends on such a challenge.

Also, this work describes the use cases of the contextualizing based on an ontological cloud threat pattern catalog (use of a query language and ontology driven threat modeling of data flow diagrams). Note, industry adoption of the use cases requires creation of software tools and modules, and redesign of existing security methodologies in order to integrate threat modeling and design itself (DFDs are not considering as a good design approach).

Mentioned above challenges form possible directions of future research.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

REFERENCES

[1] A.V. Uzunov, E.B. Fernandez, K. Falkner, "ASE: A comprehensive pattern-driven security methodology for distributed systems." Computer Standards&Interfaces 41, 2015, pp. 112-137.

[2] T. Peng, S. Wang, J. Geng, Q. Wang, "Verification of the Instantiation and Integration of Security Patterns." Journal of Web Engineering , 2020, pp. 521-556.

[3] D. Van Landuyt, L. Pasquale, L. Sion, "Threat models at run time: the case for reflective and adaptive threat management (NIER track)," SEAMS'21: Proceedings of the 16th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, 2021.

[4] B.J. Berger, K. Sohr, R. Koschke, "Automatically extracting threats from extended data flow diagrams," International Symposium on Engineering Secure Software and Systems. Springer, Cham, 2016.

[5] S. Peldszus, "Model-driven Development of Evolving Secure Software Systems," Software Engineering (Workshops), 2020.

[6] K. Tuma, at. al., "Automating the early detection of security design flaws," Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems, 2020, pp. 332-342.

[7] S. Katsikeas, et al., "An attack simulation language for the IT domain," International Workshop on Graphical Models for Security, Springer, Cham, 2020, pp. 67-86.

[8] W. Xiong, S. Hacks, R. Lagerström, "A Method for Assigning Probability Distributions in Attack Simulation Languages," Complex Systems Informatics and Modeling Quarterly, 2021, vol. 26, pp. 55-77.

[9] S. Hahner, et al., "Modeling Data Flow Constraints for Design-Time Confidentiality Analyses," 2021 IEEE 18th International Conference on Software Architecture Companion (ICSA-C), 2021, pp. 15-21.

[10] R.Y. Venkata, et al., "A Domain-agnostic Framework for Secure Design and Validation of CPS Systems," International Journal on Advances in Security, Vol. 13, Num. 3 & 4, 2020.

[11] V. Vassilev, et al., "Intelligence graphs for threat intelligence and security policy validation of cyber systems," Proceedings of International Conference on Artificial Intelligence and Applications, Springer, Singapore, 2021, pp. 125-139.

[12] M. Välja, et al., "Automating threat modeling using an ontology framework," Cybersecurity, vol. 3, no. 19, 2020.

[13] A. Shaked, Y. Reich, "Model-based Threat and Risk Assessment for Systems Design," 7th International Conference on Information Systems Security and Privacy (ICISSP 2021), 2021, pp 331-338.

[14] A. Brazhuk, "Security patterns based approach to automatically select mitigations in ontology-driven threat modelling," Open Semantic Technologies for Intelligent Systems (OSTIS), 2020, pp. 267-272

[15] H. Washizaki, et al., "Systematic Literature Review of Security Pattern Research," Information, 2021, vol. 12, no. 1.

[16] A.J. Jafari, A. Rasoolzadegan, "Security patterns: A systematic mapping study," Journal of Computer Languages, 2020, vol. 56.

[17] S. Salva, L. Regainia, "An Advanced Approach for Choosing Security Patterns and Checking their Implementation," arXiv preprint, arXiv:2007.03275, 2020.

[18] A.K. Alvi, M.A. Zulkernine, "A security pattern detection framework for building more secure software," Journal of Systems and Software, 2021, vol. 171.

[19] B. Hamid, S. Gürgens, A. Fuchs, "Security patterns modeling and formalization for pattern-based development of secure software systems," Innovations Syst. Softw. Eng. vol. 12, no. 2, 2016, pp. 109–140.

[20] M. Weigold, et al., "Pattern Views: Concept and Tooling for Interconnected Pattern Languages," Symposium and Summer School on Service-Oriented Computing, Springer, Cham, 2020, pp. 86-103.

[21] M. VanHilst, et al., "A multi-dimensional classification for users of security patterns,". J. Res. Pract. Inf. Technol. vol. 41, no. 2, 2009, pp. 87–97.

[22] A. Brazhuk, E. Olizarovich, "Format and Usage Model of Security Patterns in Ontology-Driven Threat Modelling," Russian Conference on Artificial Intelligence, Springer, Cham, 2020, pp. 382-392.

[23] N. Marko, A. Vasenev, C. Striecks, "Collecting and Classifying Security and Privacy Design Patterns for Connected Vehicles: SECREDAS Approach," International Conference on Computer Safety, Reliability, and Security, Springer, Cham, 2020. pp. 36-53.

[24] M. Papoutsakis, et al., "Towards a Collection of Security and Privacy Patterns," Applied Sciences, 2021, vol. 11, no. 4.

[25] R. Saemaldahr, et al., "Reference Architectures for the IoT: A Survey," Innovative Systems for Intelligent Health Informatics. IRICT, Lecture Notes on Data Engineering and Communications Technologies, vol 72. Springer, Cham, 2021.

[26] C. Silva, et al., "Contract-based design patterns: a design by contract approach to specify security patterns," Proceedings of the 15th International Conference on Availability, Reliability and Security, 2020, pp. 1-9.

[27] K. Kanakogi, et al., "Tracing CAPEC Attack Patterns from CVE Vulnerability Information using Natural Language Processing Technique," Proceedings of the 54th Hawaii International Conference on System Sciences, 2021.

[28] C.B. ŞAHİN, "The Role of Vulnerable Software Metrics on Software Maintainability Prediction," Avrupa Bilim ve Teknoloji Dergisi, 2021, no. 23. pp. 686-696.

[29] A.V. Uzunov, E.B. Fernandez, "An extensible pattern-based library and taxonomy of security threats for distributed systems," Computer Standards & Interfaces, 2014, vol. 36, no. 4, pp. 734-747.

[30] A. Massel, D. Gaskova, "Identification of Critical Objects in Reliance on Cyber Threats in the Energy Sector," Acta Polytechnica Hungarica, 2020, vol. 17, no. 8.

[31] K.P. Joshi, L. Elluri, A. Nagar, "An Integrated Knowledge Graph to Automate Cloud Data Compliance," IEEE Access, 2020, vol. 8.

[32] V.I. Vasilyev, A.M. Vulfin, L.R. Chernyakhovskaya, "Cybersecurity Risk Analysis of Industrial Automation Systems on the Basis of Cognitive Modeling Technology," Digital Forensic Science, IntechOpen, 2019.

[33] A. Brazhuk, "Threat modeling of cloud systems with ontological security pattern catalog," International Journal of Open Information Technologies, vol. 9, no. 5, 2021, pp. 36-41.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Theory and Practice of Multi-Agent Systems Construction

Boris Zhalezka
Dept. of Marketing
Belarusian National Technical University
Minsk, Belarus
boriszh@yandex.ru

Volha Siniauskaya
Dept. of Industrial Marketing and Communications
Belarusian State Economic University
Minsk, Belarus
olechka_si@mail.ru

Abstract. **This paper generalized the theoretical bases and example of experience of construction and developing of multi-agent systems. Standards of multi-agent systems are considered. Overview of systems with mobile agents is made. Characteristic of JADE agent platform is given; its architecture is represented. A practical example of multi-agent application construction is shown.**

Keywords: **multi-agent system, mobile agent, artificial intelligence, Java Agent developing Framework**

I. INTRODUCTION

Investigations in the sphere of intellectual agents and multi-agent systems have a long history and nowadays became one of the intensive developing directions in artificial intelligence. However, interest for these investigations increased significantly in last ten years. Multi-agent systems are able to combine different elements of artificial intelligence, distributed information systems and computer networks. Multi-agent system is as a set of intelligent agents searching for data and procedures suitable for solving user's tasks and collaborating in the process of developing these solutions. Multi-agent systems include many different components with difficult architecture, variety of mathematical methods and software used, multiple variants of interaction between agents and variety of factors of the external environment in which agents operate.

The target of this research is to consider theoretical bases of multi-agent systems developing and to show practical example of multi-agent system realization by means of JADE platform.

II. THEORETICAL BASES OF MULTI-AGENT SYSTEMS

According to the work [1], the term “intellectual agent” arose primarily due to the need of simplifying of interaction between user and software. Later, the concept of “autonomous agent” appear, which means that instead of interacting with the program by calling commands and directly manipulating, the user has an ability to jointly solve the process. With this approach, the user and the computer agent jointly influence the event management and problem solving, and jointly launch it. In this approach, when the computer agent interacts with the user in one environment, the term "personal assistant" [1, 2] is used, which nowadays replaced with the term “intellectual agent”. The term “agent” has the following generalized definition: “An agent is a system that exists in a certain environment and it is part this environment. The agent influences the environment when performing its own tasks and this environment affects agent. Thus, changes made by the agent in the environment are reflected in this agent in future” [3].

Agents may have a number of properties [4]:

- autonomy – the ability to carry out the task without the need for external intervention;
- reactivity – the ability to perceive a change in the environment and take response actions;
- focus – the ability to carry out the assigned tasks;
- stability – the ability to restore its state after abnormal termination;
- sociability – the ability to interact with other elements of the environment;
- adaptability – the ability to change its behaviour according to experience and current circumstances;
- mobility – the ability to move in the environment;
- flexibility – the ability to change the own behavior.

With accordance of abovementioned, multi-agent system (MAS) can be characterized as a set of individual agents acting together and therefore characterized by a sociability property. Working MAS must include additional components, in particular, search service. For mobile agents functioning, special software components (platforms) are installed on the hosts, which support runtime environment and

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

interactive interface between agents and hardware. Mobile agents transfer between network hosts in the process of their execution, and interact with the hosts' resources and other agents. That needs solving additional problems in organization of mobile agent systems. Mobile agent system is a distributed application, which supports mobile agent work.

With some degree of conditionality, research in the field of multi-agent systems can be divided into these main areas: (1) theory of agents, in which formalisms and mathematical methods are used to describing agents and their properties; (2) methods of agents' cooperation (organization of cooperative behavior) in the process of joint task solving or at any other variants of interaction; (3) architecture of agents and multi-agent systems; (4) programming languages for agents creation; (5) methods, languages and tools of agents' communication; (6) methods and software for agents mobility support (agents migration over the network). Research related to the development of multi-agent applications and tools to support their development technology is of particular importance. Problems of authentication (authorization) of agents and their security are also urgent.

III. STANDARDS OF MULTI-AGENT SYSTEMS

There are dozens of systems using mobile agents. For their compatibility, specifications and standards are worked out. There are two main standards nowadays, which were created by Object Management Group и Foundation for Intelligent Physical Agents.

A. MASIF standard

Object Management Group Association elaborated Mobile Agent System Interoperability Facilities (MASIF) [4]. The main attention paid to the standatization of the following problems of mobile agent technology:

- agent management (programmer has a role of system administrator and manages the different mobile agents by means of standard operations and methods – to create, to pause, to resume and to finish agent);
- agent identification (standardized syntax and semantics of agent names and agent systems allow agent systems and agents to identify each other, and allow clients to identify agents and agent systems);
- typification and targeting of agent platform (an agent cannot be transferred to another platform if the agent system type cannot support the agent; location syntax is standardized so agent systems can find each other).

According to MASIF standard [5], mobile agent system is divided on the regions (Fig. 1). The region joins platforms with common permissions. It also includes a registry containing information about platforms and agents located in this region. To be compatible with the MASIF standard, it must implement the MAFFinder interface. It defines how agents, locations, and platforms are created, deleted, and transferred.

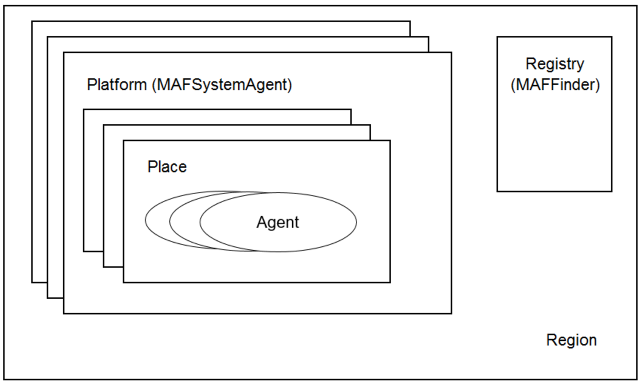


Fig. 1. Multi-agent system organization in according with MASIF standard

Platform (agent system) can create, interpret, start, transfer and delete mobile agents. To be compatible with the MASIF standard, it must implement MAFAgentSystem interface. It defines operations of receiving/transmitting, creating/deleting, interrupting/ resuming agents. The platform is identified by the name and address. The platform includes at least one location and a connection interface. The location provides the agent runtime environment on the computer. It can contain multiple agents at the same time. The connection interface implements a communication service, a name service, and a security service.

B. FIPA standard

Foundation for Intelligent Physical Agents (FIPA) – a non-profit organization established in 1996. Its main task is to develop specifications that determine the interaction of agents [6] and consider the following main topics:

- agent management (the architecture of agent platforms is unified, which include message routing and agent lifecycle management services);
- agent communication language (the syntax of a language, intended to interact between agents of different systems, is described);
- interaction with non-agent software (methods of interaction between agent and user as well as between agent and non-agent software are unified by means of shells including specified

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

ontology and dynamic mechanism of registrations);

- agent security management (key security threats in agent management are defined and possible facilities of protection are described);
- mobile agent management (transmission and start operations of agent platform are unified which are need for mobile agent management);
- ontology service (service of support of correct understanding of requests, messages, terms in the subject area context is described);
- application areas (applications using agent technology are considered, for example – network assistant, audio- and videoconferences; by means of them developed specifications are tested).

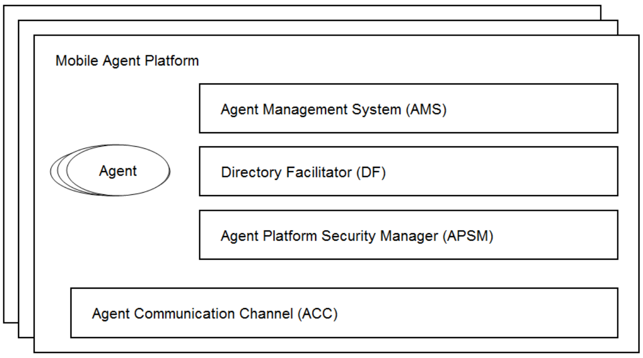


Fig. 2. Multi-agent system organization in according with FIPA standard

According to FIPA specifications, mobile agent systems consist of platforms with following components (Fig. 2):

- Agent Management System (AMS) manages the creation, removal, deactivation, restart, authentication, and migration of agents on the platform; it supports “white pages” service that stores the current location of mobile agents;
- Directory Facilitator (DF) realizes “yellow pages” service which stores agents description;
- Agent Platform Security Manager (APSM) is responsible for implementing the security policy at the transport level and verifying the execution of agent management operations;
- Agent Communication Channel (ACC) uses information from agent management system for routing messages between agents.

IV. SYSTEMS OF MOBILE AGENTS

There are several dozens of mobile agent systems. Some of them (Gypsy, JADE, Ajanta, JATLite, etc.) were elaborated in the universities with the purpose of research of this technology. Other systems (such as ASDK, JAFMAS, etc.) exist on the library level, providing the programmer with only basic classes for implementing the main components: agents, platforms, interaction mechanisms and security. Independent systems are developed on the base of them, for examples MagNet and E-Commercia. Recently commercial systems appear, such as Gossip by Tryllian company, Bee-gent and Plangent by Toshiba company. Unfortunately, the technical documentation is not available for them. Some of the projects on mobile agent technology investigation (ARA by Kaiserslautern University, Mole by Stuttgart University, Odyssey by General Magic company, etc.) are closed nowadays. There are number of mobile agent systems for solving narrowly specialized tasks (for example, Voyager System by ObjectSpace company).

V. AGENT PLATFORM JAVA AGENT DEVELOPMENT FRAMEWORK

Java Agent DEvelopment Framework (JADE) is one of the most popular agent platform. JADE project is developing by Telecom Italia Lab from 2000 year. JADE agent platform is typical middleware – software with a set of tools for creation and management of the multi-agent systems. JADE includes dynamic environment where JADE agents can live, class library which programmers can use for their own agent creation, and a set of graphical tools allowing to manage started agent activity.

JADE platform is distributed and includes a set of containers. Container is a dynamic runtime environment for multi-agent applications in which agents exists. Each container can contain multiply agents. A set of active containers is called a platform. One of the containers is always the main one, all other containers are connected to it and registered at the time of start-up. Therefore, the first container at the launch of the platform should be the main one, and all other containers should be non-main containers and should know in advance how to find the main container on which they will be registered, that is, they should have data about the host and port. Another main container running anywhere in the network is another platform on which new regular containers can register. Fig. 1 illustrates this concept based on an example showing two JADE platforms consisting of three and one container, respectively.

JADE agents are defined with unique names. Provided that they know the names of other agents, they can communicate, regardless of their actual location: in a common container (agents A2 and A3), in different containers on the same platform (agents A1 and A2), or in general on different platforms (A4 and A5). The user

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

does not have to know how the dynamic JADE environment works, but he needs to run it before starting to execute his agents.

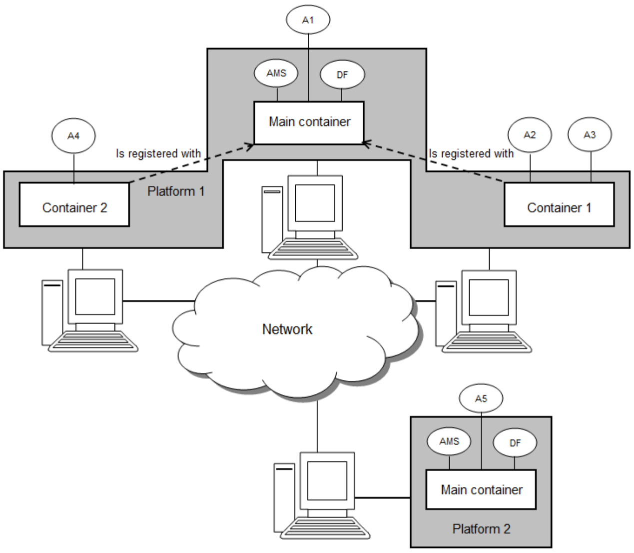


Fig. 3. JADE agent platform

In addition to the possibility of receiving registrations from other containers, the main container differs from the usual container in that it contains two special agents, which are launched automatically simultaneously with the container: AMS and DF.

JADE provides the programmer with following main facilities: (1) FIPA-compliant agent platform, including AMS and DF components which automatically activate when platform start; (2) distributed agent platform where agents are executed as Java-flows, live in containers, and only one Java-application and Java Virtual Machine can be run on each host; (3) multi-threaded execution environment with two-level scheduling; (4) object-oriented programming environment; (5) library of ready-to-use interaction protocols; (6) graphical user interface for management of several agents and containers from remote host; etc.

VI. PRACTICAL EXAMPLE OF MULTI-AGENT APPLICATION

Let us consider the example of trading multi-agent system which realized in JADE environment. The application include 4 agents:

- TraderAgent will search EmployerAgent with the best proposal in the yellow pages list and try to negotiate a deal with him;
- EmployerAgent tries to negotiate a deal with TraderAgents which interested in him and then conducts negotiations with MajorBuyerAgent, which give it the best proposal;
- MajorBuyerAgent firstly searches the best proposal from EmployerAgent, then searches interested BuyerAgent for the better deal with EmployerAgent, and at the end give it final proposal;
- BuyerAgent looks for the most profitable proposal among MajorBuyerAgents, then waits for negotiations between the MajorBuyerAgent to which it signed up and TraderAgent.

Fig. 4 illustrates an example of JADE agent platform running.

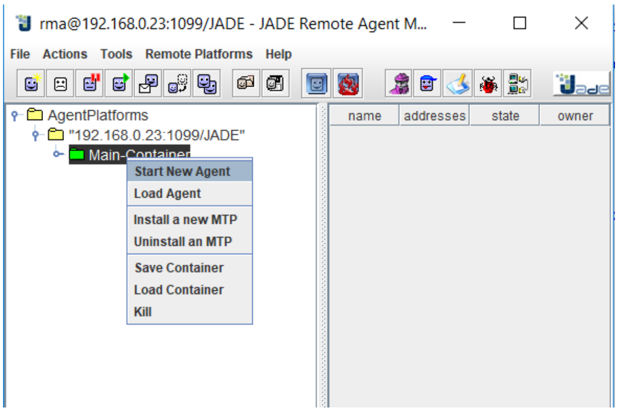
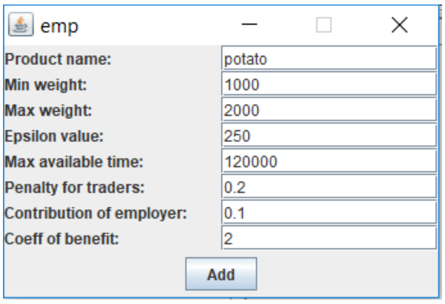


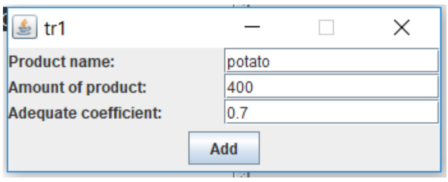
Fig. 4. JADE agent platform running

Fig. 5-9 illustrates the processes of creation of all actors of the trading platform.



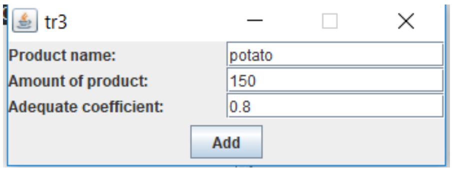
Product name: potato
Min weight: 1000
Max weight: 2000
Epsilon value: 250
Max available time: 120000
Penalty for traders: 0.2
Contribution of employer: 0.1
Coeff of benefit: 2

Fig. 5. Creation of EmployerAgent



Product name: potato
Amount of product: 400
Adequate coefficient: 0.7

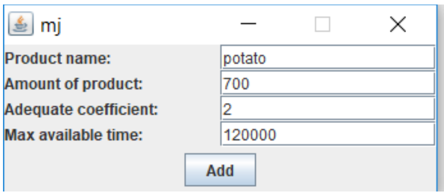
Fig. 6. Creation of TraderAgent1-2



Product name: potato
Amount of product: 150
Adequate coefficient: 0.8

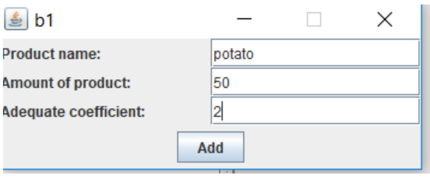
Fig. 7. Creation of TraderAgent3

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.



Product name: potato
Amount of product: 700
Adequate coefficient: 2
Max available time: 120000

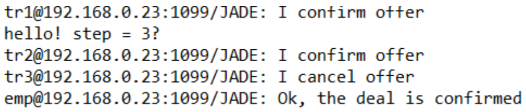
Fig. 8. Creation of MajorBuyerAgent



Product name: potato
Amount of product: 50
Adequate coefficient: 2

Fig. 9. Creation of BuyerAgent

Protocol of EmployerAgent and TraderAgent communication at the first stage of the deal is following (Fig. 10).

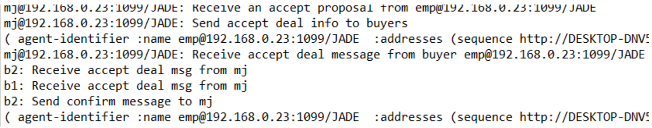


```
tr1@192.168.0.23:1099/JADE: I confirm offer
hello! step = 3?
tr2@192.168.0.23:1099/JADE: I confirm offer
tr3@192.168.0.23:1099/JADE: I cancel offer
emp@192.168.0.23:1099/JADE: Ok, the deal is confirmed
```

Fig. 10. Protocol of MAS working at the first stage of the deal

TraderAgents 1-3 have less product amount in sum than minimum established by EmployerAgent. In this case EmployerAgent raises additionally a penalty for a shortage from TraderAgents. As can be seen from the protocol, that satisfied two agents, but didn't satisfy third agent, since in this case third agent lost an unacceptable amount of profit for itself.

Fig. 11 contains listing of the MAS work protocol at the 2nd stage (communication between MajorBuyerAgent, BuyerAgent and EmployerAgent).



```
mj@192.168.0.23:1099/JADE: Receive an accept proposal from emp@192.168.0.23:1099/JADE
mj@192.168.0.23:1099/JADE: Send accept deal info to buyers
( agent-identifier :name emp@192.168.0.23:1099/JADE  :addresses (sequence http://DESKTOP-DNV
mj@192.168.0.23:1099/JADE: Receive accept deal message from buyer emp@192.168.0.23:1099/JADE
b2: Receive accept deal msg from mj
b1: Receive accept deal msg from mj
b2: Send confirm message to mj
( agent-identifier :name emp@192.168.0.23:1099/JADE  :addresses (sequence http://DESKTOP-DNV
```

Fig. 11. Protocol of MAS working at the second stage of the deal

MajorBuyerAgent and BuyerAgent1-2 were able to gain the product amount in the order exactly as much as EmployerAgent had. Therefore, the transaction was successfully completed.

VII. CONCLUSION

Analysis of the existing software and own experience of multi-agent system developing [7–9] shows, that using of intellectual agents allows simplifying such procedures as information search and processing in Internet, using of media-services, different business tasks of big data processing and other tasks related with information search, recognition and processing. The process of MAS design and developing is well standardized. The main standards in this area is MASIF and FIPA. The most popular agent platform for MAS developing and functioning is JADE. As a practical result of the research conducted the trading multi-agent system has been elaborated, taking into account the preferences of all actors in the process, flexibility, sustainability and consistency of their interactions.

REFERENCES

[1] P. Maes, "Agent that Reduce Work and Information Overload", In Communication of the ACM, vol. 37, #7, pp. 30-40, July 1994.

[2] T. Magedanz, K. Rothermel, and S. Krause, "Intelligent agents: Anemerging technology for next generation telecommunications". In IEEE Proceedings of the Fifteenth Annual Joint Conference of the IEEE Computer Societies on Networking the Next Generation (INFOCOM 96), 1996, vol. 2, pp.464–472.

[3] S. Franklin and A. Graesser, "Is It an agent, or just a program?: A taxonomy for autonomous agents", J. P. Müller, M. J. Wooldridge, N. R. Jennings (eds) Intelligent Agents III Agent Theories, Architectures, and Languages. ATAL 1996. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol 1193, Springer, Berlin, Heidelberg, 1997, pp. 21-35, https://doi.org/10.1007/BFb0013570.

[4] C. Thompson et al., "Agent Technology Glossary", Revision 0.2, Object Management Group Document, 1999, 19 p.

[5] "Agent Technology in OMA", Object Management Group Document, 1999, mode of access: http://www.objs.com/isig/agent-rfi-6.html, date of access: 7 August 2021.

[6] FIPA Abstract Architecture Specification. Foundation for Intelligent Physical Agents, 2002, mode of access: http://www.fipa.org/specs/fipa00001/XC00001J.html, date of access: 7 August 2021.

[7] A. N. Morozevich and B. A. Zhalezka, "Personal multi-agent decision support systems", In Proc. of the Sixth Intern. Conf. Pattern Recognition and Information Processing, Minsk-Szczecin, 2001, pp. 233–236.

[8] T. V. Solodukha and B. A. Zhelezko, "Developing a multi-agent system for e-commerce", In Knowledge Acquisition and Management, Research papers of Wroclaw University of Economics, Wroclaw, Publishing House of Wroclaw University of Economics, 2011, pp. 117–125.

[9] T. V. Solodukha, O. A. Sosnovskiy, B. A. Zhelezko, "Multi-Agent System for E-Commerce", In Pattern Recognition and Information Processing (PRIP'2009), Proceedings of the 10th International Conference, 19-21 May, 2009, Minsk, Belarus, Minsk, Publ. center of BSU, 2009,pp. 354-358.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Increasing the Reliability of Pattern Recognition by Analyzing the Distribution of Errors in Estimating the Measure of Proximity between Objects

Rahim Mammadov
Dept. of Instrumentation Engineering
Azerbaijan State Oil and Industry University
Baku, Azerbaijan
rahim1951@mail.ru
ORCID 0000-0003-4354-3622

Elena Rahimova
Dept. of Instrumentation Engineering
Azerbaijan State Oil and Industry University
Baku, Azerbaijan
elena1409_mk@mail.ru
ORCID 0000-0003-1921-4992

Gurban Mammadov
Azerbaijan State Scientific-Research Institute for Labor Protection and Occupational Safety
Baku, Azerbaijan
qurban_9492@mail.ru
ORCID 0000-0002-2874-6221

Abstract. **When recognizing similar or close objects in a report, the accuracy of the recognition is very low when the value of the measure of proximity between objects (MPBO) is close to the value of the error that occurs. modern algorithms are preferred instead of emprig dosturlar to improve accuracy in calculating the measure of proximity between objects. The algorithm proposed in previous research work is not effective, although it eliminates problems such as gross error, correlation coefficient, and the presence of a modular sign in formulas. Proposing a new methodology, range analysis was used instead of summarizing the results when calculating parameter values. The advantage of this system is distinguished by error reduction, more accurate recognition and efficiency. The given algorithm was modeled on a computer and the results were obtained. The processing of the results shows that, thanks to the proposed methodology, it is possible to significantly increase the accuracy of the calculation of the measure of the proximity between objects. At this time, it does not affect the running speed of the system.**

Keywords: **pattern recognition, measurement errors, interval analysis, correlation coefficient, MPBO, increase of accuracy, re-measurements**

I. INTRODUCTION

Technical vision systems are a field of artificial intelligence that teaches mobile robots to interpret and understand the visual world. Technical vision systems are used in space, aviation, surface ground, water surface and underwater mobile vehicles, which have the ability to analyze the external situation in real time. Technical vision systems have the ability to analyze what they see after accurately identifying and classifying objects using images, videos, and modeled digital images [1–5].

The reliability of the information received by the technical vision system has a significant impact on the formation of the object of recognition in terms of minimizing image recognition errors. The reliability of the information received by technical vision systems that ensure the quality of work of mobile robots are intelligent information-measuring systems is determined primarily by the quality characteristics of the sensors and the parameters of object recognition accuracy. Object recognition accuracy parameters are characterized by the accuracy of estimating the measure of proximity between objects (MPBO) determined by the calculation method. The errors allowed when measuring the values of the features of images, summing up according to the most complex law, create an error in assessing the measure of proximity between objects, which in the computer vision system is commensurate with the actual value of the distance between the features of objects. Therefore, these errors, which reduce the cost of image recognition reliability, seriously impede the use of intelligent information-measurement systems and technical vision systems for the widespread use of mobile robots in various fields.

II. PROBLEM STATEMENT

The accuracy of pattern recognition depends on the accuracy of the calculation of the measure of the proximity between objects. Manhattan, Euclid, Camberra, and many other formulas have been proposed in the field of pattern recognition. They are currently used to calculate the measure of the proximity between objects. However, due to the fact that each formula is overly integrated, there are a number of shortcomings, which maintain the relevance of the correct calculation of the error of

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

measurement of proximity between objects and, consequently, the high accuracy of image recognition.

Researches show that it is not possible to eliminate certain random and gross errors using existing empirical formulas. Even increasing the number of repeated measurements does not solve this problem. This is due to the use of an absolute sign in existing formulas (distance can never be negative). In the calculation of the measure of the proximity between objects, a gross error, statistical processing, the presence of a modular sign in the existing formulas and the correlation coefficient between the measurements can be a direct cause of incorrect results. In the proposed methodology, the recognition (input) and exemplary (reference) objects with the help of technical means are entered into the computer. The program method finds in the calculation of the numerical average, the standard deviation, the correlation coefficient and the final error of the measures of the proximity between objects values of the input and reference parameters [6–11].

Using the Manhattan formula, the compatibility of input and reference parameters is checked, so that it is simpler and more convenient than others [12–15]:

$$z = \sum_{i=1}^{n} |x_i - y_i|$$

Here is the result from the i - th re - measurement of input and benchmark parameters according to x_i and y_i. When measuring the input and reference parameters, the overage square errors σ_x and σ_y must obey the normal distribution law, so σ_z must obey the normal distribution law. However, the calculation of the absolute price violates the distribution of the final error, and as a result, the average price shifts in a positive direction. Thus, the final result is incorrect. Therefore, in the proposed algorithm, it is necessary to find out what part of this distribution the difference between the input and reference parameters falls on. The difference between the input and reference parameters is $a = x_i - y_j$. x and y are measured in n times, it is usually checked with each value of y for each value of x. That is,

$$\begin{array}{cccc} x_1 - y_1, & x_1 - y_2, & \dots & x_1 - y_n \\ x_2 - y_1, & x_2 - y_2, & \dots & x_2 - y_n \\ \dots & \dots & \dots & \dots \\ x_n - y_1, & x_n - y_2, & \dots & x_n - y_n \end{array}$$

The difference a is checked in the range -3*σ_z - +3*σ_z by every $0.5 \cdot \sigma_z$ steps ([-3*σ_z, -2,5*σ_z], [-2,5*σ_z, -2*σ_z], [-2*σ_z, -1,5*σ_z], [-1,5*σ_z, -σ_z], [-σ_z, -0,5*σ_z], [-0,5*σ_z, 0], [0, 0,5*σ_z], [0,5*σ_z, σ_z], [σ_z,1,5*σ_z], [1,5*σ_z,2*σ_z], [2*σ_z,2,5*σ_z] and [2,5*σ_z,3*σ_z]).

If a does not fall in the interval, the program checks whether it falls in other intervals. In the case of an interval falls, then as the price of **a**, the smallest price of the interval is accepted and sent to the total input and the possible deviations are minimized. In the measurement technique, errors are accepted up to ± 3*σ. Greater than it, is thrown like a gross error. Therefore, a's greater than $-3 \cdot \sigma_z$ and $+3 \cdot \sigma_z$ are not taken into account. Then the a's in the interval $[-3 \cdot \sigma_z, 0]$ and $[0, +3 \cdot \sigma_z]$ are collected and the average value is found by dividing by the number of measurements. The final values are found in the general order by the Manhattan formula and found by the operation of our algorithm. The number of measurements varies from 1 to n. The greater the number of repeated measurements of the input parameters, the greater the accuracy. However, in this case, the speed of the recognition system decreases. Therefore, the number of repeated measurements of the input and reference parameters in the proposed algorithm is taken differently. Since the repeated measurements of the reference parameters are in training mode, their number should be taken as much as possible. Because in this case, the accuracy is high and the speed of the system does not change. Since the identification and reference objects are taken from same in advance, in fact, the result must be "0".

Therefore, the use of interval analysis, taking a relatively small number of repeated measurements, both slows down the operation of the recognition system, and has the appropriate accuracy, which is reflected in this program. It is better to say that the values of *zm* (manhattan) and *zk* (proposed) in the algorithm are closer to the corresponding values obtained at the maximum value of n. Mathematical modeling of the proposed algorithm for calculating the measure of the proximity between objects and the results are given in previous scientific papers. Calculates the current a at the maximum, average, and minimum values of a given range of parameters.

The results are very good when calculating the average value of the current parameter a in a given interval, and the accuracy increases significantly when NK takes repeated measurements. . It also does not affect the processing speed. As can be seen from the given tables and diagrams, the result of the proposed algorithm is much higher than the results obtained by the classical method, and when using the proposed algorithm, the accuracy of the technical vision system increases significantly and the operating speed remains at the required level. Then the measurement errors of the Manhattan and the proposed algorithm are calculated by repeating NK times.

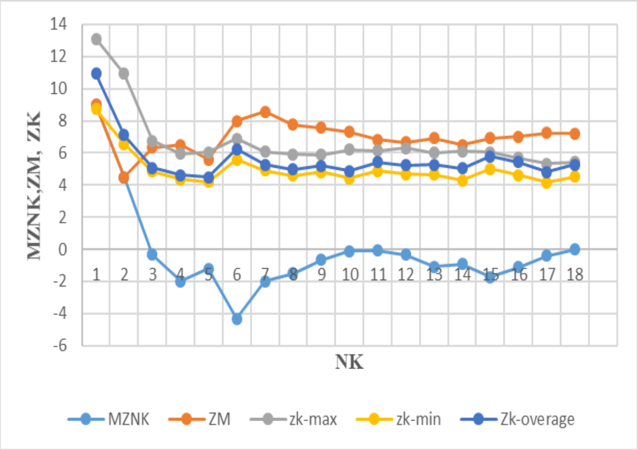


Fig. 1. Comparison of the classical method with the proposed algorithm for calculating the size of the proximity between objects

Fig. 1 is displayed non-modular calculations in blue color, calculations based on Manhattan's formula in red color, take the maximum value of the range instead of the parameters that fall into the range in gray color, take the minimum value of the range instead of the parameters that fall into the range in yellow color and take the average value of the range instead of the parameters that fall into the range in blue-purple color. The graphs show that the proposed algorithm is more accurate than the classical methods.

TABLE I. COMPARISON OF THE CLASSICAL METHOD WITH THE PROPOSED ALGORITHM FOR CALCULATING THE SIZE OF THE PROXIMITY BETWEEN OBJECTS

NK	MZNK	ZM	ZK_{min}	$ZK_{overage}$	ZK_{max}
1	9	9	8,74	10,92	13,11
2	4,5	4,5	6,55	7,1	10,92
3	-0,33	6,33	4,85	5,07	6,74
4	-2	6,5	4,37	4,61	5,94
5	-1,2	5,6	4,19	4,5	6,04
6	-4,33	8	5,58	6,23	6,87
7	-2	8,57	4,91	5,26	6,06
8	-1,5	7,75	4,58	4,96	5,9
9	-0,66	7,55	4,8	5,2	5,88
10	-0,1	7,3	4,42	4,87	6,21
11	-0,09	6,81	4,88	5,42	6,14
12	-0,33	6,67	4,67	5,23	6,29
13	-1,07	6,92	4,66	5,25	6,02
14	-0,93	6,5	4,3	5,05	6,11
15	-1,73	6,93	4,99	5,77	6,03
16	-1,125	7	4,63	5,42	5,68
17	-0,411	7,23	4,17	4,81	5,34
18	0	7,22	4,52	5,27	5,43

In this Table, the number of repeated measurements of the input parameter NK (1-18); MZNK – the manifestation of the Manhattan formula without modular sign; ZM – the measure of the proximity between objects calculated on the basis of Manhattan formula; Zk_{min}, $Zk_{average}$ and ZK_{max} according to which of the measured price falls into any of the ranges, its price is taken as the lower, middle and upper values of the range instead.

As can be seen from the table and figure, the algorithm proposed in all three options has a great advantage. But as it seems, the results are different from each other, and coming to a common opinion creates certain assumptions for the result.

Therefore, another algorithm has been proposed. According to this algorithm, the number of repeated measurements per any range is calculated and the maximum number of repeated measurements per range is decided.

III. PROBLEM SOLVING

Formulas are used in existing systems of recognition of images. However, the error is large because the formulas are too integrated. A new algorithm has been proposed using range analysis to minimize errors.

TABLE II. RESULTS OBTAINED FROM RANGE ANALYSIS WHEN CALCULATING THE MEASURE OF THE PROXIMITY BETWEEN OBJECTS

NK	MK	ZM	OAYÖ	K_max
2	18	4,5	1,9	2
3	18	6,33	1,9	2
4	18	6,5	1,9	2
5	18	5,6	1,9	2
6	18	8	0	1
7	18	8,57	0	1
8	18	7,75	0	1
9	18	7,55	0	1
10	18	7,3	0	1
11	18	6,81	0	1
12	18	6,66	0	1
13	18	6,92	0	1
14	18	6,5	0	1
15	18	6,93	0	1
16	18	7	0	1
17	18	7,23	0	1
18	18	7,22	0	1

As can be seen from Table II, NK and MK performed repeated measurements to calculate the size of the proximity between the objects and how many times each range fell. K_max indicates the range in which the values of repeated measurements fall more.

Table II shows the number of repeated measurements in each range as a result of interval analysis in the range f (1)-f (11).

The green part indicates the area where the repeated measurements fell the most. As can be seen, the recognition of the input quantity in the values of repeated measurements 2-5 falls into the 2nd interval, in the subsequent values of the number of repeated measurements the recognition falls into the 1st interval.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

TABLE III. THE NUMBER OF MEASUREMENTS PER RANGE IN THE INTERVAL ANALYSIS USED

n	f (1)	f (2)	f (3)	f (4)	f (5)	f (6)
2	3	10	4	1	7	6
3	4	12	8	2	8	9
4	5	17	8	10	12	9
5	6	22	8	17	16	9
6	20	13	13	18	12	9
7	25	16	13	20	16	9
8	26	23	18	22	18	9
9	31	26	18	24	22	9
10	36	27	21	28	26	10
11	40	31	21	29	30	11
12	45	32	25	35	31	11
13	48	33	29	37	31	15
14	53	34	33	43	32	15
15	56	35	37	45	32	19
16	61	38	37	47	36	19
17	65	42	39	51	36	22
18	72	43	40	55	37	23

Continuation of Table 3

n	f (7)	f (8)	f (9)	f (10)	f (11)
2	2	2	0	0	1
3	6	3	0	0	1
4	7	3	0	0	1
5	8	3	0	0	1
6	13	1	0	5	4
7	16	1	0	6	4
8	16	2	0	6	4
9	19	2	0	7	4
10	19	2	0	7	4
11	20	4	0	7	5
12	21	4	0	7	5
13	25	4	0	7	5
14	26	4	0	7	5
15	30	4	0	7	5
16	33	4	0	8	5
17	33	4	1	8	5
18	35	4	1	9	5

IV. CONCLUSİON

It should be noted that it is more expedient to use this method instead of the existing formulas in automatic recognition and control systems, as the intended shortcomings are eliminated. Thus, along with the elimination of shortcomings, there are a number of advantages. In this case, the errors caused by statistical processing, correlation coefficient, application of the modulus sign in the formulas and gross error are eliminated, and as a result, accuracy increases. Also, despite the increase in the number of repeated measurements, the speed of the recognition system is not affected. The classical method and the proposed algorithms were modeled on a computer and the results were obtained. As can be seen from the tables and graphs, higher results can be obtained by eliminating the uncertainties in recognition by dividing the range of distribution of measurement errors of the measure of proximity between objects into intervals and analyzing those intervals. Since these algorithms are solved on a computer, it is not more difficult to find the optimal values of the intervals. The processing of the results showed that the proposed algorithm can significantly increase the accuracy of estimating the measure of proximity between objects.

REFERENCES

[1] Bastian Hartmann, Christoph Schauer, Norbert Link, “Worker Behavior Interpretation for Flexible Production,” Engineering and Technology Intern. J. of Industrial And Manufacturing Engineering, 2009, pp. 1224–1232.

[2] Tushar Jain and Meenu, “Automation and Integration of Industries through Computer Vision Systems,” Intern. J. of Information and Computation Technology, 2013, pp. 963–970.

[3] Keith Jacksona, Konstantinos Efthymioua, John Borton, “Digital manufacturing and flexible assembly technologies for reconfigurable,” Changeable, Agile, Reconfigurable & Virtual Production Conference, 2016, pp. 274–279.

[4] F. Leighton, R. Osorio, G. Lefranc, “Modelling, Implementation and Application of a Flexible Manufacturing Cell,” Intern. J. of Computers, Communications&Control, 2011, pp. 278–285.

[5] Phansak Nerakae, Pichitra Uangpairoj, Kontorn Chamniprasart, “Using machine vision for flexible automatic assembly system,” Intern. Conf. on Knowledge Based and Intelligent Information and Engineering Systems, 2016, pp. 428–435.

[6] Petar Marić, “Computer Vision Systems For The Enhancement of Industrial Robots Flexibility,” Facta Universitatis, Ser. Mechanics, Automatic Control and Robotics, 2011, pp. 1–18.

[7] N. Herakovic, “Robot Vision in Industrial Assembly and Quality Control Processes,” Robot Vision, Edited by Ales Ude, 2010, pp. 501–534.

[8] D. G. Caldwell, “Robotics and automation in the food industry. Current and future technologies,” Woodhead Publishing Limited, 2013, pp. 523.

[9] B. Siciliano, O. Khatib., “Springer Handbook of Robotics,” Springer-Verlag Berlin Heidelberg, 2008, 1628 p.

[10] Vimal Sudhakar Bodke, Omkar S. Vaidya, “Object Recognition in a Cluttered Scene using Point Feature Matching,” International Journal for Research in Applied Science & Engineering Technology, 2017, pp. 286-290.

[11] Toshiaki Ejima, Shuichi Enokida, Toshiyuki Kouno, “3D Object Recognition based on the Reference Point Ensemble,” International Conference on Computer Vision Theory and Applications, 2014, pp. 261–269.

[12] R. K. Mammadov, T. Ch. Aliyev, “Kontrol' polozheniya 3D-ob"yektov v gibkikh avtomatizirovannykh sistemakh. Povysheniye dostovernosti raspoznavaniya”. LAP Lambert academic publishing, 2014, 90 p.

[13] Р. К. Мамедов, У. Г. Иманова, Повышение достоверности принятия решений при распознавании образов, Электрон. моделирование, Киев, 2014, т. 36, № 5, с. 115–121.

[14] R. K. Mammadov, A. S. Mutallimova, T. Ch. Aliyev, “Ispol'zovaniye momentov inertsii izobrazheniya dlya invariantnogo k affinnym preobrazovaniyam raspoznavaniya,” Vostochno-Yevropeyskiy zhurnal peredovykh tekhnologiy, 2012, pp. 4–7.

[15] В. В. Скачков и др. Минимизация доминирующей погрешности в задачах измерения информационных параметров «зашумленной» выборки сигнала, Information and Telecommunication Sciences, 2016, vol. 7, no. 2, Одесса, Украина, с. 62–69.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Using GRID for Centralized Synthesis of FPGA-based Information Security Systems

Viktor Evdokimov
Dept. of Mathematical and
Computer Modelling
Pukhov Institute for Modelling
in Energy Engineering
of NAS of Ukraine
Kyiv, Ukraine
evdokimov@ipme.kiev.ua

Anatoly Davydenko
Dept. of Mathematical and
Econometric Modelling
Pukhov Institute for Modelling
in Energy Engineering
of NAS of Ukraine
Kyiv, Ukraine
davydenko@ipme.kiev.ua

Serhii Hilgurt
Dept. of Mathematical and
Econometric Modelling
Pukhov Institute for Modelling
in Energy Engineering
of NAS of Ukraine
Kyiv, Ukraine
hilgurt@ipme.kiev.ua

***Abstract.* The functioning of such signature-based information security tools as network intrusion detection system, antivirus, anti-worms and others are based on solving in real time resource-intensive task of multi-pattern string matching. Due to rising traffic rates, increasing number and sophistication of attacks and the collapse of Moore's law, traditional software solutions can no longer keep up. Therefore, hardware approaches are frequently being used by developers to accelerate pattern matching. Reconfigurable FPGA-based devices, providing the flexibility of software and the near-ASIC performance, have become increasingly popular for this purpose. Signature databases of the current information security systems contain hundreds of thousands and even millions of rules. Every signature database update or change in network parameters forces the digital circuit of such system to be resynthesized, and the FPGA – to be reconfigured. To facilitate such reconfigurations we propose a centralized service for information security reconfigurable tools synthesis, which uses free high-performance resources of GRID infrastructure.**

***Keywords:* security, signature, FPGA, GRID, centralization**

I. INTRODUCTION

The propagation of Internet and network technologies in both industrial and civil enterprises together with the widespread availability of system hacks and viruses have made the importance of network security more significant. The increase in number and sophistication of attacks against the network infrastructure and computer systems requires more robust security solutions. Unfortunately, despite the great progress made in deep neural networking (DNN), security tools based on such methods still suffer from nonzero recognition error probability. Even as low false positive rate as 0.01% is able to disrupt the correct operation of the information system [1]. Using such systems in critical infrastructure can have disastrous consequences. Therefore, signature-based recognition methods with their theoretically exact match are still relevant when creating information security systems.

The main disadvantage of the signature-based principle is its computational complexity. Checking every byte of every packet to see if it matches one of set of hundred thousand strings becomes a major performance bottleneck in traditional software solutions which have to scan the incoming data in real time [2]. To keep up with these speeds a specialized device is required.

Different types of available hardware solutions are result in higher efficiency. Among them the Field Programmable Gate Array (FPGA) devices have commonly been proposed because they feature both the flexibility of software and the high performance of specialized hardware at a reasonable cost [3]. The reconfigurable accelerators based on FPGAs became a suitable and popular hardware platform for many security applications, including network intrusion detection/prevention systems (NIDS/NIPS) [4], antivirus, anti-worms and other signature-based information security tools. The main difficulty when using programmable logic is to synthesize the digital circuit and generate the appropriate configuration file (bitstream) for the FPGA device to give it required functionality. This process is quite complicated and requires the efforts of a highly qualified developer. Not every firm or enterprise can afford such an asset.

As a solution we propose a centralized service for security tasks hardware accelerators synthesis. Such service can use computation power of GRID or other high performance equipment to process requests

---

This work was supported in part by the Informatization program of the NAS of Ukraine in 2020-2024.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

from users to obtain desired FPGA configuration files.

II. FPGA-BASED NETWORK INTRUSION DETECTION/PREVENTION SYSTEMS

Historically the first and therefore the most studied FPGA-based tools of information security were network intrusion detection systems [5]. Without losing the generality of reasoning, consider the typical functions of reconfigurable security systems on the example of NIDS/NIPS.

The generalized structure and the content of a NIDS/NIPS based on FPGA can be compiled on the basis of a number of well-known works [5, 6] (Fig. 1).

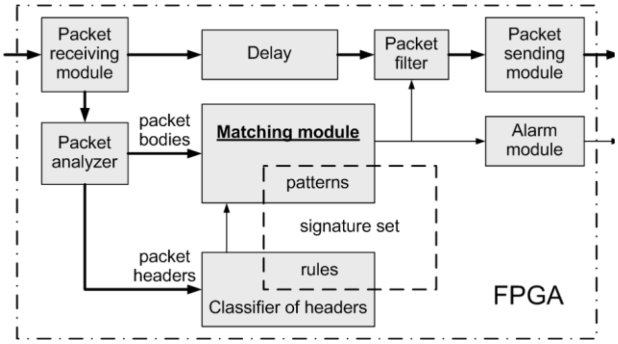


Fig. 1. The generalized structure of the FPGA-based network intrusion detection/prevention system

The Matching module is the most important component of NIDS/NIPS. It solves a computationally complex task of multi-pattern string matching, i.e. checks the content of network packets against certain sequences of symbols, so-called *patterns*, which are parts of the signatures – the descriptions of the known attacks.

Components Delay, Packet filter and Packet sending module are present only in NIPS option of the system.

III. UPDATING THE STRUCTURE OF RECONFIGURABLE SECURITY SYSTEM

As we can see when analyzing the structure shown above, the signature set is “wired” into the circuitry of the device at hardware level. This feature ensures the highest performance rate due to the maximally possible parallelism has been reached. On the other hand, any change in the circumstances of NIDS/NIPS functioning leads to the need to update the digital circuit. There are two possible reasons of such update. Firstly, new unknown attacks can emerge; consequently, new signatures have to be added into the signature set. Secondly, the operating conditions of the protected system can be changed (modification of the local network, change of its content or structure, modification of the software, etc.), resulting in necessity to add or remove whole classes of signatures, so the content of signature set is also to be changed. That is the digital circuit of the security tool needs to be re-synthesized, new bitstream to be obtained and loaded into the FPGA.

Analyzing of this structure shows that some components are independent from signature set, let's call them *constant modules*, whereas another parts have to be reconstructed obligatorily, we name them *variable modules*. The Packet receiving module, the Packet analyzer and the Alarm module belong to the first group. The Matching module and the Classifier of headers constitute the second one.

Thus, the key feature of reconfigurable information security system is the necessity occasionally to perform a so called *operational update procedure*, i.e. – to re-synthesize the variable modules and reconfigure the FPGA chip.

IV. ORGANIZING THE SYNTHESIS OF RECONFIGURABLE SECURITY TOOLS

Constructing reconfigurable devices is a complex and resource-intensive task. This task, firstly, requires high performance computer hardware. Generation of a bitstream, to be loaded into the FPGA chip, is usually fulfilled by a proprietary CAD software tool from the FPGA manufacturer and includes a number of computationally-intensive procedures such as Synthesize, Translation, Map, Place & Route and the Bitstream Generating. Depending on the complexity of the circuit and the type of FPGA, this process can take from tens of minutes to several hours. Secondly, the developers must have a high qualification and knowledge of digital device design specifics.

There are many approaches to build an FPGA-based scheme able to fulfill multi-pattern string matching [7]. Each of them has its own strengths and weaknesses. None of them demonstrates key advantages over others. The lack of the best solution, which exceeds the competitive ones by all technical parameters, makes the developers modify and elaborate the known approaches thoroughly.

That is why owners of information security systems do not have possibilities to solve this task on their own.

In this work, we propose to organize synthesizing reconfigurable tools so that complex and resource-intensive procedures are performed not locally on each individual system, but centrally, using free high-performance computer resources of GRID. Fig. 2 shows the principle of operation of a service that implements this approach.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

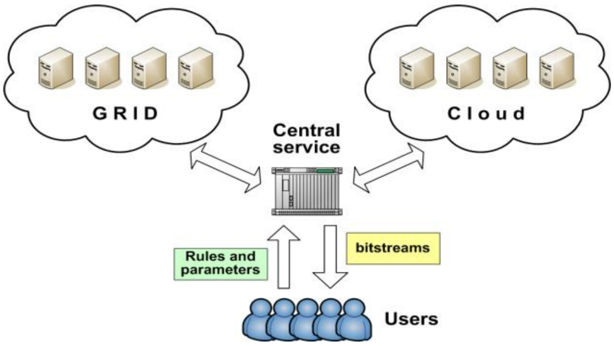


Fig. 2. Service for centralized synthesis of FPGA-based information security systems

Users send requests containing signature set (NIDS rules) and FPGA parameters to the central service. Central service using computation power of GRID infrastructure processes these requests, synthesizes digital circuits in automatic mode, compiles projects, generates bitstreams and returns them to users. To avoid a hypothetical situation when there are no free GRID resources, the service is able to use the cloud infrastructure as well.

V. GRID-SERVICE

To verify the concept provided above the service was realized as a model sample on the base of computing cluster of Pukhov Institute for Modelling in Energy Engineering of NAS of Ukraine using the resources of Ukrainian National GRID (UNG). This project was named as Security Tasks Reconfigurable Accelerators GRID-Service (STRAGS).

In order to facilitate the solution of the problem the Rainbow framework, recently developed by Ukrainian scientists, was chosen as the functional base of the service [8]. This technology, in fact, implements a cloud service at the PaaS level over the GRID infrastructure by running virtual machines (VMs) equipped with all necessary software as GRID-jobs. To provide real-time interaction between the user and VM that run at remote GRID-node, special techniques was invented and applied by the developers of Rainbow. Initially, this technology was created to run specialized software "Moldyngrid" for the virtual organization "Medgrid" in the UNG environment. However, the work turned up successful and soon found wider use.

When functioning, the STRAGS service initiates the work of several agents on remote GRID-sites, which support Rainbow platform (Fig. 3). Wherein, every agent is a VM with all the necessary instrumental software preinstalled and configured for the synthesis of reconfigurable devices and generation of the bitstreams.

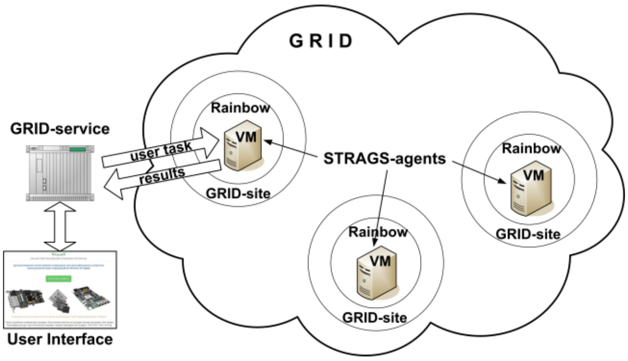


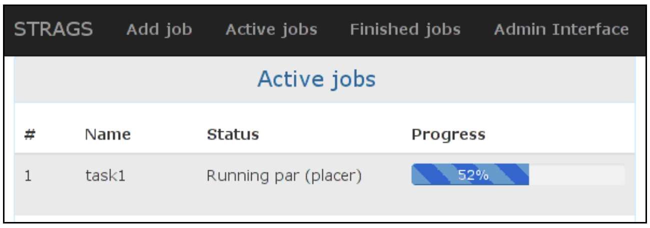
Fig. 3. STRAGS service functioning

As requests from clients are received, the service distributes tasks among active agents, maintaining their number sufficient to ensure required availability. Having received the task as a GRID-job, the agent starts the processes of automatic synthesis of the required digital circuit and synthesis of the corresponding configuration for the FPGA, and then returns the results of the work to the service.

Comparing to the traditional GRID-jobs batch submission, running an agent as a VM has several benefits:

- VMs already have FPGA synthesis software installed that simplifies the operational update procedure;
- pulling VMs from GRID-site allows to minimize latencies in synthesis task lifecycle;
- Rainbow framework interactive access feature allows extending service with debug capabilities and provides direct access to the platform with pre-installed synthesis software eliminating the need of local software copy;
- users have the ability to track the synthesis progress with a high granularity.

Fig. 4 presents one of STRAGS service user interface windows, which depicts a current security tool synthesis process progress (the substep "Placer" of the step "Place & Root" is fulfilling).



Active jobs

#	Name	Status	Progress
1	task1	Running par (placer)	52%

Fig. 4. User's job "task1" synthesis progress in the "Active jobs" window of the STRAGS service

The proposed architecture of GRID-service addresses issues of user's authorization, agent's

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

authorization and availability of a sufficient number of agents. It also implements centralized updating of VMs to ensure their relevance when distribution across the GRID infrastructure.

VI. EXPERIMENTAL RESULTS

Verifying the service on a lot of test tasks did not allow us to obtain numerical estimates. Instead, significant quality benefits were identified:

1. Increasing of system performance in general due to the use of GRID and cloud which quickly synthesize the parallel matching circuits for modern FPGAs.
2. Improving the technical characteristics of local information security systems through the division of labor. Centralization allows the use of highly qualified specialists, whose work results are used on each of the local systems.
3. Lower total cost of ownership due to reducing staffing requirements for local systems.
4. Reducing overall computational costs by grouping similar requests.

As a minor disadvantage, the complexity of the whole system and some decrease in the speed of reconfiguring hardware components can be noted, which however is fully compensated by the listed advantages.

VII. CONCLUSION

In this work, the principle of centralized creation of FPGA-based information security systems is proposed. A GRID-service that implements this principle has confirmed the viability of the idea of transferring resource-intensive procedures from a local FPGA-based security system to a remote HPC environment.

It is worth noting that such a centralized service is applicable to the synthesis of any FPGA-based digital schemes, regardless of which recognition method it uses. It can be successfully used to build systems based on the mentioned above DNN approaches, when they become robust enough. In recent years, adversarial attacks detecting have become topical issue. Hardware solutions, including FPGA-based ones, are already used when corresponding systems constructing [9, 10].

REFERENCES

[1] G. E. Dahl, J. W. Stokes, L. Deng, and D. Yu "Large-scale malware classification using random projections and neural networks," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vancouver, Canada, May 26-31 2013, New York: IEEE.

[2] B. Smyth, Computing Patterns in Strings. Essex: Pearson Addison Wesley, 2003.

[3] V. Paxson et al., "Rethinking hardware support for network analysis and intrusion prevention," presented at the USENIX First Workshop on Hot Topics in Security (HotSec), Vancouver, July 31, 2006.

[4] С. В. Казмірчук, А. О. Корченко, і Т. І. Паращук, "Аналіз систем виявлення вторгнень," Захист інформації, Т. 20, № 4, С. 259–276, 2018, doi: 10.18372/2410-7840.20.13425.

[5] T. Katashita, Y. Yamaguchi, A. Maeda, and K. Toda, "FPGA-based intrusion detection system for 10 Gigabit Ethernet," IEICE Transactions on Information and Systems, Article vol. E90D, no. 12, pp. 1923–1931, Dec 2007, doi: 10.1093/ietisy/e90-d.12.1923.

[6] Ю. М. Коростиль и С. Я. Гильгурт, "Принципы построения сетевых систем обнаружения вторжений на базе ПЛИС," Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України, № 57, С. 87-94, 2010.

[7] W. Jiang, Y. H. E. Yang, and V. K. Prasanna, "Scalable multi-pipeline architecture for high performance multi-pattern string matching," in 24th IEEE International Parallel and Distributed Processing Symposium, IPDPS 2010, Atlanta, GA, 2010, doi: 10.1109/IPDPS.2010.5470374.

[8] А. А. Сальников, В. В. Вишневский, и А. Ф. Борецкий, "«Платформа как сервис» в грид для интерактивного анализа медицинских данных," Математичні машини і системи, № 1, С. 53-64, 2015.

[9] C. Song et al., "MAT: A Multi-strength Adversarial Training Method to Mitigate Adversarial Attacks," 2018 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), Proceedings Paper, pp. 476–481, 2018, doi: 10.1109/isvlsi.2018.00092.

[10] M. Capra, B. Bussolino, A. Marchisio, G. Masera, M. Martina, and M. Shafique, "Hardware and Software Optimizations for Accelerating Deep Neural Networks: Survey of Current Trends, Challenges, and the Road Ahead," IEEE Access, Article vol. 8, pp. 225134–225180, 2020, doi: 10.1109/access.2020.3039858.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Augmentation Tools for Object Detection in Satellite Images by Using U-Net Neural Network

Ales Zhuk
Belarusian State University
Minsk, Belarus
ales.zhook@gmail.com

Sergey Ablameyko
Belarusian State University
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
ablameyko@bsu.by

***Abstract.* The paper proposes an approach to detect discrete objects on images, namely buildings using the U-NET neural network. The main idea of our approach is to use additional augmentations during model learning. The experiments carried out have shown good results.**

***Keywords:* object detection, satellite images, neural network, U-Net, data set, data augmentations**

I. INTRODUCTION

Detecting and highlighting buildings on satellite images is an important task for various applications: building maps of the area, developing city infrastructure, searching for illegally built objects. Although the manual selection of buildings on satellite images is quite accurate, with a lot of images and the need for constant monitoring, manually processing them will take a lot of time and resources. Therefore, algorithms for automatic segmentation of satellite images are being developed. The task of automatic building detection can be complicated by bad weather conditions, the variety of shapes and colors of the found structures.

In recent years, neural networks have been used for image segmentation and processing. Classical neural network for object segmentation - U-NET. It was first used in 2015 for the segmentation of medical images [1]. The training set contained 30 images with 512x512 resolution. Dataset was expanded with additional transformations (rotations 90 degrees). The segmentation results surpassed other known methods and demonstrated the effectiveness of using U-NET on small image arrays.

Article [2] is devoted to the segmentation of satellite images. The task was to select 10 classes of objects (buildings / lakes / rivers / roads / etc.) on the images. The article describes the approach that was taken in the image segmentation competition on the Kaggle platform and helped the team to take third place. The idea was to use a modified U-NET network, and properties of some image channels (the images were 16-channel). So, water and vegetation could be detected without prior training, only by extracting information from image pixels. Due to the small number of images in the training set, data augmentation (rotations and flips) was applied. It should be noted that some images in the original training set are quite similar. For example, almost all buildings have blue roofs, making it easier to learn the network.

Typically, the U-NET is trained from scratch on some sort of initialized weights. The paper [3] demonstrates the possibility of using a pretrained network. And how U-NET can be improved using pretrained encoder. The neural network U-NET is described, in which VGG11, trained on the ImageNet weights [4], with a replaced fully connected layer was used as a contracting path (encoder). As a result of the work, the conclusions were next: the pretrained models converge faster to their limiting value, and that the recognition result of such a model is better in comparison to the non-pretrained network. Since this work was aimed at showing the benefits of using pre-trained networks, rather than getting the best result, there is still room for improvement. For example, using more complex networks as an encoder, such as VGG16, ResNet, etc.

We propose to introduce additional augmentations, such as adding noise, changing the brightness and contrast of the image, transforming perspective. And we show that this allows us to improve the segmentation result.

II. TRAINING SET

To solve the problem of segmentation of buildings, the set described in [5] was used. The images cover several settlements. The training set (similarly for the test set) contains 180 color three-channel images of 5000x5000 pixels with a spatial resolution of 0.3 meters. An example of an image and its mask is shown in Fig. 1. A mask is a binary image, where, depending on the pixel value (1 or 0), we determine whether this pixel belongs to the building or not.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

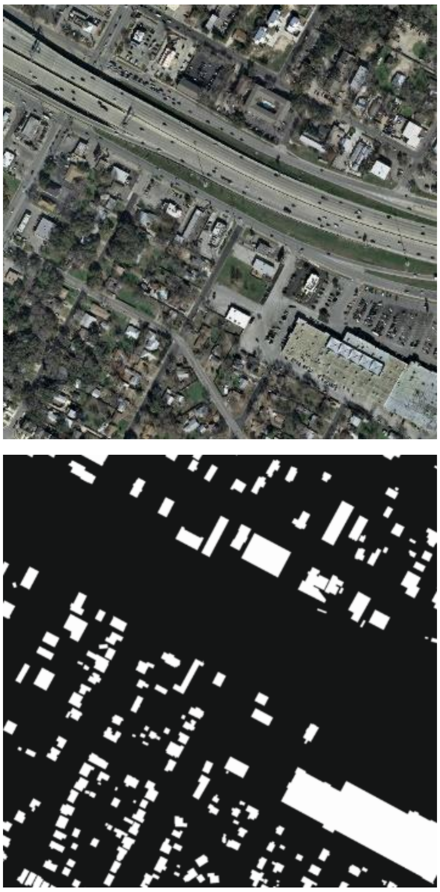


Fig. 1. Picture from training set

It should be noted that the images in the training set and the test set contain images of different cities. Training a neural network on images of some cities, and testing on others, allows you to understand how the algorithm adapts to other data of a similar nature.

III. NEURAL NETWORK ARCHITECTURE

In this work was used an architecture like the network described in [3]. U-NET consists of two parts: contracting and expanding. The contracting part is a convolutional network (convolution, activation, pooling), where the number of feature maps also increases on each layer. The expanding path is the opposite of the contracting path, where the pooling layer is replaced with an up-sampling layer, in which the image resolution is increased. U-NET also combines the features of the contracting path and with the expanding paths. The output of the U-NET network is a mask, where each pixel of the image is associated with the probability of its belonging to a particular class of objects. In our case, the probability that this pixel is a building. In our case, the contracting part was replaced by ResNet [6], pre-trained on the ImageNet weights.

IV. NEURAL NETWORK LEARNING

Network learning parameters:

1. The original set (180 images) was divided into two: training (150) and validation (30). At each iteration, for each image of the training set, a 768x768 segment is randomly cut out, all such segments are grouped into batches and transmitted to the network input. The batch size was chosen 8 (the maximum possible with this image size and the provided graphics card).
2. Focal loss [8] was used as a cost function. If y_{ij} is a true value that determines the class of a particular pixel, $\overline{y_{ij}}$ is the probability of a pixel belonging to a class with label 1 obtained by the model. Let:
$$\mathrm{p}_t = \begin{cases} \overline{y_{ij}} & , y_{ij} = 1, \\ 1 - \overline{y_{ij}} & , y_{ij} = 0. \end{cases} \quad (1)$$
Then the cost function can be written as:
$$C = -\frac{1}{n*m} \sum_{i=1}^{n} \sum_{j=1}^{m} -\alpha * (1 - \mathrm{p}_t)^{\gamma} * \ln \mathrm{p}_t \ , \quad (2)$$
where α, γ are parameters, which in our case are equal to 0.25 and 2, respectively. Focal loss can be characterized as a weighted cross-entropy function. Adding $\alpha * (1 - \mathrm{p}_t)^{\gamma}$ to the cost function reduces the value of the function for well-classified objects and, consequently, improves the learning result for negative cases. The use of this function is necessary in the case of an unbalanced dataset when one of the classes prevails or concedes to the others. In our case, buildings occupy 0.15 of the images in the training set. And the use of focal loss avoids preprocessing associated with building a balanced training set.
3. At the stage of network learning image augmentations (described in the next section) were additionally performed.
4. We used the Adam optimization algorithm [9] with a learning factor of 0.0001. (During the training of the model, the learning factor was decreased several times. The criterion for the decrease is that the validation metrics stop improving or changing.)

V. AUGMENTATIONS

Standard data augmentations used in satellite imagery segmentation tasks are rotations by angles divisible by 90 degrees and image flips. In this work, we propose and apply additional image transformations. The final set of augmentation is listed below:

- Rotate at a random angle multiple of 90 degrees,

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

- Vertical flip,
- Horizontal flip,
- Adding Gauss noise,
- Change hsv. (Random change in the hue, saturation, and brightness of the color of an image within a certain range),
- Change the brightness and contrast of the image (Random change in the brightness and contrast of the image in a certain range),
- Transformation of the perspective of the image.

At each training iteration, for a particular image, the probability that some augmentation would be applied was 0.25.

To test and implement augmentations, we used the library described in [10].

VI. EXPERIMENTAL RESULTS

In our work, we took advantage of the cloud computing capabilities provided by Google Colaboratory [12]. We were allocated a Tesla P100 graphics card with 16 GB of memory. This allowed us to use a more complex architecture of the neural network, increase the batch and size of images at the input of the neural network, carry out experiments faster and do more iterations during training. The result was assessed using next metrics: accuracy (3) and Jaccard coefficient (4)

$$A = \frac{1}{n*m}\sum_{i,j}^{n,m}\begin{cases}1, \text{где}\ y_{ij} = \bar{y}_{ij}, \\ 0\end{cases} \qquad (3)$$

$$J = \frac{1}{n*m}\sum_{i,j=1}^{n,m}\frac{y_{ij}*\bar{y}_{ij}}{y_{ij}+\bar{y}_{ij}-y_{ij}*\bar{y}_{ij}}, \qquad (4)$$

where y_{ij} is the true pixel value, $\overline{y_{ij}}$ is the model predicted value, $n * m$ is the image size. The TABLE 1 below shows a comparison of the results of several experiments: using standard augmentation, additional augmentation and solution [11]. This metrics show that the best building segmentation result was obtained when we used additional image augmentations during network learning.

VII. DISCUSSION AND CONCLUSION

A few examples of building segmentation in Fig. 2.

It should be noted that to improve the result of the selection of discrete objects in satellite images, various transformations of images should be used. In our case, changing the brightness and saturation of colors, adding noise, changing the perspective of the image helped to increase the resulting metric (Jaccard coefficient) from 74.12 to 75.78.

The use of other neural networks can be considered as further improvements. For example, instead of the pre-trained ResNet34 network, take DenseNet or SE-ResNet as a basis. Try to predict the result using an ensemble of several networks, that is, determine the class of each pixel not according to the output of one of the networks, but based on a certain rule and the output of several networks at once. And to practice more with image augmentations.

Fig. 2. Some recognition results

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

TABLE 1. Results Comparison Table

Model described above with standard augmentations						
Metrics	***Bellingham***	***Bloomington***	***Innsbruck***	***San Francisco***	***Tyrol***	***Overall***
Jaccard index	69.03	73.44	74.50	75.02	76.54	**74.12**
Accuracy	96.89	97.45	96.88	91.31	97.87	96.08
Model described above with additional augmentations						
Jaccard index	69.95	75.19	75.46	77.29	77.69	**75.78**
Accuracy	96.96	97.61	97.06	92.26	98.00	96.38
Solution [11]						
Jaccard index	69.75	72.04	74.64	74.55	77.40	**73.91**
Accuracy	96.77	97.13	96.83	91.14	97.92	95.96

REFERENCES

[1] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI, Springer, 2015, pp. 234–241.

[2] V. Iglovikov, S. Mushinskiy, V. Osin, Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition, arXiv preprint, 2017, 6 p.

[3] V. Iglovikov, A. Shvets, TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation, arXiv preprint, 2018, 5 p.

[4] O. Russakovsky, J. Deng, H. Su, et al., ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision – Springer, 2015, pp. 211–252.

[5] E. Maggiori, Y. Tarabalka, G. Charpiat, P. Alliez, Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark, IEEE International Symposium on Geoscience and Remote Sensing Symposium (IGARSS), July 23-28, IEEE, 2017, pp. 3226–3229.

[6] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for

[7] Image Recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, June 27-30, IEEE, 2016, pp. 770–778.

[8] T. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, Focal Loss for Dense Object Detection, IEEE International Conference on Computer Vision (ICCV), Venice, 22-29 Oct., IEEE, 2017, pp. 2999–3007.

[9] D. Kingma, J. Ba, Adam: A method for stochastic optimization, International Conference on Learning Representations, Banff, Canada, April 14-16, 2014, 15 p.

[10] A. Buslaev, A. Parinov, E. Khvedchenya, V. I. Iglovikov, and A. A. Kalinin, Albumentations: Fast and flexible image augmentations, Information, MDPI, 2020, vol. 11, 4 p.

[11] Girard, N., Polygonal Building Segmentation by Frame Field Learning, Nicolas Girard, Dmitriy Smirnov, Justin Solomon, and Yuliya Tarabalka, arXiv preprint, 2020, 30 p.

[12] Google Colaboratory [Electronic resource]

[13] https://colab.research.google.com. Date of access: 06.09.2020

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Preprocessing of SAR Image for Building Detection

Egor Busko
FAMCS of
Belarusian State University
Minsk, Belarus
busko.egor@gmail.com

Alexander Nedzved
FAMCS of
Belarusian State University
Minsk, Belarus
NedzvedA@tut.by
ORCID 0000-0001-6367-5900

Shiping Ye
Zhejiang Shuren University
Hangzhou, China
zjsruysp@163.com
ORCID 0000-0002-9771-7168

***Abstract.* In this paper, we propose high-resolution SARS image processing algorithm to improve quality of the YOLO v4 algorithm by special processing and alignment of different polarization channels.**

***Keywords:* SAR image, buildings detection, image processing, high-resolution SARS images**

I. INTRODUCTION

Detection and highlighting buildings in satellite images is very useful for many applications. It’s helpful in building maps, creating a territory building plan, finding malicious and illegally constructed objects, etc. The main difficulty of the SAR image analysis task lies in the large number of different structures recognition. The task is complicated cause of the various shape, color characteristics and size of the objects.

The basis of SAR image establishment is the reflection features of the scattering of the emitted radar signal by various surface types [5]. The total intensity of signal reflection (pixelgray value) is affected by the characteristics of flatness and regular properties of objects. As a result of processing, the "raw" SAR data is converted into a gray image. The gray value in a pixel in SAR images, which is not affected by lighting, chemical composition (except salt and ice) and temperature of objects, depends on three factors: the SAR system, the SAR processing, and the properties of object. So, an object can be classified by properties of it, such as geometry, dielectric properties, and so on. Volumetric objects (for example, vegetation or other index decoration) correspond to an average level of gray value and texture, surfaces (for example, a calm water surface) shifts brightness to a dark value and buildings to be a bright value. The dielectric properties of the material affect the intensity of the reflected signal. The difference in the coefficient values for different materials makes it possible to identify by SAR.

There have been developed many algorithms for recognizing buildings in optical satellite images. Most of those algorithms are based on the analysis of object shape, texture, shadow, s boundary, etc. [1, 2]. Recently, neural network methods have been used for buildings extraction in optical satellite images [3]. The training dataset consists of satellite original images and their masks. They are binary images or contours of regions, where the pixel value corresponds to classes of objects. But there are only a few neural networks that can get quality results, for example, Fully convolutional network (FCN) [8], Mask R-CNN [9], CNNs [10]. Such methods have been successfully employed in computer vision and remote sensing fields for optical image classification, but few applied to SAR images.

The preparing of mask or region plays a very important role in quality of detection by neural network method. The size of mask is depending on size of shadow. Therefore, each building should be covered by a mask that changes according to patterns of shadow.

In this paper we propose preprocessing of SAR images to detect buildings for analysis by neural networks, where the training areas for detection are selected for each building whose shadow is detected by using the shadow shape. It allows us to improve recognition of regions of buildings The experiments were performed on set of high-resolution SARS images.

II. BUILDINGS REPRESENTATIONS IN SAR IMAGES

Buildings in SAR images have their own characteristics that allow to be detected. Each type of building has its own characteristics of shadow. So, the geometry of shadow is pattern for buildings recognition.

The basic feature for building detection in SAR image is shadow. The structure of shadow depends on features of buildings. The shadow, which is formed by location of building details with conductive properties angle and speed of satellite motion reflections of waves from corners of construction element of building.

The paper [5] proposed that SAR building shadow is described by geometrical point of view, then defined an evaluation function implementing the ratio of exponentially weighted averages (ROEWA) which is used for the matching between the predicted structure and the observed SAR image.

The projection of the three-dimensional building into the two-dimensional slant image plane influence shadow and produces effects such as layover, shadow and foreshortening. In addition, there are specular

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

reflections for close building for urban regions. The paper [5] proposes dense alignment for the buildings images along the radar look of sight (RLOS) for compensation the multi-path reflection effects [6] like as in Fig 1.

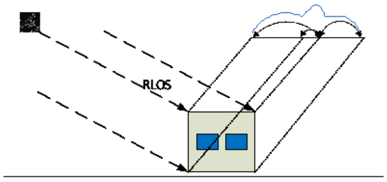


Fig. 1. Geometry scheme of SAR image generation for one floor building with flat roof

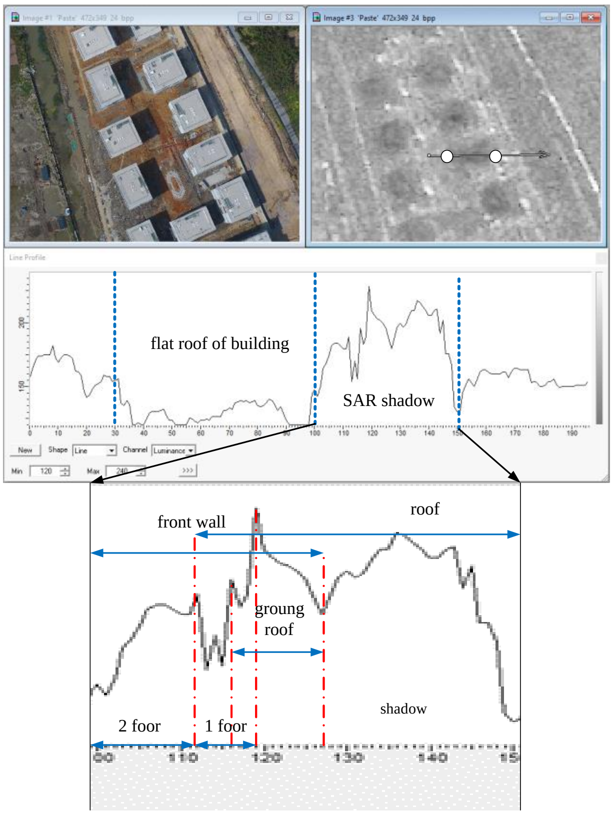


Fig. 2. Sample of profile of building on SAR image with shadow description

Profiles of shape of building shadow in the SAR image are divided into two basic types: the flat roof and the gable roof building. For example, the single bounce generated by the isolated flat roof of one floor building reflections from the ground, front wall and roof. The structure of one store building with flat roof is not complex. For multistory building, brightness profile has more complex structure, like as Fig. 2.

For urban buildings shadow has regular character where each local maximum corresponds to one floor.

Azimuthal and lateral resolution has different concepts. In this case, there is different brightness representation, but the structural elements of the SAR shadow are preserved.

III. PREPARING OF DATASET FOR TEACHING PROCEDURE

A. *Data creation*

Common SAR image processing consists of seven basic steps: creation, calibration, multilooking, selection of region of interest, indexing, segmentation and classification [7].

The first step is data creation (Fig 3). Imperially, it was found that the double polarization of VV+VH [4] allows to obtain more accurate results than the double polarization of HH+HV. Therefore, we use only combination VV+VH. In this step two branchs of processing are created for every type of SAR polarization.

Fig. 3. Sample of contrasting of a SAR satellite image

B. *SAR Image Calibration*

The second step is calibration. The calibration radiometrically corrects the SAR image that the pixel values are changed to the backscattering value of the radar beam from the reflecting surface. It is automatically determined based on satellite image metadata. Calibration radiometrically corrects a SAR image so that the pixel values truly represent the radar backscatter of the reflecting surface. The calibration corrections are realized by the SNAP software that automatically determine what corrections need to be applied to the image. Calibration is essential for quantitative estimation of SAR images.

Multilooking is used to produce a product with a conditional pixel size of the image. Its accumulation is formed by averaging the pixel resolution in range and azimuth, increasing radiometric resolution, but deteriorating spatial resolution. As a result, the image has an approximate square pixel size. It corresponds to conversion from inclined range to ground range.

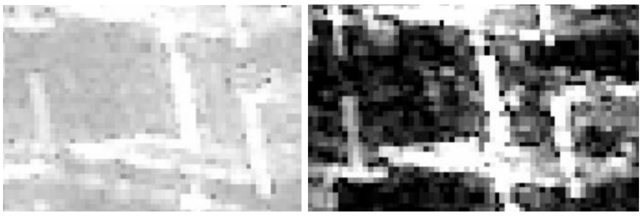
The multilooking is an optional step. It is not required when the image is adjusted for terrain. But we do it in common processing scheme.

Speckle reduction remove specific noise that is caused by random constructive and destructive interference during construction of the image. The resulting image of a particular pixel is obtained by adding a set of values from antenna sensors. Speckle noise is represented as graininess caused by chaotic alternation of light and dark pixels. The presence of it makes difficult to analyze radar images. Speckle filters are applied to SAR images to reduce the amount of speckle at the cost of blurred features.

The order of recording the return signal values in radar sensing of the earth's surface depends on the direction of motion of the satellite and the direction of sight, as a result of which the original images are not always correctly oriented relative to the cardinal directions. Terrain Correction is decoding of the image by correcting SAR geometric distortions. It includes geocoding and orthocorrection.

Geocoding is coordinate reference of the original or converted radar image without removing distortions for the relief. Such transformation converts an image from Slant Range or Ground Range Geometry into a Map Coordinate System.

Orthocorrection includes not only the coordinate reference, but also the elimination of distortions associated with the terrain, which uses a digital elevation model. As a rule, geocoding and orthocorrection of radar images is performed by orbital data. Such transformation is used a Digital Elevation Model (DEM) to correct for inherent SAR geometry effects on images such as shadow, foreshortening, layover. For terrain correction use range-Doppler algorithm.



a) b)

Fig. 4. Image of the building: a) before contrasting, b) after contrasting

Then SAR image usually contrasted for best representation. This operation corresponds to expression of intensity like as logarithmic transformation. It is possible spend analysis and comparation SAR images after such transformation because all distortions should be corrected (Fig. 4).

C. *Selecting data samples by region of interest*

The ROI selection is important step that depends on basic task. It is necessary doing because usually SAR satellite image has very big size and it is not possible to download it memory of computer. We use traditional way where the image is cut into tiles.

We use the method of cutting the image into sections. For realization of some functions, we use software QGIS [11]. The subdivided areas represent neighborhoods of houses with small buildings. We also left one skyscraper in one image.

After we have sliced our large satellite image into many small images, we need to label them. We will train the YOLO network to recognize 3 classes of houses: small cottages (private_building), multistory houses (multistory_building) and tall skyscrapers (skyscraper).

Using the tool for graphic annotation of images LabelImg [12] we mark our images with classes (Fig. 5).

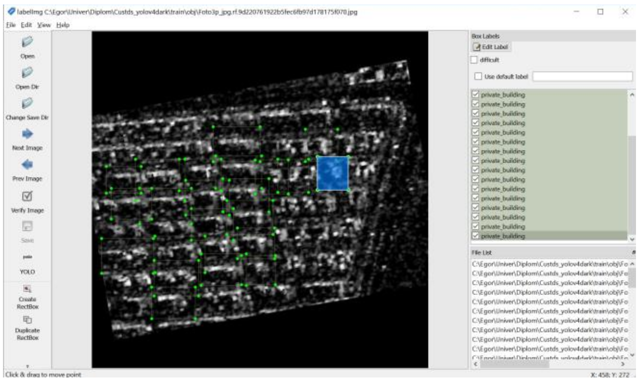


Fig. 5. Example of image marking

With the help of this utility we select our objects with rectangles. Each shape contains information about the location of an object in the image.

After labeling, each new image is supported by an annotation txt file in the same directory and with the same name, which includes the object's number and coordinates of the object in this image, for each object on a new line: <object-class> <x> <y> < width> <height>. The first field object-class is an integer representing the class of the object. It ranges from 0 to (number of classes - 1). In our current case since we have 3 classes from 0 to 2. This annotation was classified into the following classes:

- multistory_building;
- private_building;
- skyscraper.

The second and third x and y records are the x and y coordinates of the center of the bounding box, normalized (divided) by the image width and height, respectively. The fourth and fifth records, width and height are respectively the width and height of the

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

bounding box, again normalized (divided) by the width and height of the image, respectively.

We do not have classes of wooden and metal houses, due to the lack of a training set. Such objects are detected by the brightness threshold. Wood and metal buildings can be identified using a uniform brightness control area with a high value for metal buildings and a low value for wood structures. For metal buildings, this is a small range of maximum brightness values of about 5% of the brightness histogram. For wooden buildings, this is a small range of maximum brightness values of about 10% of the brightness histogram.

IV. BUILDING DETECTION

A. Segmentation and classification

Two last steps of building detection are segmentation and classification. They are very complicated procedures that have many realizations.

The segmentation is the process of partitioning a SAR image into multiple regions (connected sets of pixels, that correspond to objects). The goal of segmentation is to simplify the representation of an image into more meaningful to analyze. The classification is defining visual content to segmented regions. It is final step for detection building on SAR image. These two steps can be combined through the use of a CNN.

For our research we use images from RadarSAT-2 satellite. The resolution of such images about 1.7 meter per pixels. Type of such pixels is float.

B. Selection of buildings by CNN

Many detection systems repurpose classifiers or localizers to perform detection by using R-CNN, VGGNet, ResNet, Inception, and so on. They use the model only for images with selected scales in separate location. As rule, such algorithms use images with very high resolution.

YOLO based on a single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region. Such box can include parts of different class of building with SAR shadow. These bounding boxes are weighted by the predicted probabilities. The basic method for detection of buildings we define as YOLO V4. It is multi-object detector. In our solution three types of images (VV, VH and RVI) are merged by concatenation. A layer that concatenates two inputs is along a specified axis, which corresponds to the concatenate layer. The inputs must be of the same shape except for the concatenation axis. Concatenation takes as input a list of tensors, all the same shape except for the concatenation axis, and returns a single tensor, the concatenation of all inputs. It converted multiple inputs into same shape layer. It is possible to realize by adding additional Dense layer to inputs of VV image and to result of such merging. The same shape outputs are used for every concatenation layer and for input. The basic architecture YOLO is shown at Fig. 6.

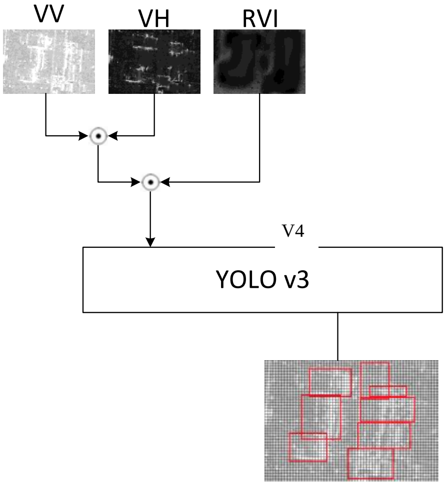
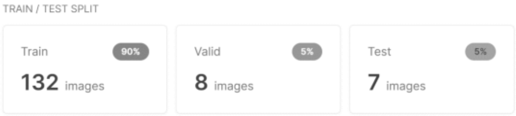


Fig. 6. Modification of YOLO v4 network for VV, VH and RVI inputs

C. Formation of a dataset for the YOLOv4 network

We will use the tool for systematization, preparation and improvement of training data Roboflow [13]. For the YOLO network, it is desirable to use images whose sizes are multiples of 32. Therefore, this problem was solved by adding black stripes at the edges of the images. In addition, we have a fairly small set of images to train the network. To enlarge the data, we use the rotation of the images by 15, 165, 90, 180, 270 degrees, as well as the reflection vertically and horizontally. All our markings are automatically corrected according to the changes made. After that, we form a dataset, which will consist of a set of images for training the network (train), images for verification (validation) and images for testing (test) our trained network (Fig. 7). We assigned most of the images to the training set (90% percent) and 5% to the testing set.



TRAIN / TEST SPLIT

Train 90%	Valid 5%	Test 5%
132 images	8 images	7 images

Fig. 7. Proportional partitioning of our dataset

You can also find out how many images of buildings of different classes appeared in our dataset (Fig. 8).

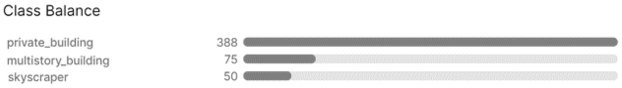


Fig. 8. The number of buildings of different classes in our set

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Most of all there are houses of the private_building class in the SAR image, the skyscraper class turned out to be the rarest. Our datasets are now properly formatted for training and validation.

V. CONCLUSION

For the high-quality operation of the YOLO v4 algorithm, we optimize the data input through special processing and alignment of different polarization channels. The approach described in this paper has proven to be effective for detecting buildings. This approach can be used to detect any discrete objects. The key is to provide the correct training kit. This set must be balanced. The problems with our result are related to the lack of data for training. The dielectric properties of wooden structures form the low brightness of these objects. As a result, timber houses have a low detection rate. Complex buildings are often found not as one, but as several buildings. These problems are very difficult to solve. Thus, if the set is well balanced, then the neural network training result is better.

ACKNOWLEDGMENT

This work is supported by the National Foreign Experts' Program (G20200216025) and Public Welfare Technology Applied Research Program of Zhejiang Province under Grant (No.LGF18F030004, LGJ19F020002 and LGF19F020016).

REFERENCES

[1] G. Cheng, J. Han, A survey on object detection in optical remote sensing images, ISPRS 117, 11–28 (2016).

[2] Salar Ghaffarian, Saman Ghaffarian "Automatic building detection based on supervised classification using high resolution google earth images" International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 2014, 40(3), pp. 101-106.

[3] X. Zhuo, F. Fraundorfer, F. Kurz, P. Reinartz, "Building Detection and Segmentation Using a CNN with Automatically Generated Training Data", IGARSS 2018 – 2018 IEEE Intern. Geoscience and Remote Sensing Symp., Valencia, 2018, pp. 3461-3464. doi: 10.1109/IGARSS.2018.8518521.

[4] Morton John Canty, Image Analysis, Classification and Change Detection in Remote Sensing, 3rd Edition 2014, ImprintCRC Press, 576 p.

[5] Z. Wang, L. Jiang, L. Lin, and W. Yu, "Building Height Estimation from High Resolution SAR Imagery via Model-Based Geometrical Structure Prediction", Progress In Electromagnetics Research M, vol. 41, 2015, pp. 11-24.

[6] A. Ferro, D. Brunner, L. Bruzzone, G. Lemoine, "On the relationship between double bounce and the orientation of buildings in VHR SAR images," IEEE Geosci. Remote Sens. Lett., vol. 8, no. 4, 2011, pp. 612–616.

[7] W. Liu, F. Yamazaki, " Building height detection from high-resolution TerraSAR-X imagery and GIS data," Proceedings of 2013 Joint Urban Remote Sensing Event, Sao Paulo, Brazil, CD-ROM, 2013, pp. 33-36.

[8] K. Bittner, S. Cui, P. Reinartz. "Building extraction from remote sensing data using fully convolutional networks." ISPRS – Intern. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII- 1/W1, 2017, pp. 481-486.

[9] Kang Zhao, Jungwon Kang, Jaewook Jung, et al. Building Extraction from Satellite Images Using Mask R-CNN with Building Boundary Regularization[C], 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2018.

[10] R. Hamaguchi, S. Hikosaka, Building Detection from Satellite Imagery using Ensemble of Size-Specific Detectors[C], 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2018.

[11] QGIS: Free open source geographic information system [Electronic resource], Documentation Center, https://qgis.org/ru/site/. – Date of access: 29.06.2021.

[12] LabelImg: graphical image annotation tool [Electronic resource], Documentation Center, https://tzutalin.github.io/labelImg/. – Date of access: 29.06.2021.

[13] Roboflow Modify Dataset [Electronic resource], Documentation Center, https://app.roboflow.com/datasets. – Date of access: 29.06.2021.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Retinal Image Analysis for Diabetic Retinopathy Grading: Preliminarily Results

Marina Lukashevich
Dept. of Electronic Computing Machines
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
lukashevich@bsuir.by

Yuliya Golub
United Institute of Informatics Problems of NAS of Belarus
Minsk, Belarus
6423506@gmail.com

Aun Irtaza
Dept. of Computer Science
University of Engineering and Technology
Taxila 47050, Pakistan
ORCID 0000-0001-7757-5839

Valery Starovoitov
United Institute of Informatics Problems of NAS of Belarus
Minsk, Belarus
ORCID 0000-0001-7190-761X

Abstract. **Many eye diseases manifest themselves in the retina. Advances in Artificial intelligence (AI), especially in deep learning, improve pathological image analysis in routine clinical practice. We have developed a method for retina image analysis, based on image preprocessing stage and deep neural network as machine learning model. This is the preliminary results of big project for retina image analysis. The main focus was done for diabetic retina images. Also the scheme of new technology for retina image analysis was presented.**

Keywords: **diabetic retinopathy, retina image analysis, image preprocessing, machine learning, convolution neural network**

I. INTRODUCTION

There are many different eye diseases. The fund's image analysis is used to detect different stages of the 3 main diabetes-based eye diseases, i.e. Diabetic Retinopathy, Diabetic Macular Edema, and Glaucoma. In our research, we have considered Diabetic Retinopathy because it causes blindness (the most severe consequences). Diabetic retinopathy (DR), also known as diabetic eye disease, is a medical condition in which damage occurs to the retina due to diabetes mellitus. It is a leading cause of blindness. Diabetic retinopathy affects up to 80 percent of those who have had diabetes for 20 years or more. Diabetic retinopathy often has no early warning signs. Digital pathology nowadays plays an increasingly important role in basic, translational, and clinical pathology research and in routine clinical practice. Retinal (fundus) photography with manual interpretation is a widely accepted screening tool for diabetic retinopathy, with performance that can exceed that of in-person dilated eye examinations.

The clinical signs of DR include (1) multiple cotton wool spots (accumulations of axoplasmic debris within adjacent bundles of ganglion cell axons); (2) venous beading and/or looping; (3) microaneurysms (deep round and blot haemorrhages); (4) hard exudates (lipid deposits); and (5) intraretinal microvascular abnormalities (dilated preexisting capillaries). Artificial intelligence (AI), particularly machine learning (ML), has been widely applied to the pathological image analysis and has provided significant support for medical research and clinical practice. ML can infer image analysis rules from data representations and typically does not require manual algorithm adaptation to different data sets or images [1, 2].

II. DEVELOPMENT OF NEW TECHNOLOGY FOR RETINA IMAGE ANALYSIS

A. *The Scheme of New Technology*

Development of new technology for retina image analysis is the main goal of our international scientific research. The plan for this issue have been developed. It includes the next stages:

1. *Collecting dataset of retina images* from public datasets. It is necessary to collect a variable dataset of retina images obtained in different conditions and using different equipment. Thanks to this it is possible to develop a new technology, but not a new method or approach suitable for limited amount of images.
2. *Quality image analysis.* This stage will allow to select images for the next stages of image processing and decision making. It will increase the quality of decision making in the framework of retina image analysis.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

3. *Image preprocessing.* Low quality of retina images is the main reason of preprocessing stage. This stage can include different options. Main of them is image filtering, circle (mask) cropping, brightness correction, etc.
4. *Machine learning model development* for decision making (image classification). At this stage the main direction for our future research is a neural network approach (deep learning). And the promising approach is using the convolution neural network.
5. *Model evaluation* using different metrics. The vast majority of retina image dataset is unbalanced. It needs to find a suitable metric for unbalanced classes.

The scheme of new technology development for retina image analysis is presented on Fig. 1. We have marked those stages that were explored during the first preliminary stage of our project. Also the next stages and directions of scientific researches are presented in this scheme.

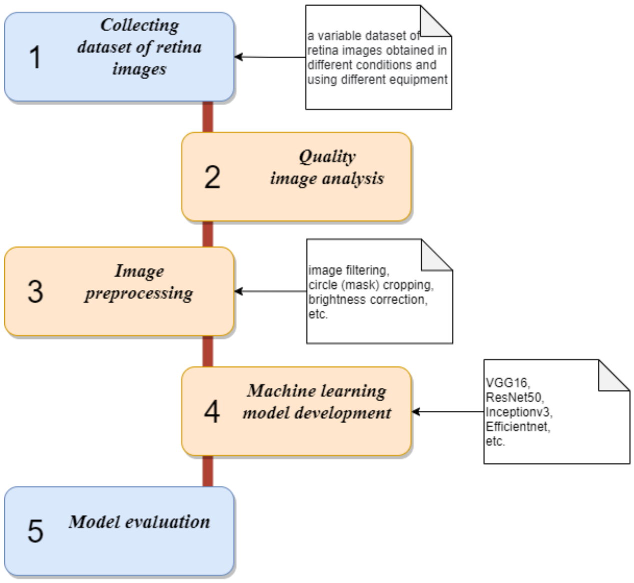


Fig. 1. The scheme of new technology development for retina image analysis

B. Quality Image Analysis

For quality retina image analysis 16 measures described in foreign and domestic literature were selected: BEGH, BISH, BREN, CMO, CURV, FUS, HELM, EBCM, KURT, LAPD, LAPL, LAPM, LOCC, LOEN, SHAR, WAVS [3]. 8 tests were carried out for the correspondence of the obtained values of local quality measures to the normal distribution. Experiments have shown that local estimates do not correspond to the normal distribution of the data. Therefore, Weibull distribution parameters were used to assess the quality of retinal images. The parameter scale (scale) of the FUS measure allows you to divide images into two classes - satisfactory and unsatisfactory quality. Fig. 2 shows examples of retinal images and quality values for each image.

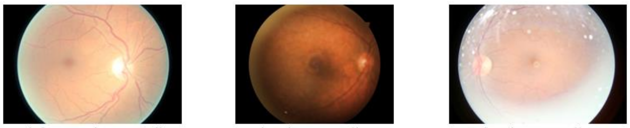

satisfactory picture quality — 0,9068
low image quality (too dark) — 0,6509
low image quality (overexposure and noise) — 0,0024

Fig. 2. The examples of retinal images and quality values for each image

III. EXPERIMENTAL ENVIRONMENT

The main goal of our research on this stage was to develop machine learning model for classifying samples from retina image dataset into 5 classes (0 – No DR, 1 – Mild, 2 – Moderate, 3 – Severe, 4 – Proliferative DR). There are two main approaches in image analysis and classification: traditional approach and deep learning approach. In the framework of traditional approach we perform feature extractions and decision making [4]. In deep leaning neural network extracts features and makes decision themselves. Deep neural networks consist of a large number of layers, have complex and difficult for interpretation. But it is a good variant for large datasets with variable samples. We have selected deep learning approach for this research, because it is promising direction for variable data.

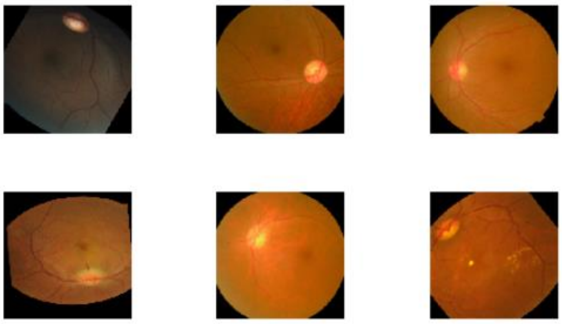


Fig. 3. Examples of the images from examineted dataset

We obtained images for training, validation and testing from a Kaggle competition [5]. It provides a large set of retina images, taken using funds photography under a variety of imaging conditions. A clinic has rated each image for the severity of diabetic retinopathy on a scale of 0 to 4 (0 – No DR, 1 – Mild, 2 – Moderate, 3 – Severe, 4 – Proliferative DR). Like any real-world data set, noise is presented in both the images and labels. Images may contain artifacts, be out of focus, underexposed, or overexposed. The images were gathered from multiple clinics using a variety of

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

cameras over an extended period of time, which will introduce further variation. The examples of the images from this dataset are shown on Fig. 3.

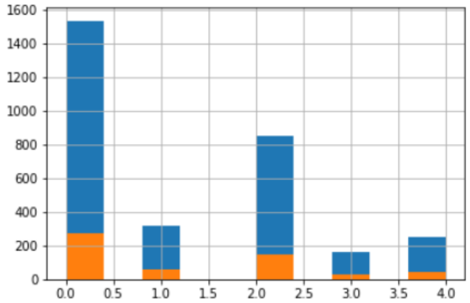


Fig. 4. The amount of samples from each classes (5 classes)

The dataset is highly imbalanced, with many samples for level 0, and very little for the rest of the levels. It was divided into two parts for training and testing machine learning model, Fig. 4.

IV. METHODOLOGY OF EXPERIMENTS

A. *Retina Images Preprocessing*

As we mentioned above retina images contain different types of noises and artefacts. That is why the first stage of the proposed method is image preprocessing. Improvement of the quality of input data can improve the performance of machine learning model. This stage includes reducing lighting-condition effects and cropping uninformative area. The registrations of digital retina images were conducted with many lighting conditions. Some images are very dark and difficult for visualization.

Image smoothing techniques help in reducing the noise. Using different image prospecting libraries, image smoothing (also called blurring) could be done in many ways. In this method, we have performed image smoothing using the Gaussian filter with sigma parameter equal 10. Gaussian filters have the properties of having no overshoot to a step function input while minimizing the rise and fall time. In terms of image processing, any sharp edges in images are smoothed while minimizing too much blur [6].

Cropping is a quite typical step for such kind of images. To solve our case, one method would be to look for rows and columns that have at least one pixel along rows and columns that is greater than some lower limit or threshold as a pixel value. So, if we are sure that the *black* areas are absolute zeros, we can set that threshold as *0*. Thus, if *img* represents the image data, we would have correspondingly two boolean arrays: *(img>tol).any(1)* and *(img>tol).any(0).*

The results of this stage are presented on Fig. 5 (several samples of images before preprocessing) and Fig. 6 (several samples of images after preprocessing).

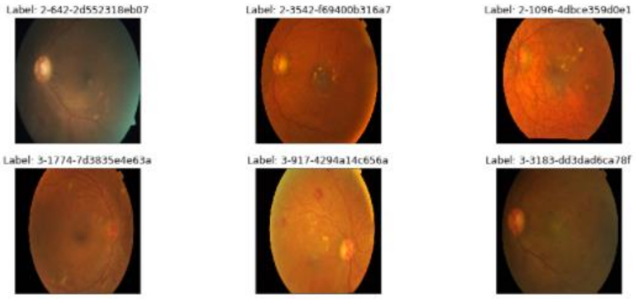


Fig. 5. Samples of retina images before preprocessing

B. *Experimental Details*

Colaboratory from Google and Keras framework were used for experiments. Colaboratory, or "Colab" for short, allows to write and execute Python in browser, with zero configuration required, free access to GPUs and easy sharing. It is a good choice for scientific recaches in the fild of AI [7]. Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow. It was developed with a focus on enabling fast experimentation.

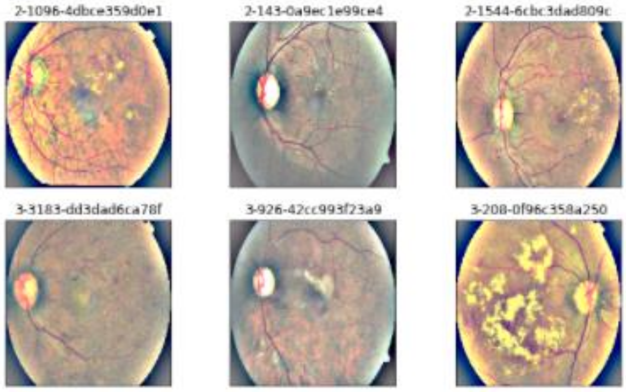


Fig. 6. Samples of retina images after preprocessing

C. *Machine Learning Model Development and Evaluation*

Deep learning approach has become more popular in digital image processing. There are a lot of promising model for classification tasks. Since AlexNet won the 2012 Image Net competition, convolution neural networks have become more accurate by going bigger. While the 2014 ImageNet winner GoogleNet achieves 74.8% top-1 accuracy with about 6.8M parameters, the 2017 ImageNet winner SENet (achieves 82.7% top-1 accuracy with 145M parameters. Recently, GPipe further pushes the state-of-the-art ImageNet top-1 validation accuracy to 84.3% using 557M parameters: it is so big that it can only be trained with a specialized pipeline parallelism library by partitioning the network and spreading each part to a different accelerator. There are several state-of-art CNN models suitable for our

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

task: VGG16 [8], ResNet50, Inceptionv3 and EfficientNet. We are planning to exam all of these architectures. In the framework of our preliminary stage we have built machine learning model using VGG16.

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is an annual computer vision competition [9]. Each year, teams competed on two tasks. The first is to detect objects within an image coming from 200 classes (object localization). The second is to classify images, each labeled with one of 1000 categories (image classification). This model won the 1st and 2nd place on the above categories in 2014 ILSVRC challenge. This model achieves 92.7% top-5 test accuracy on ImageNet dataset which contains 14 million images belonging to 1000 classes. VGG-16 architecture map is presented on Fig. 7.

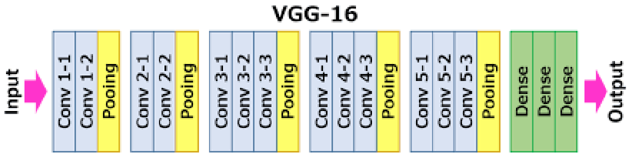


Fig. 7. VGG-16 architecture map

The input to the network is image of dimensions (224, 224, 3). The first two layers have 64 channels of 3*3 filter size and same padding. Then after a max pool layer of stride (2, 2), two layers which have convolution layers of 256 filter size and filter size (3, 3). This followed by a max pooling layer of stride (2, 2) which is same as previous layer. Then there are 2 convolution layers of filter size (3, 3) and 256 filter. After that there are 2 sets of 3 convolution layer and a max pool layer. Each has 512 filters of (3, 3) size with same padding. This image is then passed to the stack of two convolution layers. In these convolution and max pooling layers, the filters use filters with the size (3, 3). In some of the layers, it also uses (1, 1) pixel which is used to manipulate the number of input channels. There is a padding of 1-pixel (same padding) done after each convolution layer to prevent the spatial feature of the image.

After the stack of convolution and max-pooling layer, there is a (7, 7, 512) feature map. We flatten this output to make it a (1, 25088) feature vector. After this there are 3 fully connected layer, the first layer takes input from the last feature vector and outputs a (1, 4096) vector, second layer also outputs a vector of size (1, 4096) but the third layer output a number of classes. Then after the output of 3rd fully connected layer is passed to softmax layer in order to normalize the classification vector. All the hidden layers use ReLU as its activation function. ReLU is more computationally efficient because it results in faster learning and it also decreases the likelihood of vanishing gradient problem [10].

The pre-trained model was trained on ImageNet and not on medical images. In experiments all layers for image feature extraction are used. The decision making layers of pre-trained model were deleted. Batch normalization, flatten and dense layers were added in a new model, Fig. 8.

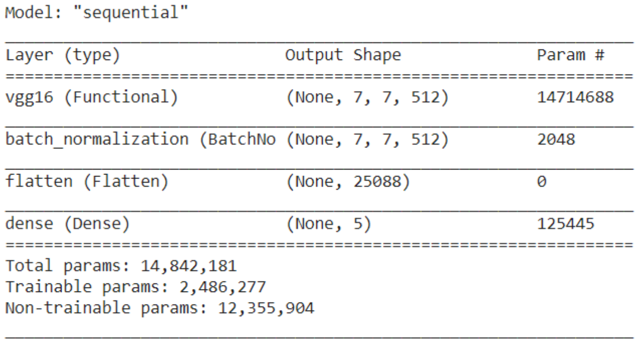

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
vgg16 (Functional)           (None, 7, 7, 512)         14714688
_________________________________________________________________
batch_normalization (BatchNo (None, 7, 7, 512)         2048
_________________________________________________________________
flatten (Flatten)            (None, 25088)             0
_________________________________________________________________
dense (Dense)                (None, 5)                 125445
=================================================================
Total params: 14,842,181
Trainable params: 2,486,277
Non-trainable params: 12,355,904
_________________________________________________________________
```

Fig. 8. Used neural network model

The initial image dataset was divided into two sets for training, and testing (70% and 30 %). The validation set consists of 100 images. Details about training processing are presented on Fig. 9, 10. Numerical results are presented in Table I.

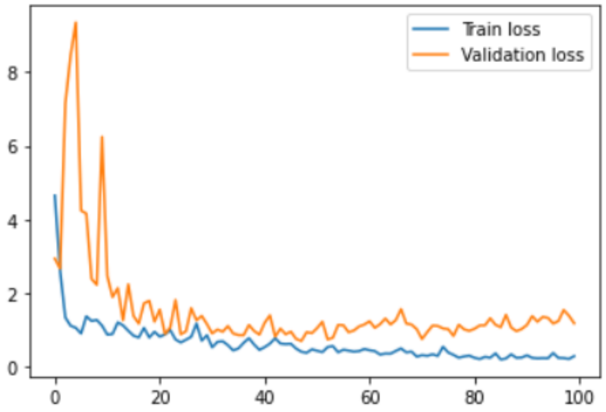


Fig. 9. Model Loss on Training and Validation Datasets

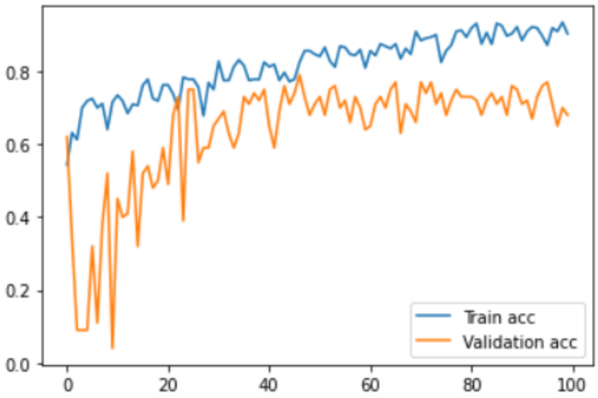


Fig. 10. Model Accuracy on Train and Validation Dataset

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

TABLE I. TABLE TYPE STYLES

Model	Pre-trained dataset	Precision	Recall
VGG16	ImageNet	0.9230	0.8763

V. DISCUSSIONS

We have obtained preliminary acceptable results from the first phase of the study. It is obvious that the deep learning approach is a priority in computer vision in general and in our applied problem in particular. The neural network architecture used is not the best, but it allows you to get preliminary results and understand the range of problems. It is also necessary to carefully implement the stages of image preprocessing (noise reduction), since this may impair the classification results.

VI. CONCLUSIONS

The method for retina image analysis has been developed, based on image preprocessing stage and deep neural network as machine learning model. This is the preliminary results of big projects for retina image analysis. The main focus was done for diabetic retina images. Also the scheme of new technology for retina image analysis was presented.

ACKNOWLEDGMENT

The study was supported in part by the Belarusian Republican Foundation for Fundamental Research under Grant F21PACG-001.

REFERENCES

[1] V. Gulshan, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, Jama, 2016, 316 (2), pp. 2402-2410.

[2] R. Gargeya and T. Leng, “Automated identification of diabetic retinopathy using deep learning,” pp. 962–969, 2017.

[3] Y. I. Golub, F. V. Starovoitov, V. V. Starovoitov Comparative analysis of no-reference measures for digital image sharpness assessment. Doklady BGUIR. 2019; 7(125): 113-120.

[4] M. Haloi, S. Dandapat, and R. Sinha, “A gaussian scale space approach for exudates detection, classification and severity prediction,” arXiv preprint arXiv:1505.00737, 05 2015.

[5] Kaggle. APTOS 2019 Blindness Detection (Data); 2019. Available from: https://www.kaggle.com/c/aptos2019-blindness-detection.

[6] L. G. Shapiro, G. C. Stockman, "Computer Vision", Prentice Hall, 2001, pp. 137, 150.

[7] Welcome To Colaboratory - Google Research. Available from: https://research.google.com/colaboratory/

[8] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition - arXiv preprint arXiv:1409.1556, 2014.

[9] The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) Available from: https://www.image-net.org/challenges/LSVRC/

[10] Abien Fred Agarap, Deep Learning using Rectified Linear Units (ReLU) - arXiv preprint arXiv: 1803.08375, 2018.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Correctness of Control Systems with Concurrency Behavior

Liudmila Cheremisinova
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
cld@newman.bas-net.by

Dmitry Cheremisinov
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
cher@newman.bas-net.by

Abstract. **The discussed problem is to verify whether a reactive control system design with concurrency behavior meets its specification. A model of the desired behavior is in the form of parallel automaton that describes concurrent control algorithms. It is proposed to generate test patterns in the process of simulating the design specification of a concurrent system, which includes an algorithm for the behavior of not only the system itself, but also the environment of the designed device.**

Keywords: **concurrent algorithm, hardware verification, test pattern, simulation, PRALU language**

I. INTRODUCTION

The paper deals with the problem of verification of digital systems with parallelism of behavior. In the functional verification phase, it is established whether the designed device implements the desired behavior, i.e. whether it works according to the requirements set in its specification. The interest in the problem is motivated by the fact that with the growth of the complexity of the designed control systems, the labor costs for their testing also grow. The testing and debugging phase accounts for up to 60 - 80% or more of the total cost of developing control systems.

The most common approach to verification is simulation testing, which requires a test system, which is a specialized software environment that solves three main tasks: generating a test sequence; verifying the correct behavior of the component under test and evaluating the completeness of testing relative to the original specification. A test is a sequence of sets of signal values applied to the input of a device under test and a set of expected signal values generated by it. The purpose of the verification test is to identify errors as a result of situations where the expected results do not coincide with the results of the device under test when the corresponding test sequence is submitted [1].

The quality of testing directly depends on the test sequences used. The methods for constructing test sequences traditionally used in testing practice are based on manual, random and directed test generation. Although such tests allow detecting a significant number of errors in the design, they do not give an estimate of the completeness of the coverage of the operation area of the device under test. In this sense, more effective is the verification of control systems based on model checking [2, 3]. In the case, a test sequence is generated based on a model describing the desired system behavior, which is specified by a device design specification in a certain language. Tests are built on the basis of the specification of the designed system in an algorithmic way; the responses of the device under test are compared with the expected values derived from the specification.

If the model is correct and the device under test must implement the specified behavior (and only it), then the successful passing of tests, generated appropriately based on this model, can serve as a sufficient guarantee of the system correct implementation. The description of the specification is assumed to be correct. The internal structure of an implementation under test can be viewed as a black box. The assessment of the completeness of testing is determined by the degree of coverage of the scenarios of the device operation, specified by the specification.

The verification problem is considered for the case of reactive systems [4]. The peculiarity of these systems (as opposed to systems of transformational type) lies in the continuous (and, in the general case, infinite) exchange of signals with the external environment to accomplish the task. The most popular model of reactive systems is the state machine [5, 6], which describes sequential behavior and is widely used to describe protocols

However, there are a number of systems in which the expressive means of finite state machines are insufficient. The most important property of such systems is the inherent parallelism of the processes occurring in them. The problem of model-based

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

verification for devices with concurrent behavior has not yet been sufficiently studied. One of the most studied types of models of such devices is a system consisting of simultaneously operating components, which is modeled as a network of finite state machines or labeled transition systems (LTS) [3]. Approaches to testing labeled transition systems were proposed [7], in which this problem was considered as checking the system for input-output conformance - ioco relation. The device under test meets the specification in respect to ioco relation if, after any sequence of inputs allowed by the specification, the observed responses of the device under test meet the values expected in the specification.

There are also known approaches to generating test sequences for systems using models of "true parallelism", where some actions are performed in parallel and in the same component, and when it is necessary to control the order of execution. These approaches are based on Petri nets. Test cases for such systems are generated on the basis of the Petri net reachability graph [8], according to which test cases are generated by traversing it [9–11], similarly to how it is done for the case of finite state machines [12]. Reachability graph-based verification is one of the most studied approaches to verifying systems with behavior parallelism. The disadvantage of this approach is the exponential growth of the size of the space of possible states of the system and, accordingly, the size of the reachability graph. As a result, the reachability graph faces the problem of an explosion in the number of states, which negatively affects the performance of testing complex systems.

The paper considers the problem of constructing a test system for verifying the circuit (or software) implementation of a control device with parallelism of behavior. As an example of such systems, control systems in industry, where it is necessary to take into account the parallelism present in control objects, can be mentioned. Within the framework of this system, the specification of the designed device is set in the PRALU language for describing parallel control algorithms [4]. The same language describes the behavior of an object controlled by a designed device. The control object is considered as a part of the test environment. The device implementation is viewed as a black box for which only inputs and outputs are available. Test sequences are formed on the basis of the described algorithms for the behavior of the device and the control object dynamically - in the process of simulating the control algorithm.

II. LANGUAGE TO SET SPECIFICATIONS FOR DESIGNING DEVICES WITH PARALLELISM OF BEHAVIOR

Concurrency in a specification arises for a variety of reasons. For example, it can be a multi-block system in which some actions are performed in parallel, but in different components. And finally, systems with "real parallelism", when some actions are performed in parallel and in the same component,

The problem of designing control devices is one of the most important in the automation of production processes in various industries. When solving the problem of implementing control devices, one has to deal with the parallelism present in control objects. The aim of such object control is to ensure the interacting components to work in parallel and asynchronously in a coordinated manner. The parallelism present in control objects is reflected in the functional model of digital devices that control these objects. It is also inherent in digital devices of this class that control actions and signals about the state of control objects are described by Boolean variables, and only a small percentage of all information is numerical. At present, networks of interacting finite state machines and languages based on Petri net are used as the language for setting the specification for the design of control devices.

To set the specification for the design of devices with parallelism of behavior, it is proposed to use parallel logic control algorithms, which are widely used in the design and testing of digital systems. One of such languages is the PRALU language [4] for describing logic control algorithms. Algorithms in the PRALU language are represented in the form of causal dependencies between events occurring in a technical system, the behavior of which is described in terms of binary variables: control actions and signals about the state of the control object are Boolean variables.

The main operations of the PRALU language waiting are acting operations. The waiting operation "$- k^{in}$" boils down to waiting for the moment in time when the conjunction k^{in} takes the value 1. The acting operation "$\rightarrow k^{out}$" is performed by assigning the variables that form the conjunction k^{out} to values that turn it into 1 In one of the interpretations of the acting operations in the language, it is assumed that all internal variables (if any in the description) and output (or control) variables retain their values until any of the acting operations changes them. The waiting and acting operations can be interpreted as polling the states of the sensors of the control object and issuing commands to the executive and signal equipment.

A control algorithm on PRALU is represented by an unordered set of sentences, each of which opens with a label and consists of one or several equally

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

labeled linear chains of language operations, ending with transition labels: "μ_i: $l_i \rightarrow \nu_i$", where l_i denotes some linear algorithm consisting of language operations; μ_i and ν_i are initial and final labels, which are non-empty subsets of elements from the set $M = \{1, 2, \ldots, m\}$, which can be interpreted as partial states (in the sense that they can exist simultaneously).

The order of execution of the chains in the process of a control algorithm implementation is determined by the start set N [4], its current values are $N_t \subset M$. Among the algorithm proposals, one is distinguished as the initial; its label is entered into the set N before the implementation of the algorithm.

In the process of a control algorithm implementation, the chains are started independently of each other. If at some moment in time for some chain "μ_i: $l_i \rightarrow \nu_i$" the condition $\mu_i \subseteq N_t$ is satisfied and the event k_i^{in} is realized, with the expectation of which the chain l_i begins, then it is started. In this case, N_t is replaced by $(N_t \setminus \mu_i) \cup \nu_i$, and after the end of the chain, the new state of N_t becomes equal to $(N_t \setminus \mu_i) \cup \nu_i$. The syntactically parallel algorithm is characterized by the presence of labels $|\mu_i| > 1$, $|\nu_i| > 1$. An alternative branching is provided by the constraint

$$(i \neq j) \wedge (\mu_i \cap \mu_j \neq \varnothing) \rightarrow (k_i^{in} \wedge k_j^{in}=0).$$

As an example, we will give a parallel algorithm describing the cycle of the manipulator, which consists in moving it between the extreme positions recorded by the sensors r and l. Movements to the left and to the right are initiated by signals L and R, respectively. In the initial position, the manipulator is in position r and starts the working cycle after pressing s button. The manipulator is controlled by the buttons of the control panel "enable" - s and "disable" - e. These buttons can be pressed during the working cycle in any sequence, the manipulator reacts to them only in the position r: it continues the working cycle if the s button was pressed last, or stops if e.

Description of the control algorithm in the PRALU language, defined on the sets $\{s, e, r, l\}$ and $\{L, R\}$ of input (or conditional) and output (control) variables:

```
RUNNING_CYCLE($s, e, r, l$ / $R, L$)
1: $\rightarrow 2.3$
2: $- g \rightarrow L - l \rightarrow \bar{L} \rightarrow R - r \rightarrow \bar{R} \rightarrow 2$
3: $- s \rightarrow g - e \rightarrow \bar{g} \rightarrow 3$
```

The control algorithm is cyclical: once started, it can function indefinitely. Input variables of the algorithm are the variables s, e, r, l, these variables fix the state of the environment. Output variables L, R initiate movement to the left and to the right. In addition, there is one more internal variable g, introduced to remember the fact of pressing the s or e button during the operating cycle of the manipulator: $g = 1$, if s button was pressed last, and $g = 0$, if e.

III. SIMULATION OF REACTIVE SYSTEMS WITH PARALLELISM OF BEHAVIOR

In [4], it was proposed to characterize digital devices by the type of algorithmic description. Devices, the model of which are classical algorithms (scheduling algorithms), belong to the type of transformation. Their purpose is to compute some result from the original data through a finite sequence of steps. Examples of such systems are processors, programming language compilers, web servers.

The purpose of a reactive system [4] is to interact with the environment. The behavior of a reactive system is set by a control algorithm. The functioning of reactive systems, ideally, never ends. It follows that the algorithm of a reactive system is not an algorithm in the sense of the classical theory of algorithms (there is no sign of a finite number of steps). In current literature, control algorithms are called communication protocols. Nevertheless, to formalize these algorithms, one can use the same approach as for transformation systems - description by specifying a formal language and an abstract computational mechanism. Examples of such devices are controllers of computer peripheral devices connected to a common bus, embedded systems and equipment control devices. Recently, the term "reactive system" began to be used to designate software systems in which data streams are processed asynchronously, the volume of which is not predetermined [13].

Traditionally, a protocol has been modeled as a set of interacting processes, where each process is described as an extended state machine that has a finite number of states. In modern verification systems, the interaction of processes is represented as communication, in which the acts of communication are transactions through common data structures called channels. The transaction-level model (TLM) is a performance-enhancing tool (up to 1000 times faster than RTL). The most popular and widely used language for simulating TLM level is SystemC (IEEE 1666 standard), which is an extension of the C ++ language. An executable program that results from compiling a SystemC model with any ANSI-compliant C ++ compiler implements a simulator with integrated simulation controls. Concurrency in SystemC has the semantics of interleaving the operations of sequential processes. SystemC concurrent processes are threads that are scheduled for sequential execution by the native scheduler SystemC based on cooperative multitasking cooperative multitasking.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The multithreading model has significant non-determinism, and SystemC processes are specially organized to eliminate this non-determinism. Atomic operations of processes, which are transactions, are linearizable. Linearizability is a property of a program in which the result of any parallel execution of operations is equivalent to some of their sequential execution [14]. For any other thread, the execution of the linearizable operation is instantaneous: the operation has either not started, or it has completed.

Synchronization of TLM processes at the transaction level is carried out by a barrier mechanism [15]. The barrier is the points in the source code where each process must pause and wait for all processes in the group to reach the barrier. In the SystemC TLM model, barrier points are set by calls to the wait (.) function.

IV. SIMULATION OF DEESCRIPTIONS OF REACTIVE SYSTEMS IN PRALU LANGUAGE

Algorithms in the PRALU language can be interpreted by the TLM model; this requires a refinement of the semantics of waiting and acting operations. The essence of the refinement is to extend the definition of the partial order of the implementation of operations, given by the original parallel algorithm, to a linear order. An interleaving parallelism model is used, in which concurrency is understood as the ability to order operations in an arbitrary way. Waiting and acting operations are considered as compositions of some elementary operations. In this interpretation, algorithms on PRALU have the property of linearizability, i.e. the result of parallel execution of the algorithm is equivalent to some sequential execution of atomic operations. Transactions in algorithms on PRALU are represented by waiting and acting operations that have a common variable and describe the interaction event [16].

The data structure in the TLM model of the PRALU algorithm is a vector of variable values, the components of which are pairs representing the current and planned values for each variable. Access to the vector components of variables is carried out through the operation of setting the planned value of the algorithm variable and the operation of checking the value of the conditional variable. The implementation of a waiting operation for an algorithm on the PRALU consists of the sequential execution of the operations of algorithm suspending and checking the values of the variables in the vector of current values, the acting operation consists of performing the operations of setting the planned values of the variables.

When describing the scheduling procedures for computations associated with the linear ordering of partially ordered operations, the concept of a branch is traditionally used as a set of sequential subprocesses starting with a given operation. A sequential subprocess is usually called the maximum chain of operations of a process that are in a direct sequence. A branch is a dynamic object generated by the operation of its formation and destroyed by the operation of its termination.

Synchronization of concurrent chains of the PRALU algorithm is carried out using a barrier mechanism. Barrier points are set by the operation to suspend execution of branches. The data structure of the synchronization barrier is represented in the memory by the queue QR of branches ready for execution and the queue QW of waiting branches. The operation of forming a branch consists in entering its first operation into the QR. The meaning of the branch termination operation is clear from its name: the branch is removed from the QR.

The fundamental point in simulating algorithms on PRALU is the agreement on the duration of the execution of the operations of the language; in particular, it concerns the acting operations. This convention significantly affects the degree of conformity of signal changes produced by the emulator and appearing at the outputs of the circuit implementation. The implementation (as well as simulation) of PRALU algorithms is performed under some assumption about the duration of the execution of the language operations. The most natural assumption is that all operations (and, in particular, acting operations) have the same duration.

One of the ways to increase the speed of computations is to assume that operations have zero duration. In this case, calculations of operations of one branch are performed until their continuation requires a change in the states of conditional variables. This means that branches will only be suspended while waiting operations are in progress. For a hardware implementation, it is more natural to assume the same, but not zero, duration of the execution of acting operations. In this case, branch execution is suspended after acting operations.

When simulating the control algorithm, branches are sequentially extracted from the QR and executed until suspended. The execution of branch *G* is suspended if^

1) if its initial fragment "$-k^{in} \rightarrow k^{out}$" cannot be executed on the set of current values of algorithm variables;

2) if its initial fragment "$-k^{in} \rightarrow k^{out}$" is already executed.

In the first case, the branch *G* is transferred to the queue QW. In the second case, a new branch is entered into the queue QW, starting with the operation that should be executed next in the branch *G*. The barrier is reached when the queue becomes empty. When the barrier is reached, the following processes are started:

1) transferring elements from the queue QW into the queue QR (QW becomes empty);

2) entering the next values of variables as planned values (if the system is not closed);

3) sending the planned variables values to the current ones for each component of the vector of variables. Then the first operation from the queue QR is started.

Reaching the barrier fixes the clock cycles of the emulator, and the changes in the values of the variables (marked in the vector of the planned values) correspond to the changes in the signal values at the outputs of the circuit implementation of the control system when the signal values corresponding to the values in the vector of the planned values are fed to its inputs. Thus, the verification process of a control algorithm can be performed in two ways:

1) dynamically in the process of debugging the control algorithm;

2) on the test sequence obtained after simulation of the control algorithm.

The transformation of the description of the algorithm in the PRALU language into the TLM model is carried out by translating it into expressions of the intermediate procedural language, which is carried out by substitutions of compositions from elementary operations instead of the operations of the PRALU language [16]. In the PRALU TLM model, there is no need to explicitly specify the barrier points (unlike the SystemC model). The barrier is formed automatically during the broadcast. Synchronizing processes takes longer than computing, especially in distributed computing. The operations of forming, terminating and suspending branches are related to the overhead of organizing computations.

Barrier synchronization is considered to be quite memory and runtime expensive mechanism. However, in the TLM model in the PRALU language, each of the operations of the barrier mechanism can be implemented in modern microprocessors with one command, including the suspension, which is analogous to the wait (.) function in SystemC. A separate scheduler is not required.

V. METHODOLOGY OF CONSTRUCTING A TEST SYSTEM BASED ON THE PRALU LANGUAGE

Let us demonstrate the process of constructing a test sequence in the process of simulation and debugging the specification for the design of a control device given in PRALU language. As an example, consider the above algorithm RUNNING_CYCLE(*s*, *e*, *r*, *l* / *R*, *L*) of manipulator operation.

In the PRALU language, one can describe the functioning of the system as a whole, including not only setting the control algorithm, but also describing the behavior of its environment (as an object of control). This makes it possible to simplify the simulation of the control algorithm, since in this case it is sufficient to change the values of only those variables, the change in the values of which is not fixed in these two algorithms (they are external variables for the system). In our case these are the variables *s* and *e*. The description of the behavior of the environment is as follows:

OC (*R*, *L* / *r*, *l*/)

$1\colon - L \to \bar{r} \to l - R \to \bar{l} \to r \to 1$

The above algorithms on PRALU describe the functioning of the manipulator system as a whole; changes are recorded not only in the internal variable *g*, but also in the values of other variables, except for the variables *s* and *e*. Before starting the manipulator, the variables have the following values: $s = e = 0$, $r = 1$, $l = g = R = L = 0$, $r = 1$, set by the vector 0010000.

Let us demonstrate the process of simulation of the described manipulator functioning for the case of a synchronous implementation of the control algorithm. For this case, 10 branches are allocated in the algorithms:

$1\colon - L \to \bar{r}\,|_4 \to l\,|_5 - R \to \bar{l}\,|_6 \to r \to 1$

$2\colon - g \to L\,|_7 - l \to \bar{L}\,|_8 \to R\,|_9 - r \to \bar{R} \to 2$

$3\colon - s \to g\,|_{10} - e \to \bar{g} \to 3$

Table I shows the cycles of simulation of the PRALU manipulator algorithms. Each of the subtables corresponds to one simulation cycle spawned by the reached barrier B_i. The columns of each of the subtables specify the numbers of the branches G_j, selected from the queue QR at each cycle; the numbers of the branches in the queues QR and QW; the vector of current values of the algorithm variables at the corresponding step C_i, and vectors P_i of the planned values of variables after considering the branches.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The first row of the subtable shows the states of QR, QW, C_i and P_i at the beginning of the corresponding cycle. When passing from the *i*-th cycle to the (*i*+1)-th we set the values to be: $\mathrm{QR}_{i+1} = \mathrm{QW}_i$, $C_{i+1} = C_{i+1} = P_i$. The vector T_i generates with its components corresponding to the conditional and internal variables the test action, and the vector C_i obtained after the execution of the cycle generates a reference response to this action, set by the components corresponding to the control variables.

TABLE I SIMULATION STEPS

Simulation steps	Branches G_j	Queues		Variable values	
		QR_i	QW_i	T_i	P_i
B_1		1,2,3	∅	**1**0 100 00	00 100 00
	1	2,3	1		00 100 00
	2	3	1,2		00 100 00
	3	∅	1,2,10		00 10**1** 00
B_2		1,2,10	∅	00 101 00	00 101 00
	1	2,10	1		00 101 00
	2	10	1,7		00 101 0**1**
	10	∅	1,7,10		00 101 0**1**
B_3		1,7,10		00 101 01	00 101 01
	1	7,	4		00 **0**01 01
	7	10	4,7		00 **0**01 01
	10	∅	4,7,10		00 **0**01 01
B_4		4,7,10	∅	00 001 01	00 001 01
	4	7,10	5		00 0**1**1 01
	7	10	5,7		00 0**1**1 01
	10	∅	5,7,10		00 0**1**1 01
B_5		5,7,10	∅	00 011 01	00 011 01
	5	7,10	5		00 011 01
	7	10	5,8		00 011 0**0**
	10	∅	5,8,10		00 011 0**0**
B_6		5,8,10	∅	00 011 00	00 011 00
	5	8,10	5		00 011 00
	8	10	5,9		00 011 **1**0
	10	∅	5,9,10		00 011 **1**0
B_7		5,9,10	∅	**1**0 011 10	00 011 10
	5	9,10	6		00 0**01** 10
	9	10	6,9		00 0**01** 10
	10	∅	6,9,3		00 0**00** 10
B_8		6,9,3	∅	0**1** 001 10	00 001 10
	6	9,10	1		00 **1**01 10
	9	10	1,9		00 **1**01 10
	3	∅	1,9,3		00 **1**00 10
B_9		1,9,3	∅	00 100 10	00 100 10
	1	9,3	1		00 100 10
	9	3	1,2		00 100 **00**
	3		1,2,3		00 100 **00**
B_{10}		1,2,3	∅	00 100 00	00 100 00

The required tests are represented by pairs of vectors: a five-component vector of the test pattern, the components of which correspond to the values of the variables *s*, *e*, *r*, *l*, *g*, and a three-component vector of reactions, the components of which correspond to the values of the variables *g*, *R*, *L*.

For the simulation fragment, shown in Table 1, the following test sequence is obtained for verifying the control device from the initial state:

10100 / 100; 00101 / 101; 00001 / 101; 00011 / 100;

00011 / 110; 10011 / 010; 01001 / 010; 00100 / 000.

Here the first part of each test sets a test pattern and the second part shows the expected responses of the device under test after feeding it by the test pattern.

VI. CONCLUSION

The behavior of an embedded control device is highly dependent on the object it controls and the environment in which it operates. Simulation of the control device for verification purposes should be carried out in the area of its planned operation. Using the PRALU language to describe control algorithms makes it possible to specify the behavior of the control system as a whole. Currently, there is software support for design automation and debugging of control systems based on PRALU, which includes simulation tools and synthesizers of the PRALU language in the hardware model in the Verilog and C languages [16].

REFERENCES

[1] A. Kamkin, M. Chupilko, “Obzor sovremennykh tekhnologiy imitatsionnoy verifikatsii apparatury” (Overview of modern technologies for simulation verification of equipment), Programmirovaniye, 2011, № 3, pp. 42–49.

[2] L. Hoffman, “Talking Model-Checking Technology”, Communications of the ACM, vol. 51, 2008, № 07/08, pp. 110–112.

[3] Tretmans J. Model based testing with labelled transition systems. Formal Methods and Testing: Lecture Notes in Computer Science (Springer), vol. 4949, 2008, pp. 1–38.

[4] A. D. Zakrevskij, “Parallel'nye algoritmy logicheskogo upravlenija” (Parallel Logic Control Algorithms), Minsk: Institut tehnicheskoj kibernetiki Nacional'noj akademii nauk Belarusi, 1999, 202 p. (in Russian).

[5] D. Lee, M. Yannakakis, “Principles and methods of testing finite state machine – a survey”, Proceedings of the IEEE, vol. 84, 1996, № 8, pp. 1090–1123.

[6] B. Kanso, O. Chebaro, “Compositional testing for FSM-based models”, International Journal of Software Engineering & Applications (IJSEA), vol. 5, 2014, № 3, pp. 9–23.

[7] H. Ponce de Leon, S.H. Delphine Longuet, “Model-based Testing for Concurrent Systems with Labeled Event Structures”, Software Testing, Verification & Reliability, vol. 24, 2014, № 7, pp. 558–590.

[8] H. Watanabe, T. Kudoh, “Test Suite Generation Methods for Concurrent Systems based on Coloured Petri Nets”, Proceedings of the 2nd Asia-Pacific Software Engineering Conference, 1995, pp. 242–251.

[9] U. Farooq, C. P. Lam, H. Li, “Towards Automated Test Sequence Generation”, Proceedings of the 19th Australian Conference on Software Engineering, 2008, pp. 441–450.

[10] H. Zhu, X. D. He, “A Methodology of Testing High-level Petri Nets”, Information and Software Technology, vol. 44, 2002, pp. 473–489.

[11] J. Liu, X. Ye1, J. Zhou, X. Song, “I/O Conformance Test Generation with Colored Petri Nets”, Applied Mathematics and Information Sciences, vol. 6, 2014, № 6, pp. 2695–2704.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

[12] I. B. Burdonov, A. S. Kosachev, V. V. Kuljamin, “Neizbytochnye algoritmy obhoda orientirovannyh grafov. Determinirovannyj sluchaj” (Irredandant algorithms for traversal of directed graphs, The determinate case), Programmirovanie, 2003, № 5, pp. 11–30 (in Russian).

[13] N. Halbwachs, “Synchronous Programming of Reactive Systems”, Springer-Verlag, 2010, 192 p.

[14] M. P. Herlihy, J. M. Wing, “Linearizability: A Correctness Condition for Concurrent Objects”, ACM Trans. Program. Language Systems, 1990, pp. 463–492.

[15] Y. Solihin, “Fundamentals of Parallel Multicore Architecture”, CRC Press, 2015, 494 p.

[16] D. I. Cheremisinov, “Proektirovanie i analiz parallelizma v processah i programmah” (Design and the Analysis of Parallelism in Processes and Programs), Minsk: Belaruskaja navuka Publ, 2011, 300 p. (in Russian).

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Some Trends in Updating the Nomenclature of Specialties of Researchers in the Areas of Research Methods and Means of Intellectualization

Aleksei Rozhnov
V. A. Trapeznikov Institute of
Control Sciences of RAS
Moscow, Russia
rozhnov@ipu.ru

Abstract: **for discussion, some features of understanding modern concepts in the extremely interesting developing field of breakthrough technologies are proposed, with the definition of characteristic (primary, verifiable, reliable, etc.) features and (also and/or later) elements of the so-called concept of "Artificial Intelligence". From the standpoint of eliminating administrative restrictions, the main focus is on popularization and some trends in updating the nomenclature of specialists of researchers in the areas of research methods and means of intellectualization (Artificial Intelligence (AI), Machine Learning (ML), Intelligent Transport Systems (ITS), Robotic Systems (RS), and others).**

Keywords: **artificial intelligence, machine learning, intelligent transport systems, robotic systems, technological gap, administrative decisions, popularization, new technologies**

I. INTRODUCTION

Few States currently have extensive knowledge and significant potential in the field of Artificial Intelligence (AI). At the same time, the development of new technologies based on AI ideas is proceeding at a very high pace, and therefore there are reasonable fears that in the foreseeable future the emerging technological gap will not only not decrease, but, on the contrary, rapidly increase. In addition, although the desired potential is concentrated only in limited subjects, the consequences of the introduction of AI are not limited and will not be limited only to those countries that have similar potential. So, in connection with the dynamic development of the situation that has arisen, a significant number of problematic issues appear, although many of them are obviously outside the IP policy, and are related, for example, to such areas as labor policy, ethics, human rights, etc. [1]. The given highlighted list of issues largely corresponds to the competencies of WIPO – in the context of intellectual property, innovation and the results of creative activity. However, the interesting question is "are there any other important areas of activity or additional ones in the field of artificial and/or hybrid intelligence?". We present several issues currently being discussed from the previously proposed positions of forming responsibility for administrative decisions in the theory and practice of AI [2].

It should be noted that since July 1, 2020, an experiment has been conducted in Moscow to establish special regulations to create the necessary conditions for the development and implementation of artificial intelligence technologies in the subject of the Russian Federation – the federal city of Moscow, as well as the subsequent possible use of the results of the use of artificial intelligence [3].

So, the structure of this discussion report contains the following key information points of AI popularization [4]:

- Artificial Intelligence (AI) and Machine Learning (ML);
- Intelligent Transport Systems (ITS);
- Robots, Mechatronics and Robotic Systems (RS);
- and some aspects of the problems of the Technological Gap in the field of Advanced Systems (new technologies) with reliable signs of Artificial Intelligence.

It should be especially noted that, along with our conference, the workshop "Maturity of artificial intelligence: system integration and management problems" successfully functions permanently as a development of the all-Moscow seminar "Control Science of Autonomous Systems".

II. ACCENTUATION OF RESEARCH DIRECTIONS ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

In the context of the preferences of natural science research, we will begin to review the directions of AI and ML.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Field of science: natural sciences. Group of scientific specialties: 1.2. Computer science and informatics (the name of the branch of science in which academic degrees are awarded: Physical and mathematical sciences). At the current stage, according to the author, the formulations used in the draft nomenclature are very far from the academic level. However, it is precisely such a presentation that will presumably allow us to highlight the most important moments of the development of the subject area.

The section specifies key concepts and definitions, as well as lists the directions of AI and ML research according to the draft nomenclature of scientific specialties.

A. Basic concepts and definitions

For the purposes including this experiment, the relevant Federal Law (RF) uses the following basic concepts:

Artificial Intelligence (AI) is a complex of technological solutions that allows simulating human cognitive functions (including self-learning and finding solutions without a predetermined algorithm) and obtaining results comparable, at least, with the results of human intellectual activity when performing specific tasks.

The framework of such technological solutions includes information and communication infrastructure (including information systems, information and telecommunications networks, other technical means of information processing), software (including those that use ML methods), processes and services for data processing and solution search.

Artificial Intelligence Technologies are technologies based on the use of AI (including computer vision, natural language processing, speech recognition and synthesis, intelligent decision support and promising AI methods).

B. Research directions of the scientific specialty 1.2.1 Artificial Intelligence and Machine Learning

1. Natural science foundations and methods of artificial intelligence.

2. Research in the field of evaluating the quality and effectiveness of algorithmic and software solutions for artificial intelligence and machine learning systems. Methods of comparison and selection of algorithmic and software solutions under many criteria.

3. Methods and algorithms for modeling cogitative processes: reasoning, argumentation, recognition and classification, formation of concepts. Research in the field of neuromorphic methods of data analysis, simulation modeling of the structure and functions of the brain, including using machine learning methods. Neuroinformatics and methods of modeling biological nervous systems.

4. Development of methods, algorithms and creation of artificial intelligence and machine learning systems for processing and analyzing texts in natural language, for images, speech, biomedicine and other special types of data.

5. Methods and technologies for the search, acquisition and use of knowledge and laws, including empirical ones, in artificial intelligence systems. Research in the field of joint application of machine learning methods and classical mathematical modeling. Methods and means of using expert knowledge.

6. Formalization and formulation of management tasks and (support) decision-making based on artificial intelligence and machine learning systems. Development of control systems using artificial intelligence systems and machine learning methods, including control of robots, cars, UAVs, etc.

7. Development of specialized mathematical, algorithmic and software for artificial intelligence and machine learning systems. Methods and means of interaction of artificial intelligence systems with other systems and a human operator.

8. Multi-agent systems and distributed AI.

9. Methods and means of using parallel, quantum computing, etc. for solving artificial intelligence and machine learning problems.

10. Research in the field of ethical problems related to the creation and implementation of AI systems, including modeling of expected social and economic consequences.

11. Research in the field of "strong AI", including the formation of a conceptual base and elements of mathematical formalism necessary for the construction of an algorithmic apparatus.

12. Research in the field of "trusted" AI class systems, including the problems of forming test samples of use cases, reliability, stability, retraining, etc.

13. Methods and means of generating arrays of data and use cases, including "big data", necessary for solving problems of artificial intelligence and machine learning. Problem-oriented data collections for important application areas.

14. Methods and means of forming arrays of conditionally real data and precedents necessary for solving problems of artificial intelligence and machine learning.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

15. Mathematical research in the field of statistics, logic, algebra, topology, function analysis and other fields, focused on solving problems of artificial intelligence and machine learning.

16. Research in the field of special optimization methods, problems of complexity and elimination of iteration, dimension reduction.

17. Research in the field of multilayer algorithmic structures, including multilayer neural networks.

Related specialties (within the group of scientific specialties):

1.2.2 Mathematical modeling, numerical methods and software packages;

1.2.3 Theoretical computer science, cybernetics.

III. SUBSTANTIATION OF RESEARCH DIRECTIONS FOR INTELLIGENT TRANSPORT SYSTEMS AND ROBOTICS SYSTEMS

The section highlights the most interesting points of substantiation of the directions of research of Intelligent Transport Systems (ITS) and Robotics Systems (RS).

Field of science: Technical sciences. Groups of scientific specialties: 2.5. Mechanical engineering; 2.9. Transport systems (the name of the branch of science in which academic degrees are awarded: technical).

A. Research directions of the scientific specialty 2.9.8 Intelligent Transport Systems

1. Architectures of intelligent transport systems and their subsystems.

2. Integration platforms and buses of intelligent transport systems.

3. The life cycle of intelligent transport systems and intelligent road infrastructure.

4. Cooperative intelligent transport systems and their subsystems.

5. Subsystems, on-board and infrastructure telematics of intelligent transport systems.

6. Information and communication technologies and elements of artificial intelligence in intelligent transport systems.

7. Communication systems and means in intelligent transport subsystems.

8. Intelligent systems of technical diagnostics of elements and devices, control, monitoring, management of technological and production processes in transport.

9. Big data management in the transport complex.

10. Connected, highly automated and unmanned vehicles.

11. Elements of intelligent transport infrastructure, technical means of intelligent transport systems.

12. Traffic management and automated vehicle traffic control systems in intelligent transport systems, creation of highly automated and unmanned traffic systems.

13. Promising transport systems based on intelligent passenger and cargo mobility services, built on the tools of intelligent transport systems.

14. Transport planning and simulation of transport systems.

15. Digital models of linear objects of transport infrastructure and transport systems.

16. Digital doubles of roads and transport routes, transport infrastructure, vehicles, transport processes.

17. Systems for ensuring information, functional safety, as well as road safety in intelligent transport systems.

18. Human-machine interface of a highly automated and unmanned vehicle.

19. Transport psychology and psychophysiology.

20. Management of mobility in agglomerations and transport behavior.

21. Regulatory regulation of the development and implementation of intelligent transport systems.

Related specialties (within the group of scientific specialties):

2.9.1 Transport and transport-technological systems of the country, its regions and cities, organization of production in transport;

2.9.3 Railway rolling stock, train traction and electrification;

2.9.4 Management of transportation processes;

2.9.5 Operation of road transport;

2.9.6 Air navigation and operation of aviation equipment;

2.9.7 Operation of water transport, waterways and hydrography;

2.9.9 Logistics transport systems;

2.9.10 Technosphere safety in transport systems.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

B. *Research directions of the scientific specialty 2.5.4 Robots, Mechatronics and Robotic Systems*

1. Development of theoretical foundations and methods of analysis, structural and parametric synthesis and computer-aided design of robots and robotic systems.

2. Theory and methods of creating robots and mechatronic devices based on new physical effects and phenomena, principles and methods of their construction for various conditions and environments of application.

3. Development of principles and methods of construction of mechatronic devices and systems as a result of the synergetic combination of precision mechanics units, electrical, electro-pneumatic, electrohydraulic, electronic and computer components to design and application of qualitatively new machines, systems and modules with highly efficient digital control of their functional motions.

4. Mathematical and semi-natural modeling of mechatronic and robotic systems, analysis of their characteristics based on the results of modeling.

5. Methods, algorithms, software and hardware for controlling robots, robotic and mechatronic systems, including adaptive, optimal, distributed, intelligent and supervisory control.

6. Mathematical and software, computer methods and means of information processing in real time in robots, robotic and mechatronic systems.

7. Methods of experimental research, creation of prototypes and experimental stands for the development of robots, robotic and mechatronic systems.

8. Planning and implementation of movements and actions, individual and group control of mobile robots of land, air, underwater and space-based.

9. Methods of calculation and design of mechatronic servos, executive, sensor and control components of robots, robotic and mechatronic systems.

10. Interfaces and methods of human interaction with robots. Methods of effective and safe joint work of humans and robots.

11. Research, improvement of the efficiency and safety of operation of automated technological processes created based on robotic and mechatronic systems, including in collaborative robotics. Methods and tools for computer-aided design, analysis and optimization of robotic complexes, cells and lines.

Related specialties (within the group of scientific specialties):

2.5.2 Machine Science;

2.5.11 Ground transport and technological facilities and complexes;

2.5.21 Machines, aggregates and technological processes.

IV. PRELIMINARY CONCLUSIONS

A. *Related Research Areas*

Considering the characteristic features of interdisciplinary research for the fields of science (1. Natural Sciences; 2. Technical Sciences), we will distinguish for groups of scientific specialties (1.2. Informatics and Computer Science; 2.5. Mechanical Engineering) related research areas in their respective specialties. For AI and RS – shown in Table I.

TABLE I. RELATED RESEARCH AREAS FOR AI AND RS

No.[a]	**Scientific Specialty**	
	scientific specialty, name	***the code of the nomenclature***
I	Mathematical modeling, numerical methods and software packages	1.2.2
	Theoretical Informatics, Cybernetics	1.2.3
II	Machine Science	2.5.2
	Ground transport and technological facilities and complexes	2.5.11
	Machines, aggregates and technological processes	2.5.21

[a] For the recommendation of scientific specialties in the dissertation councils being created

For interrelated scientific specialties (1.2.1 Artificial Intelligence and Machine Learning; 2.5.4 Robots, Mechatronics and Robotic Systems), respectively, – groups I and II. The summary Table II shows additional groups of scientific specialties in the context of ITS and IT.

TABLE II. RELATED RESEARCH AREAS FOR IT AND ITS

No.[a]	**Scientific Specialty**	
	scientific specialty, name	***the code of the nomenclature***
III	Vacuum and plasma electronics	2.2.1
	Technology and equipment for the production of materials and electronic devices	2.2.3
	Photonics	2.2.7
	Design and technology of instrumentation and radio-electronic equipment	2.2.9
IV	Automation and control of technological processes and production facilities	2.3.2

No.[a]	**Scientific Specialty**	
	scientific specialty, name	***the code of the nomenclature***
	Management in organizational systems	2.3.4
	Informatics and information processes	2.3.8
V	System analysis, control and processing of information	2.3.1
	Management in organizational systems	2.3.4
	Mathematical and software support of computers, complexes and computer networks	2.3.5
	Methods and systems of information protection, information security	2.3.6
VI	Automation and control of technological processes and production facilities	2.3.3
	Mathematical and software support of computers, complexes and computer networks	2.3.5
	Computer modeling and design automation	2.3.7
VII	Transport and transport-technological systems of the country, its regions and cities, organization of production in transport	2.9.1
	Railway rolling stock, train traction and electrification	2.9.3
	Management of transportation processes	2.9.4
	Operation of road transport	2.9.5
	Air navigation and operation of aviation equipment	2.9.6
	Operation of water transport, waterways and hydrography	2.9.7
	Logistics transport systems	2.9.9
	Technosphere safety in transport systems	2.9.10

Thus, we will summarize the preliminary result of comparing related research areas (the field of science:

2. Technical Sciences), for groups of scientific specialties:

2.2. Electronics, photonics, instrumentation and communications (scientific specialty:

2.2.2 Electronic component base of micro-and nanoelectronics, quantum devices);

2.3. Information technologies and telecommunications (scientific specialties:

2.3.1 System analysis, control and information processing;

2.3.2 Computer systems and their elements;

2.3.8 Informatics and information processes) and

2.9. Transport Systems (scientific specialty:

2.9.8 Intelligent Transport Systems). In the table, they are sequentially placed under the numbers from III to VIII. By the time of the conference, the content of the nomenclature can be significantly changed when clarifying the directions of research [4]. Factual data and illustrations will be argued in the presentation of this report. But at the conference, it is proposed to discuss the prospects for cooperation in the creation of joint dissertation councils in the presented areas. Other discussion points, due to the limited volume of the report, it is possible to submit for prolonged discussion within the log of the Workshop “Maturity of artificial intelligence: system integration and management problems,” in ResearchGate [https://www.researchgate.net/~project/Workshop-Maturity-of-artificial-intelligence-system-integration-and-management-problems].

B. Prefinals Clauses

Some features of understanding modern concepts in an extremely interesting emerging field are proposed for discussion: protection of intellectual property – objects related to the problem of breakthrough technologies to a greater or lesser extent with the definition of characteristic (primary, verifiable, reliable, etc.) features and elements of so-called AI systems [2]. The short-term perspective includes specialized accentuation [5–7]: Addresses the issue of Cognitive Semantics’ aspects that cannot be represented by traditional digital and logical means; Discusses necessary conditions for purposeful and sustainable convergence of Decision-Making.

ACKNOWLEDGMENT

The author is grateful to the regular participants of the all-Moscow scientific seminar “Control Science of Autonomous Systems” for the valuable ideas expressed when discussing the local AI problems. The author also thank the anonymous reviewers for the endorsement of the non-trivial target setting of this report on the popularization of AI innovations in the nomenclature of specialties of researchers. The work was partially supported by the RGNF, project No. 17-11-01353.

REFERENCES

[1] WIPO discussion on “Intellectual Property and Artificial Intelligence,” WIPO/IP/AI/3/GE/20, Geneva, Switzerland, November 4, 2020.

[2] Rozhnov, A. “Technological Gap in the Field of New Technologies and Features of Intellectual Property Protection of Systems with Reliable Signs of Artificial Intelligence,” Proceedings of the 28th International Conference “Problems of Security Management of Complex Systems

(“PUBSS'2020,” Moscow), ICS RAS, Moscow, pp. 124–129, 2020, (in Russian), [doi: 10.13140/RG.2.2.25497.03682].

[3] “On conducting an experiment to establish special regulation in order to create the necessary conditions for the development and implementation of Artificial Intelligence Technologies in the subject of the Russian Federation – the federal city of Moscow and amending Articles 6 and 10 of the Federal Law “On Personal Data”,” the Federal Law of the Russian Federation, Adopted by the State Duma on April 14, 2020, Approved by the Federation Council on April 17, 2020.

[4] Draft passports of scientific specialties, nomenclature of scientific specialties (“unpublished”), HAC of the Russian Federation, [URL: https://drive.google.com/drive/folders/1xqoWINSPHH48_IA2Iw1uuWt3qkMQc5E0], (access mode without restrictions).

[5] A. Rozhnov, “Justification of the Export Control Internal Program of Telecommunications and Information Processing Technologies in the Artificial Intelligence Area,” preprint (“in press”), 2021, [doi: 10.13140/RG.2.2.17021.90088].

[6] S. Ablameyko, V. Krasnoproshin, “Development and Current State of the Scientific Direction “Pattern Recognition and Image Processing” in Belarus,” Pattern Recognition and Image Analysis, 31, pp. 117-118, 2021, [doi: 10.1134/S1054661821010028].

[7] A. Raikov, Cognitive Semantics of Artificial Intelligence: A New Perspective, Series “SpringerBriefs in Computational Intelligence,” Springer Singapore, 2020, [doi: 10.1007/978-981-33-6750-0].

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Robustification of Sequential Statistical Decision Rules for Stochastic Data Flows Analysis

Alexey Kharin
Department of Probability Theory and Mathematical Statistics
Belarusian State University
Minsk, Belarus
KharinAY@bsu.by

Ton That Tu
Division of Applied Mathematics
University of Danang
Danang, Vietnam
tthattu@gmail.com

Dai Yukun
Faculty of Applied Mathematics and Computer Sciences
Belarusian State University
Minsk, Belarus
daiyukun0905@gmail.com

Wang Yumin
Faculty of Applied Mathematics and Computer Sciences
Belarusian State University
Minsk, Belarus
wangyumin1994@gmail.com

***Abstract*. In data analysis the issues of statistical decision making on parameters of observed stochastic data flows are important. To solve the relevant problems, here sequential statistical decision rules are used. The sequential statistical decision rules traditionally used loose their performance optimality in situations that are common in practice, when the hypothetical model is distorted. Here the robustified sequential decision rules are constructed for three models of observation flows: independent homogeneous observations; observations forming a time series with a trend; dependent observations forming a homogeneous Markov chain.**

***Index Terms*: sequential decision rule, time series with trend, homogeneous Markov chain, distortion, robustness**

I. INTRODUCTION

In applications, a problem of stochastic data flows analysis is often important [1], aiming decisions on one of two possible modes of a system that generates such a flow. The mode corresponds to a parameter value (or a value of parameters vector) that defines the probabilistic properties of the observed data.

To construct efficient decision rules, one of possible approaches is the sequential statistical analysis [2]. It exploits the key concept concerning the number of observations needed to make a decision with given small levels of error probabilities. It is not fixed a priori, and is defined through the observation process on the basis of the observed random values, so it is random itself [3]. The number of observations is tailored for each situation with the observed data flow, and this feature makes effective the resulting decision rule. There are two admissible decisions after each observation received: to stop the process and to decide in favor of one of two defined hypotheses, or to collect the next observation, as the requested accuracy is not reached at the moment. Although theoretical analysis of sequential decision rules is not trivial, it is widely used in medicine, finance, quality control, and other fields, where the cost of each observation is not ignorable.

Here we use the approach developed in [4], [7], to construct robustified sequential decision rules, i.e. the rules that are robust [4], [5] under distortions of the hypothetical model of data [6]. In other words, their performance characteristics are essentially less influenced by the distortions as compared to the sequential decision rules that are traditionally used.

II. SEQUENTIAL DECISION RULE FOR THE FLOW OF INDEPENDENT HOMOGENEOUS OBSERVATIONS

Let a data flow of independent random vectors $x_1, x_2, \dots$ be observed with a probability distribution P_θ that has the probability density function $p_\theta(x)$, $x \in U \subseteq R^N$, where $\theta \in \Theta = \{0, 1\}$ is a parameter value which is not observed.

There are two hypotheses corresponding to two modes of the system, in terms of the parameter value:

$$H_0: \ \theta = 0, \ H_1: \ \theta = 1.$$

A sequential decision rule is defined as a pair of components: stopping moment rule, and acceptance (terminal decision) rule. Consider a family of sequential decision rules $\delta_\lambda = (\tau_\lambda, d_\lambda)$ based on function $\lambda(\cdot)$: $U \to R$, where

$$\tau_\lambda = \inf\{n: \ \Lambda_n \notin (C_-, C_+)\}$$

is the stopping moment rule (the result depends on $x_1, \dots, x_n$ and that is why it is random), and

$$d_\lambda = 1_{[C_+, +\infty}(\Lambda_n)$$

is the terminal decision rule in favor of the hypothesis H_i, if $d_\lambda = i$, $i \in \{0, 1\}$.

The test statistic is

$$\Lambda_n = \Lambda_n(x_1, \dots, x_n) = \sum_{t=1}^{n} \lambda(x_t), \ n \in N = \{1, 2, \dots\}.$$

The research is supported by the State research program “Digital and space technologies, human and state safety”.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The parameters $C_-, C_+ \in R$, $C_- < C_+$ are called thresholds and calculated according to:

$$C_- = \log \frac{\beta}{1-\alpha},\ C_+ = \log \frac{1-\beta}{\alpha},$$

where α, β are values given by a user for admissible levels of error type I (H_0 is true, but declined) and II (H_1 is true, but declined) probabilities.

For the traditionally used sequential probability ratio test

$$\lambda(u) = \log \frac{p_1(u)}{p_0(u)},\ u \in U.$$

Let the described above hypothetical model be distorted:

$$\bar{P}_k(x) = (1-\epsilon_k)P_k(x) + \epsilon_k \tilde{P}_k(x),\ x \in U,\ k \in \{0,1\}$$

be the factual data flow observations probability distribution, representing a mixture of the hypothetical probability distribution P_k and of the contaminating probability distribution $\tilde{P}_k$, where $\epsilon_k \in [0, \frac{1}{2})$ is the probability of contamination (contamination level).

To construct the robust sequential decision rule, a family of modified sequential tests $\delta_g = (\tau_g, d_g)$ is developed:

$$\tau_g = \inf\{n :\ \sum_{t=1}^{n} g(\lambda_W(x_t)) \notin (C_-, C_+)\},$$

$$d_g = 1_{[C_+,+\infty}(\sum_{t=1}^{n} g(\lambda_W(x_t))),$$

where

$$g(z) = g_- 1_{(-\infty,g_-)}(z) + z \cdot 1_{[g_-,g_+]}(z) + g_+ 1_{(g_+,+\infty)}(z),$$
$$z \in R.$$

Here $g_-, g_+ \in R$ are extra parameters of the developed sequential decision rules. Using this paprameters, the robustified sequential decision rules are constructed with minimax criterion w.r.t. the risk function.

III. INHOMOGENEOUS DATA FLOWS FORMING TIME SERIES WITH A TREND

Consider the model of stochastic data flow, where inhomogeneous independent observations forming a time series with a trend [8].

Let $x_1, x_2, \dots$ be observations of a time series with a trend:

$$x_t = \theta^T \psi(t) + \xi_t,\ t \geq 1,$$

where $\psi(t) = (\psi_1(t), \dots, \psi_l(t))$, $t \geq 1$, are the vectors of the trend basic functions, $\theta = (\theta_1, \dots, \theta_l)^T \in \mathbf{R}^l$ is a vector of coefficients, their values are not known along with the observation process, $\{\xi_t,\ t \geq 1\}$ is a sequence of independent identically distributed random variables from $\mathcal{N}_1(0, \sigma^2)$.

To give more flexibility in the decision making, the case of M simple hypotheses is considered w.r.t. the vector θ. The following two sequential test were analyzed.

A. M-ary sequential probability ratio test. It uses the posterior probabilities of the hypotheses. The stopping time N_a and the final decision d_a for this test are defined by the equations:

$$N_a = \inf\Big\{n \geq 1 :\ \exists m \in \{1, \dots, M\},$$
$$P\{\mathcal{H}_m \mid x_1, \dots, x_n\} > \frac{1}{1+A_m}\Big\},$$
$$d_a = \arg \max_{1 \leq m \leq M} P\{\mathcal{H}_m \mid x_1, \dots, x_{N_a}\},$$

where $A_m \in (0,1]$ are some specified constants, $m \in \{1, \dots, M\}$, $d_a = m$ means that the decision in favor of the hypothesis H_m is made.

B. Matrix sequential probability ratio test. Denote

$$\Lambda_n(i,j) = \ln \left(\prod_{t=1}^{n} \frac{n_1(x_t; (\theta_i)^T \psi(t), \sigma^2)}{n_1(x_t; (\theta_j)^T \psi(t), \sigma^2)} \right);$$

$$\tau_i = \inf\{n \in \mathbf{N} :\ \Lambda_n(i,j) > b_{ij},$$
$$\forall j \in \{1, \dots, M\} \setminus \{j\}\},\ i \in 1, \dots, M,$$

where $B = (b_{ij})$, $i,j \in \{1, \dots, M\}$, is the matrix of the test thresholds (using them, the error probabilities of the test are controlled by the user f the decision rule). For this test the stopping time N_b and the final decision d_b are defined as follows:

$$N_b = \min\{\tau_i :\ i \in \{1, \dots, M\}\},\ d_b = \arg \min_{i \in \{1,\dots,M\}} \tau_i.$$

For the two sequential tests defined above, the termination with probability 1 property and the finiteness of all moments of the random stopping time are proved under a condition reasonable and affordable for practice. For the M-ary sequential probability ratio test, upper bounds for the error probabilities are derived.

A robustified version of the matrix sequential probability ratio test based on change limitation for the test statistics is constructed and its properties are analyzed via numerical experiments.

IV. DEPENDENT DATA FLOW FORMING A HOMOGENEOUS MARKOV CHAIN

Consider here the situation, where observations are dependent and forming a homogeneous Markov chain [4].

Let the data flow be dependent observations forming a homogeneous Markov chain $x_1, x_2, \dots$, with possible values in the set $V = \{0, 1, \dots, M-1\}$. Denote the vector of initial states probabilities by $\pi = (\pi_i)$, $i \in V$, and the one-step transition probabilities matrix by $P = (p_{ij})$, $i,j \in V$, that are: $\mathrm{P}\{x_1 = i\} = \pi_i$, $\mathrm{P}\{x_n = j \mid x_{n-1} = i\} = p_{ij}$, $i,j \in V$, $n > 1$.

There are two hypotheses concerning the Markov chain parameters introduced above: $\mathcal{H}_0$: $\pi = \pi^{(0)}$, $P = P^{(0)}$ with the alternative $\mathcal{H}_1$: $\pi = \pi^{(1)}$, $P = P^{(1)}$, where $\pi^{(0)}$, $\pi^{(1)}$ are the given values of the initial states probabilities vector,

$P^{(0)} \neq P^{(1)}$ are the one-step transition probabilities matrices for correspondent hypotheses. Denote also:

$$\lambda_1 = \ln \frac{P_1\{x_1\}}{P_0\{x_1\}},\ \lambda_k = \ln \frac{P_1\{x_k \mid x_{k-1}\}}{P_0\{x_k \mid x_{k-1}\}},\ k > 1,$$

$$\Lambda_n = \sum_{k=1}^{n} \lambda_k,\ n \in \mathbb{N},$$

where $P_s\{x_1\}$ is the probability to observe the value x_1, $P_s\{x_k \mid x_{k-1}\}$ is the probability to observe x_k at the moment k provided at the moment $k-1$ the value x_{k-1} was observed, if hypothesis $\mathcal{H}_s$, is true $s \in \{0, 1\}$.

As it was done above, construct the sequential decision rule to decide in favor of $\mathcal{H}_0$ or $\mathcal{H}_1$. According to this decision rule, with given thresholds values $C_-, C_+ \in \mathbb{R}$, $C_- < 0$, $C_+ > 0$, hypothesis $\mathcal{H}_0$ is accepted on the basis of n observations, if $\Lambda_n \leq C_-$. Hypothesis $\mathcal{H}_1$ is accepted, if $\Lambda_n \geq C_+$, otherwise the observation process is not stopped, and $(n+1)$-th observation is requested.

Correspondent families of modified sequential decision rules are developed. Within the developed families, the robustified sequential decision rules are constructed with the minimax risk criterion [10].

V. CONCLUSION

The approach is applied to analysis of COVID-19 incidence dynamics process in the Republic of Belarus to identify types of trajectories: growth, horizontal fluctuation, decrease [11]. Also cases of composite hypotheses can be treated with the discussed approach [9].

REFERENCES

[1] N. Mukhopadhyay, and B. de Silva. Sequential Methods and Their Applications. New York, Marcel Dekker, 2009, 504 p.

[2] A. Wald. Sequential Analysis. New York, John Wiley and Sons, 1947, 212 p.

[3] T. Lai. “Sequential analysis: Some classical problems and new challenges,” StatisticaSinica, 2001, vol. 11, pp 303–408.

[4] A. Y. Kharin. Robustness of Bayesian and Sequential Statistical Decision Rules. Minsk, BSU, 2013, 207 p. (In Russ.).

[5] P. Huber, and E. Ronchetti. Robust Statistics. New York, Wiley, 2009, 380 p.

[6] A.Y. Kharin, and D.V. Kishylau. “Robust sequential test for hypotheses about discrete distributions in the presence of ”outliers”,” Journal of Mathematical Sciences, 2015, vol. 205(1), pp 68–73.

[7] A. Kharin, T.T. Tu. “Performance and robustness analysis of sequential hypotheses testing for time series with trend,” Austrian Journal of Statistics, 2017, vol. 46(3-4), pp 23–36.

[8] A.Y. Kharin, T.T. Tu. “On error probabilities calculation for the truncated sequential probability ratio test,” Journal of the Belarusian State University. Mathematics and Informatics, 2018, no. 1, pp 68–76.

[9] A.Y. Kharin. An approach to asymptotic robustness analysis of sequential tests for composite parametric hypotheses. Journal of Mathematical Sciences, 2017, vol. 227(2), pp 196–203.

[10] T.T. Tu, and A.Y. Kharin. Sequential probability ratio test for many simple hypotheses on parameters of time series with trend. Journal of the Belarusian State University. Mathematics and Informatics, 2019, no. 1, p 35–45.

[11] Y. S. Kharin, V. I. Malugin, V. A. Voloshko, O. V. Dzernakova, and A.Y. Kharin. Statistical forecasting of the dynaics of epidemiological indicators for COVID-19 incidence in the Republic of Belarus. Journal of the Belarusian State University. Mathematics and Informatics, 2020, no. 3, pp 36–50.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Semantic Logging of Repeating Events in a Forward Branching Time Model

Valerian Ivashenko
Intellectual Information
Technologies Department
Belarusian State University of
Informatics and Radioelectronics
Minsk, Belarus
ivashenko@bsuir.by

Nikita Zotov
Intellectual Information
Technologies Department
Belarusian State University of
Informatics and Radioelectronics
Minsk, Belarus
nikita.zotov.belarus@gmail.com

Maksim Orlov
Intellectual Information
Technologies Department
Belarusian State University of
Informatics and Radioelectronics
Minsk, Belarus
orlovmassimo@gmail.com

Abstract. **The tasks of knowledge logging in the form of semantic networks of the model of the unified semantic knowledge representation are considered. The formal model of a semantic log of repeating events in knowledge processing and algorithms for adding and retrieving logged events from the log are presented. The spatial-time structure of logged processes should satisfy a forward branching time model.**

Keywords: **semantic logging, repeated events, semantic networks, branching time, full persistence, unified semantic knowledge representation model, multi-agent system, cognitive architecture**

I. INTRODUCTION

Semantic logging (SL) is a mechanism designed to provide the intellectual system with introspective capabilities in order to endow it with the qualities of an artificial consciousness, including the ability to explain one's own work and its results [1]. It should take into account the non-factors of knowledge [3] including its incompleteness, uncertainty, hypotheticality as well as requirements for working in real time [10]. SL consists in recording in the knowledge representation language, for example, in the form of a semantic network, a knowledge about an order (system) of actions and events (phenomena) occurred in processes of knowledge processing.

II. OVERVIEW

SL can be considered as one of the approaches to a process introspection [6]. SL approach was descrbied in several previous works [4, 5, 9–11, 16]. From the point of view of the becoming structure or structure of time, there are several types of process time models: linear time, tree or branching time, directed acyclic time structure, arbitrary time structure [2, 9–11, 13]. From the point of view of the data structure processing there are also several types of persistent data structures (DS): partially persistent DS, fully persistent DS, confluently persistent DS [7, 9]. One of the approaches to achieve DSs persistence (structures when some changes are made to them retain all their previous states and access to those states) is using CoW data structures [8].

There are also several points of view for knowledge processing (KP) in multi-agent an AI systems: concurrent and distributed KP [6] without a certain global state or common knowledge base such as in actor model; coordinated and partially synchronized KP via such means as blackboard systems [15], fully synchronized KP using global states as in state space search models and algorithms (using various logics) [2].

Some algorithms and models have been proposed for acyclic process structures [9–11]. All considered models do not takeevents repetitions into account.

Their basic algorithms solve mostly SL generation and information retrieving tasks.

A. Basic procedures

1) Generation: The following log (pre-existing and possibly empty) generation algorithm (Fig. 1) pushes pre-created event identifiers with links providing full persistence of the structure of the semantic log as a set of its integrated versions.

```
$\langle LE, log, e, n\rangle \leftarrow$
  if ($undefined(log)$)
    $\langle log, g\rangle \leftarrow createLeafLink(\langle LE, log, e, n, 2\rangle)$
  else
    $l \leftarrow getFirstLink(log)$
    $c \leftarrow getCoefficient(\langle log, l\rangle)$
    $\langle log, g\rangle \leftarrow createLeafLink(\langle LE, log, e, n, 2/c\rangle)$
    while ($hasNextLink(\langle log, l\rangle) \wedge (c = 1)$)
      $\langle f, p, c\rangle \leftarrow \langle l, getNextLink(\langle log, l\rangle), 1\rangle$
      $l \leftarrow p$
      if ($hasNextLink(\langle log, l\rangle)$)
        $l \leftarrow getNextLink(\langle log, l\rangle)$
        $c \leftarrow getCoefficient(\langle log, l\rangle)$
      $g \leftarrow createPair(\langle log, f, p, 2/c, g\rangle)$
    $g \leftarrow appendLink(\langle log, l, g\rangle)$
$\leftarrow log$
```

Fig. 1. Log link generation algorithm

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

2) Retrieving information: There are two basic searchalgorithms in such SL structures (Fig. 2, 3).

```
⟨log, i⟩ ←
⟨l, c, s⟩ ← ⟨getFirstLink (log) , 1, 1⟩
j ← getIdentifier (⟨log, l⟩)
while (hasNextLink (⟨log, l⟩) ∧ less (⟨i, j⟩))
  n ← getNextLink (⟨log, l⟩)
  ⟨d, k⟩ ← ⟨n, 1⟩
  while (k < c) ⟨k, d⟩ ← ⟨k + k, getDownLink (⟨log, d⟩)⟩
  j ← getIdentifier (⟨log, d⟩)
  if (¬less (⟨j, i⟩))
    s ← s + c
    c ← c ∗ getCoefficient (⟨log, l⟩)
    l ← n
while (1 < c)
  l ← getDownLink (⟨log, l⟩)
  d ← l
  if (hasNextLink (⟨log, l⟩))
    n ← getNextLink (⟨log, l⟩)
    d ← n
  ⟨c, k⟩ ← ⟨c/2, 1⟩
  while (k < c) ⟨k, d⟩ ← ⟨k + k, getDownLink (⟨log, d⟩)⟩
  j ← getIdentifier (⟨log, d⟩)
  if (¬less (⟨j, i⟩)) ⟨s, l⟩ ← ⟨s + c, n⟩
if (less (⟨j, i⟩) ∨less (⟨i, j⟩)) ← s
← ⟨s, l⟩
```

Fig. 2. Get event by identifier algorithm

```
⟨log, n⟩ ←
⟨l, c, s⟩ ← ⟨getFirstLink (log) , getCoefficient (⟨log, l⟩) , 1⟩
while (hasNextLink (⟨log, l⟩) ∧¬(n < s + c))
  ⟨s, l⟩ ← ⟨s + c, getNextLink (⟨log, l⟩)⟩
  c ← c ∗getCoefficient (⟨log, l⟩)
c ← c/getCoefficient (⟨log, l⟩)
while (1 < c)
  ⟨l, c⟩ ← ⟨getDownLink (⟨log, l⟩) , c/2⟩
  if (hasNextLink (⟨log, l⟩) ∧¬(n < s + c))
  | ⟨s, l⟩ ← ⟨s + c, getNextLink (⟨log, l⟩)⟩
if (s = n) ← getEvent (l)
← nothing
```

Fig. 3. Get event by index algorithm

B. Applied tasks and questions

These algorithms are the platform to solve more complicated problems and applied tasks. There are several levels of such tasks and problems: level of relations between events and their repetitions [9]; level of spatial-time relations between phenomenons [13]; level of applied tasks of analysis and synthesis of external and internal phenomenons. The problems of the first level include problems of determining: event anteriority, primary event (Fig. 4) and last common event (LCE). For mentioned models, there are four cases of anterioity of two events: event coincidence, event alternativeness (synchronicity), the first event antecedence, the second event antecedence [2]. Tasks of the third level are the memorizing and supervising of internal and external processes, process mining [6] including history analysis and navigation, process reproduction (inductive programming). These tasks can arise in such areas as education, software development and interaction within control version systems, (digital) music composition and others [4, 6, 12, 15]. As for history navigation as a general applied tasks, let's consider the following approach. Let's consider an agent named "locator". This agent has five properties: two initial events (or its repetitions) – minor and major, two margin events (or event repetitions) – minor and major, chosen event (event repetition). Also this agent has behavior implemented by several procedures: initializing initial, margin and chosen events; transferring a chosen event to a user (another agent); updating minor or major margin event by the chosen event and choosing new event; updating minor margin event either by the major margin event or by the minor initial event and choosing new event; updating major margin event either by the minor margin event or by the majorinitial event and choosing new event.

```
log ←
⟨l, c, s⟩ ← ⟨getFirstLink (log) , 1, 1⟩
while (hasNextLink (⟨log, l⟩))
  c ← c ∗ getCoefficient (⟨log, l⟩)
  ⟨s, l⟩ ← ⟨s + c, getNextLink (⟨log, l⟩)⟩
while (1 < c)
  ⟨l, c⟩ ← ⟨getDownLink (⟨log, l⟩) , c/2⟩
  if (hasNextLink (⟨log, l⟩))
    ⟨s, l⟩ ← ⟨s + c, getNextLink (⟨log, l⟩)⟩
← s
```

Fig. 4. Primary event log index determining algorithm

III. PROPOSITION

A. Repetition logging

Previous models did not take into account the repetitions of events. Event (periodic) repetitions can be interpreted as event occurrences in (locally) cyclic spatial-temporal model. It is important to admit that all further events under consideration will have no more than one immediate predecessor. In the case of repeating events, the previous models are inappropriate. Thus, we need to identify event repetitions but not events. Therefore, we have many repetition identifiers for one event. ll of them should be enumerated and distinguished by an event repetition number. Another problem is that we might want to know, "Is there any event occurrence in the log or not?" without any interest in a particular event repetition. To solve the last problem (occurrence problem), global (event) occurrence identifiers (GOI) can be used. For an event, its first event repetition identifier may be used as a GOI. Thus, a global occurrence number is an event repetition number which equals 1 for each event. If we have a log structure with partial persistence only then these are all essential differences.

The Fig. 5, 6 show the basic algorithms for KP under a log structure where event repetitions are taken into account. Algorithm for getting a last event repetition number can be obtained replacing *getCoefficient*(⟨*log*, *l*⟩) calls at Fig. 4 by 2.

```
⟨log, r⟩ ←
⟨l, c, s⟩ ← ⟨getFirstLink (log) , 1, 1⟩
while (hasNextLink (⟨log, l⟩))
| c ← c + c
| ⟨s, l⟩ ← ⟨s + c, getNextLink (⟨log, l⟩)⟩
 ⟨t, k⟩ ← ⟨l, c⟩
while (1 < c)
| ⟨l, c⟩ ← ⟨getDownLink (⟨log, l⟩) , c/2⟩
| if (hasNextLink (⟨log, l⟩))
|  | ⟨s, l⟩ ← ⟨s + c, getNextLink (⟨log, l⟩)⟩
|       else ⟨t, k⟩ ← ⟨l, c⟩
|
l ← newNextLink (⟨log, t⟩)
while (1 < k) ⟨k, l⟩ ← ⟨k/2, newDownLink (⟨log, l⟩)⟩
← appendRepetitionLog (⟨log, l, r⟩)
```

Fig. 5. Algorithm of the queuing of an event repetition

```
⟨log, n⟩ ←
⟨l, c, s⟩ ← ⟨getFirstLink (log) , 1, 1⟩
while (hasNextLink (⟨log, l⟩) ∧¬(n < s + c + c))
|  c ← c + c
|  ⟨s, l⟩ ← ⟨s + c, getNextLink (⟨log, l⟩)⟩
while (1 < c)
|  ⟨l, c⟩ ← ⟨getDownLink (⟨log, l⟩), c/2⟩
|  if (hasNextLink (⟨log, l⟩))
|  | ⟨s, l⟩ ← ⟨s + c, getNextLink (⟨log, l⟩)⟩
if (s = n) ← getOccurenceStructure (l)
← nothing
```

Fig. 6. Algorithm of the getting of occurrence structures by number of anevent repetition (getOccurrences)

B. Full persistence problems

However, if we work with full persistence SL structures then there will be a set of other problems. There will be: enumerating all repetitions of a certain event in the log (for partial persistence it can be efficiently solved with log-specific event repetition identifiers and additionally segregated enumeration queues), an occurrence problem that requires more complex solution than for a partial persistence SL structure.

C. Model

SL model of repeating events having no more than one immediate predecessor is defined by (1) and satisfies (2).

$$\langle e, d, l, i, g, o, r, h, n, p, P, R\rangle \in E_{+}{}^{I} \times S_{+}{}^{E\times \mathrm{N}} \times \times L^{S} \times I^{S} \times P^{E} \times \{\bot, \top\}^{S\times P} \times R_{+}{}^{S\times P} \times \times I_{+}{}^{L\times \mathrm{N}} \times \mathrm{N}_{+}{}^{L\times \mathrm{N}} \times \mathrm{N}_{+}{}^{L\times I} \times 2^{I} \times 2^{L} \qquad (1)$$

where E – events, I – identifiers, S – descriptors structures with occurrence log sets (implemented as CoW critbit trees), L – logs, P – primary (occurrence) identifiers, R – repetition sublogs, h – event identifier by index mapping, n – repetition number by index mapping, p – index by identifier mapping.

$$g = (d \cap ((E \times \{1\}) \times S)) \circ i. \qquad (2)$$

IV. COMPUTATION COMPLEXITY ANALYSIS

It can be shown also by experimental confirmations that generation algorithms have $O\ (log\ (n)\ f\ (n))$ time complexity. Whereas time complexities of basic search algorithms do not exceed $O\ (log^{2}\ (n) * f\ (n))$, where $f\,(n)$ – time of the access to an element of logs.

V. EFFICIENT SOLVABLE PROBLEMS AND APPLICATION

Anteriority and LCE problems for repeating events that satisfy given restrictions can be efficiently solved with followingalgorithms (Fig. 7, 8 and 9).

```
⟨e, n⟩ ←
log ← getRepeatedEventLog (⟨e, n⟩)
if (defined (log))
| ← getIdentifier (getFirstLink (log))
← emptyIdentifier
```

Fig. 7. Algorithm for getting an event repetition identifier (getRepeatedEventIdentifier)

```
⟨e, n, a, m⟩ ←
l ← getRepeatedEventLog (⟨e, n⟩)
g ← getRepeatedEventLog (⟨a, m⟩)
i ← getRepeatedEventIdentifier (⟨e, n⟩)
j ← getRepeatedEventIdentifier (⟨a, m⟩)
if (coincide (⟨e, a⟩))
|← coincidence
if (interdependent (⟨e, a⟩))
|← interdependence
if (less (⟨i, j⟩))
| if (retrieveREByIdentifier (⟨g, i⟩))
| |← antecedenceFirst
  else
|
|  if (less (⟨j, i⟩))
|  | if (retrieveREByIdentifier (⟨l, j⟩))
|  | |← antecedenceSecond
|       else
|
|  |← coincidence
← potentialAlternativeness
```

Fig. 8. Anteriority determining algorithm

These algorithms are primarily oriented to navigation task applications. Anteriority problem

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

seems to be more complicated for cyclic process structures. That is particularly why, there are five cases of anterioity of two events: event coincidence, event alternativeness (synchronicity), the first event antecedence, the second event antecedence, interdependency (antecedence for both of events) (Fig. 8) [9–11, 13]. To solve anteriority problem for the last case, authors suggest to combine an incremental structures for merge-find set approach [14] and the proposed SL models.

```
$\langle e, n, a, m\rangle \leftarrow$
$x \leftarrow$ getPrimaryRepeatedEventIndex ($\langle e, n\rangle$)
$y \leftarrow$ getPrimaryRepeatedEventIndex ($\langle a, m\rangle$)
$l \leftarrow$ getRepeatedEventLog ($\langle e, n\rangle$)
$i \leftarrow$ retrieveREByIndex ($\langle l, x\rangle$)
$i \leftarrow$ getRepeatedEventIdentifier ($i$)
$g \leftarrow$ getRepeatedEventLog ($\langle a, m\rangle$)
$j \leftarrow$ retrieveREByIndex ($\langle g, x\rangle$)
$j \leftarrow$ getRepeatedEventIdentifier ($j$)
$k \leftarrow$ getRepeatedEventIdentifier ($\langle a, m\rangle$)
$h \leftarrow$ getRepeatedEventIdentifier ($\langle e, n\rangle$)
if (less ($\langle i, j\rangle$) $\vee$less ($\langle j, i\rangle$) $\vee$less ($\langle k, i\rangle$) $\vee$less ($\langle h, j\rangle$))
$| \leftarrow$ nothing
if (less ($\langle x, y\rangle$)) $\langle c, d, p\rangle \leftarrow \langle\langle e, n\rangle, \langle a, m\rangle, x\rangle$
    else $\langle c, d, p, l\rangle \leftarrow \langle\langle a, m\rangle, \langle e, n\rangle, y, g\rangle$

$|\langle q, r\rangle \leftarrow \langle 1, 1\rangle$
while ($r < p$)
    $k \leftarrow$ retrieveREByIndex ($\langle l, r\rangle$)
    if (anteriority ($\langle d, k\rangle$) = antecedentFirst) $p \leftarrow r$
        else

    $|q \leftarrow r$
    $r \leftarrow \lceil (p+q)/2 \rceil$
 $\leftarrow$ retrieveREByIndex ($\langle l, r\rangle$)
```

Fig. 9. Last common event repetition determining algorithm

VI. CONCLUSION

The proposed model and algorithms are able to provide prompt response to the main issues related to the study of the order of logged events and their repetitions. Basic algorithms were implemented using an integration platform which is the implementation platform for the reference-testing system [15]. The use of implemented algorithms is focused on supporting the automation of interaction with the knowledge base with reference materials and analyzing the history of responses by the user to test questions.

REFERENCES

[1] D. McDermott et al. Mind and Mechanism. Cambridge (Mass), MIT Press. (xv +), 2001, pp. 262.

[2] J. F. Allen, Time and time again: the many ways to represent time, Intern. J. of Intelligent Systems, 6, 1991, pp. 341–355.

[3] A. S. Narinyani, Non-factors: inaccuracy and underdetermination – difference and interrelation. Izv. RAN (RAS). Ser. Teoriya i sistemy upravleniya, 5, 2000, pp. 44–56 (in Russian).

[4] Sorin Ilie, Mihnea Scafes, Costin Badica, Thomas Neidhart, Rani Pinchuk. Semantic logging in a distributed multi-agent system. 2010.

[5] G. Pavlin, M. Kamermans, and M. Scafes. Dynamic process integration framework: Toward efficient information processing in complex distributed systems. In Proceedings of 3rd International Symposium on Intelligent Distributed Computing – IDC'2009, vol. 237 of Studies in Computational Intelligence, Springer, 2009, pp. 161–174.

[6] W. Gaaloul, K. Gaaloul, S. Bhiri, A. Haller, M. Hauswirth. Log-based transactional workflow mining. Distributed and Parallel Databases 25, 2009.

[7] J. R. Driscoll, D. D. Sleator, R. E. Tarjan, Fully Persistent Lists withCatenation. J. ACM 41(5), 1994, pp. 943–959.

[8] O. Rodeh. B-Trees, Shadowing, and Clones. ACM Transactions onStorage 3 (4), 2008, 1.

[9] V. P. Ivashenko, Semantic logging of knowledge processing. Proceedings of the workshop "Information Technologies and Systems 2017 (ITS 2017)", Minsk, Belarus, 25 October 2017, pp. 110–111 (in Russian).

[10] V. P. Ivashenko. Algorithms for semantic logging of knowledge processing, BIG DATA & Advanced Analytics, Minsk, Belarus, 3-4 May 2018, pp. 267-273. (in Russian).

[11] V. P. Ivashenko, Semantic logging of knowledge processing based on binary generated. PRIP, 2019, pp. 172–177.

[12] D. A. Pospelov, Situational management: theory and practice, Nauka, Moskva, 1986, p. 288. (in Russian).

[13] V. P. Ivashenko. Ontological model of space-time relations for events and phenomena in the processing of knowledge. Vestnik BrGTU 5 (107), 2017, pp. 13-17. (in Russian).

[14] Z. Galil, G. Italiano. Data structures and algorithms for disjoint set union problems. ACM Computing Surveys 23 (3), (1991), pp. 319—344.

[15] V. P. Ivashenko, Reference and testing system based on the unified semantic representation of knowledge, ITS 2020, 2020, pp. 80–81. (in Russian).

[16] C. Ringelstein, S. Staab. Logging in distributed workflows. Proceedings of the ISWC'07 Workshop on Privacy Enforcement and Accountability and Semantics (PEAS 2007), 320, 2008, pp. 19–30, 2007.

Development of Molecular Autoencoders as Generators of Protein Inhibitors: Application for Prediction of Potential Drugs Against Coronavirus SARS-CoV-2

Mikita Shuldau
United Institute of Informatics
Problems of NAS of Belarus
Minsk, Belarus
nickshuldov29@gmail.com

Artsemi Yushkevich
United Institute of Informatics
Problems of NAS of Belarus
Minsk, Belarus
artsemi.yushkevich@gmail.com

Ivan Bosko
United Institute of Informatics
Problems of NAS of Belarus
Minsk, Belarus
vanya_384@mail.ru

Alexander Tuzikov
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
tuzikov@newman.bas-net.by

Alexander Andrianov
Institute of Bioorganic Chemistry
of NAS of Belarus
Minsk, Belarus
alexande.andriano@yandex.ru

Abstract. **A generative autoencoder for the rational design of potential inhibitors of the SARS-CoV-2 main protease able to block the catalytic site of this functionally important viral enzyme was developed.**

Keywords: **SARS-CoV-2, X77, main protease, deep learning, generative autoencoder, semi-supervised learning, virtual screening, molecular docking, binding free energy calculations, anti-SARS-CoV-2 drugs**

I. INTRODUCTION

To date, computer-aided drug design has become an important tool allowing one to significantly reduce the time and costs required for developing novel therapeutic agents. In recent years, computer-assisted mathematical and statistical models, such as machine learning, are increasingly being used for drug design and discovery. Despite these methods becoming more common in chemoinformatics, their potential in this field is yet to be revealed.

Generative models have proven to be promising in tasks of text [1] and image [2] generation, including generation of medical images like X-ray ones. Despite the traditional similarity-based virtual screening of chemical databases, such as PubChem (https://pubchem.ncbi.nlm.nih.gov/) [3], provide wide possibilities for identification of novel potential drugs, it has certain disadvantages compared to generative statistical models. One of the major incentives to use generative models is a better exploration of a molecular feature space. Similarity-based search provides exploration of focused chemical space, limited by search space diversity of compounds at disposal, while generative statistical models allow one to cover molecular feature space of much wider chemical diversity. The second reason for generative model superiority is conditional sampling. Generation of new molecules from a chemical space is not the only option: predicted binding free energy could be used as an additional dimension, which allows one to generate molecules from a subset of investigated chemical space with a preset binding affinity.

This study is devoted to the development of the generative autoencoder based on a linear molecule representation in the Simplified Molecular Input Line Entry System (SMILES) format [4]. One of the core ideas behind using SMILES, or more precisely, vectorized SMILES for model training was the recovery capabilities of such data. As was shown in a study [5], generative models provide a decent ground for screening results enrichment, however use of descriptors like fingerprints may complicate the recovery of the chemical structures themselves. In contrast, SMILES-based embeddings are supposed to be a good alternative to using fingerprints in deep learning chemoinformatics approaches when the ability to restore the structure of chemical compounds is important. That is why the SMILES embeddings were chosen as the architectural basement for the constructed generative autoencoder to generate potential inhibitors of the selected protein target.

II. MATERIALS AND METHODS

A. Training Set Preparation

The developed generative autoencoder is built to be specific for a target protein, and, therefore, the training

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

dataset should include compounds potentially active against the selected protein. As noted before, the autoencoder developed here was adopted and applied for generation of potential anti-SARS-CoV-2 inhibitors, and, accordingly, a virtual compound library of potential anti-SARS-CoV-2 agents was formed for preparation of a training dataset. The preparation procedure of this molecular library was as follows:

a) Pharmacophore-based Virtual Screening: To identify small-molecule compounds potentially active against SARS-CoV-2 main protease (M^{pro}), the pharmacophore-based virtual screening was performed using the Pharmit server software (http://pharmit.csb.pitt.edu/) [6]. Seventeen pharmacophore models were built based on 6 peptidomimetics and 10 small-molecule inhibitors of SARS-CoV reported in a study [7], using web-server PharmaGist [8]. Virtual screening was performed in the nine Pharmit molecular libraries containing over 213.5 million chemical structures, resulting in a set of 711102 compounds that satisfied one of the seventeen constructed pharmacophore models. The Pubchem API wrapped in Python 3 (https://www.python.org/) module PubChemPy (https://pubchempy.readthedocs.io/) was used to additionally enrich the screened dataset with potential inhibitors based on the selected PubChem compounds by the similarity search with a Tanimoto similarity coefficient of 0.8.

b) Molecular Docking: Compounds identified by the pharmacophore-based virtual screening and PubChem similarity search were subject to the preliminary molecular docking with the unliganded SARS-CoV-2 M^{pro} structure. The compounds were then filtered based on the values of the docking scoring function with the threshold of –7 kcal/mol, which corresponds to the standard activity threshold of 10 µM commonly used in vitro screening. The dataset of 353467 potentially active compounds was subject to the refining molecular docking with the unliganded SARS-CoV-2 M^{pro} structure. Analysis of the distribution of scoring function values after the refining docking resulted in the filtration of successfully docked compounds above the selected threshold of –6 kcal/mol.

c) SMILES space revision and vectorization: Based on a linear SMILES notation, the dataset of selected compounds was cleared from those containing non-recognizable atoms, non-abundant isotopes, other than druglike (H, C, N, O, P, S, F, Cl, Br, I) atoms or those which molecular weight was above the selected threshold of 1000 Da. Structure representations of the prepared compounds in the linear notation SMILES were obtained by Python 3 using the RDKit (http://www.rdkit.org/) module which was also used previously for the described dataset cleaning. Based on the frequency distribution of SMILES elements in the prepared dataset, compounds possessing at least one SMILES element with frequency less than 0.001 were filtered out. Finally, distribution of SMILES lengths was investigated and compounds with SMILES representation longer than 120 characters were eliminated. After all the filters applied, the dataset consisted of 342102 distinct ligands and their corresponding SMILES. The SMILES were vectorized into a matrix according to the maximum length and symbols vocabulary size, with the added start and end symbols represented by "!" and "E".

The obtained 342102 compounds combined with the corresponding values of molecular docking scoring function formed the dataset which was split into the training, validation, and test sets comprising 70%, 15% and 15% of the original dataset, respectively. When forming the subsets, a stratified split was used to preserve equal energy distributions within all 3 sets. The validation set was used to evaluate the model's ability to reconstruct the input SMILES during training, while test SMILES were used to sample new compounds from by adding distortion to their latent representation. Thus, the corresponding datasets for model training, validation and generation of new molecules were prepared.

B. 3D Structures Generation for Generated Molecules

To evaluate the ability of deep learning model to generate novel compounds active towards the target protein, molecular docking of these molecules should be performed. In doing so, 3D structures of the generated molecules are required. To obtain these 3D structures from a linear notation SMILES, a script was developed in Python 3 using the RDKit module. The generation pipeline included the following steps: SMILES input, SMILES validity check, 2D coordinates generation, 3D coordinates generation, optimization of the structure in the MMFF94 force field, addition of hydrogen atoms, and re-optimization in the MMFF94 force field. Generation of 3D coordinates was performed using the ETKDGv3 [9] algorithm.

C. Molecular Docking

a) Preparation of Protein Structure: The crystal structure of the unliganded SARS-CoV-2 M^{pro} was taken from the Protein Data Bank (PDB ID: 6Y84; https://www.rcsb.org/pdb/). This SARS-CoV-2 M^{pro} structure was prepared by adding hydrogen atoms and annotating atoms with partial charges by Gasteiger scheme [10] followed by the structure optimization in the UFF force field [11] using the OpenBabel

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

software [12]. The structure of SARS-CoV-2 M^{pro} prepared in this way was used for the preliminary and refining molecular docking both during the dataset preparation as well as for molecular docking of the generated compounds.

b) Preparation of Ligand Structures: Prior to the preliminary docking, preparation of the ligand structures was the same as described for the SARS-CoV-2 M^{pro} structure. This procedure was performed using OpenBabel but included an additional step of rotatable bonds identification which is auto-made by this software. However, prior to the refining molecular docking, the ligand structures were prepared via the following two steps: i) optimization in the MMFF94 force field [13, 14] to remove steric clashes and addition of hydrogen atoms that are absent in the initial structure, both using the RDKit module in Python 3, ii) addition of partial charges by Gasteiger scheme and rotatable bonds identification using MGLTools (http://mgltools.scripps.edu/). It should be noted that before subjecting the generated compounds to molecular docking, 3D structures of novel ligands were obtained from the generated linear SMILES notations, as described above. Further preparation steps of the generated compounds for molecular docking included the addition of partial charges by Gasteiger scheme and rotatable bonds identification performed by MGLTools.

c) Molecular Docking Settings: The preliminary molecular docking was performed using the QuickVina 2 [15] program, and refining molecular docking was carried out by AutoDock Vina [16] software, both in the approximation of rigid receptor and flexible ligands. In both cases, the grid box included the catalytic site of SARS-CoV-2 M^{pro} with the following parameters: $\Delta X = 19$ Å, $\Delta Y = 21$ Å, $\Delta Z = 23$ Å centered at $X = -20$ Å, $Y = 19$ Å, $Z = -26$ Å. Thus, the grid box volume was $19\times 2 \times 23 = 9177$ Å^3. The value of the exhaustiveness parameter defining the number of individual sampling "runs" was set to 10 and 50 for preliminary and refining docking, respectively.

D. Deep learning

a) Models Architectures: Two deep learning models have been developed, namely an unsupervised SMILES-based Long Short-Term Memory (LSTM) [17] autoencoder (model I) and a semi-supervised SMILES-based LSTM autoencoder (model II) in which the value of binding free energy was used as an additional dimension in latent space to learn from the docked compounds and a value to sample around in the generation mode.

Model I (Fig. 1, I) takes vectorized SMILES as input, which follow through the LSTM layer. The peculiarity of this model is defined by the fact that LSTM output itself is not used, instead the hidden and cell states vectors are derived, which are concatenated together and then put through a dense layer. The output of this dense layer serves as a latent vector or SMILES embeddings in the context of the autoencoder. The embeddings are fed to two dense layers in parallel, creating initial hidden and cell state inputs for the LSTM layer in the decoder part. There is also the decoder input layer used as input for the decoder LSTM, which in the training mode receives

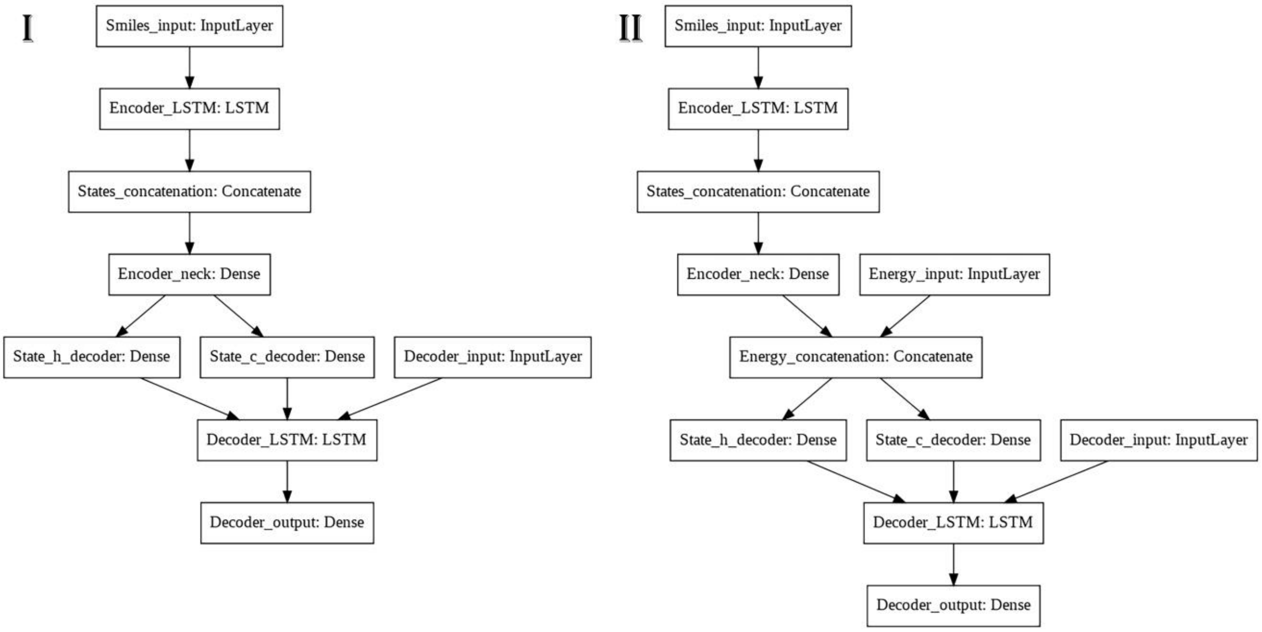


Fig. 1. Architectures of the autoencoder models: (I) Unsupervised (embeddings) model; (II) Semi-supervised (embeddings and energy) model

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

the same vectorized SMILES as the encoder input, and, as a conventional LSTM generative model, it predicts the next symbol. However, in the generation mode, decoder input starts the generation process with a start symbol only, embeddings are used to predict the initial states of decoder LSTM and they basically define which kind of SMILES will be generated.

The so-defined energy model (Fig. 1, II) differs from model I in the neuron responsible for the binding affinity value situated on the latent layer of the model. While model I allows one to generate molecules from random SMILES embeddings as well as adding noise to SMILES embeddings of ligands with predicted affinity, the energy model enables one to generate new ligands with a preset property of binding affinity, in addition to attempts to manipulate SMILES embeddings of the given ligands to try to improve their structures after decoding and thus increase the value of binding affinity.

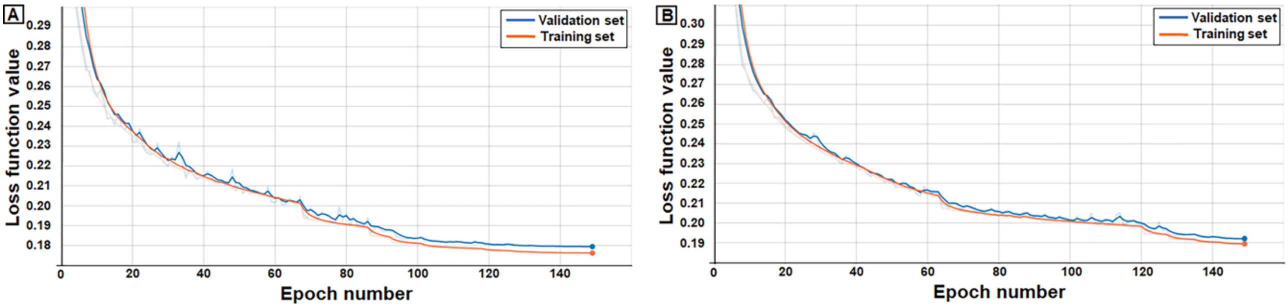


Fig. 2. Train and validation losses for autoencoder model: (A) Semi-supervised (embeddings & energy) model; (B) Unsupervised (embeddings) model

b) Models Training: Both models were composed layer by layer using TensorFlow 2.1 (https://www.tensorflow.org/) high-level API. The models were subject to 150 epochs of training, additionally “Reduce learning rate on a plateau” and “Early stopping” callbacks were used to help the model converge to a better local minima and also avoid overfitting. Stochastic gradient descent optimization method Adam [18] was used as an optimizer with 0.005 learning rate initial value and the categorical cross-entropy loss function was chosen. The loss score progress for both models is shown in Fig. 2.

E. Deep Learning-based Compounds Generation

Two methods of generation were subjects of our consideration: generation from random numbers drawn from normal distributions, where distributions parameters were derived using test data distribution on the latent layer for each vector component. For this method, the process of generation for model II implied setting an a priori value of binding free energy to approximate generated ligands with. Experiments for different thresholds were carried out. The major idea of the second method of generation was to sample best ligands from the test set, to try to add noise to their embeddings. This approach is supposed to change the reconstructed ligand, and, in the case of model II, also increase the predicted binding affinity, forcing the generation of more promising ligands. The combinations of two autoencoder models and two generation methods are summarized in Table I.

III. RESULTS AND DISCUSSION

As noted above, both models were tested using each of two generation methods. The results obtained were evaluated based on the values of binding affinity predicted by molecular docking, as well as by comparing these values with those calculated for the reference compounds used in the virtual experiments as a positive control.

The results of compound generation common for all of the conducted simulations are summarized in Table II.

A. Embeddings Model, Random Vectors Generator

Despite the fact, that this model does not use both reference compounds for generation and values of binding energy, it is capable of generating new potential inhibitors of the selected target only by generating compounds from the embeddings distributions inferred from the training data. In the set of generated compounds, the share of molecules with the predicted values of binding free energy less than –9 kcal/mol was 3.2%, which exceeds the same share in the training dataset (1.8%) by almost 2 times.

B. Embeddings Model, Test Set Compounds Used To Generate New Compounds From

This generation method utilizes compounds available and tries to generate new potential inhibitors from them. The share of generated compounds with high binding affinity is considerably larger, with 38%

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

of compounds exhibiting the predicted values of binding free energy less than –8 kcal/mol and 4% of compounds showing the predicted values less than –9 kcal/mol.

TABLE I. DESCRIPTION OF INVESTIGATED COMBINATIONS OF GENERATIVE LSTM AUTOENCODER MODELS AND GENERATION METHODS

Model	**Generation starting point description**	**Generation process description**
Unsupervised (Embeddings model)	Random number vectors drawn from fitted normal distributions	Random numbers are used as embeddings and fed to the decoder
Unsupervised (Embeddings model)	Compounds with free binding energy less than –9 kcal/mol, sampled from test set	Embeddings for these compounds are calculated, then distortion is added and updated embeddings are fed to the decoder
Semi-supervised (Energy model)	Random number vectors drawn from fitted normal distributions and a preset free binding energy value	Random vectors are used as embeddings and are passed as latent layer inputs along with a preset free binding energy value
Semi-supervised (Energy model)	Compounds with free binding energy less than –8 kcal/mol, sampled from test set and improved free binding energy values	Embeddings for these compounds are calculated, then distortion is added and updated embeddings along with improved free binding energy values are passed to the decoder

C. Energy Model, Generation from Random Numbers with a Set of Preset Thresholds of Binding Affinities

The semi-supervised model utilizes a version of “style and content” disentanglement for molecular data. According to the results obtained, 31% and 64% of generated compounds showed the values of binding energy within the deviations from the pre-defined energy value equal to 1 kcal/mol and 2 kcal/mol, respectively.

D. Energy Model, Test Set Compounds Used as Starting Points to Generate More Compounds, Energy Threshold is Shifted Towards Lower Energy by 0.5 and 1.0 kcal/mol Steps

This combination of the model and method proved to generate the top compounds throughout all four modes of generation. 52% of generated compounds are located within 1 kcal/mol deviation from the reference compounds, while 16% of generated compounds have the values of binding free energy lower than –9 kcal/mol.

IV. CONCLUSION

Two generative autoencoder models for prediction of novel drugs against SARS-CoV-2 were developed and applied to identify potential inhibitors able to block the catalytic site of the coronavirus main protease. The designed generative models combined with molecular docking proved their great potential to enrich screening pipelines with new compounds with desired properties. The generative power of the designed models is confirmed by the fact that out of 4805 successfully generated compounds only one compound was found in the original dataset. This indicates the richness of unexplored chemical space and proves an importance of development and application of generative models in drug design and discovery.

TABLE II. COMPOUNDS GENERATION RESULTS USING TWO GENERATIVE LSTM AUTOENCODER MODELS AND TWO GENERATION METHODS

Model name, generation method	**Number of generated compounds**	**Number of successfully docked compounds**	**Number of compounds with the predicted binding free energy less than –8 kcal/mol**	**Lowest predicted binding free energy, kcal/mol**	**Fraction of generated compounds with a lower binding affinity compared to reference compounds or energy threshold**
Embeddings, random vectors	1000	986	277	–10.6	–
Embeddings, reference compounds	2543	2518	967	–10.3	> 10.0 %
Energy, random vectors and energy value	600	594	161	–9.3	> 17.4 %
Energy, reference compounds and improved energy values	662	658	266	–10.4	> 12.2 %

ACKNOWLEDGMENTS

This study was financed by a grant of the Belarusian Republican Foundation for Fundamental Research (project F21COVID-002). The authors are grateful to EPAM Systems for support of the project and interest in the study.

REFERENCES

[1] T. Iqbal and S. Qureshi, “The survey: Text generation models in deep learning,” Journal of King Saud University - Computer and Information Sciences, April 2020.

[2] V. Sorin, Y. Barash, E. Konen, and E. Klang, “Creating Artificial Images for Radiology Applications Using Generative Adversarial Networks (GANs) – A Systematic Review,” Academic Radiology, vol. 27, no. 8, pp. 1175–1185, August 2020.

[3] S. Kim et al., “PubChem in 2021: new data content and improved web interfaces,” Nucleic Acids Research, vol. 49, no. D1, pp. D1388–D1395, January 2021.

[4] D. Weininger, “SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules,” J. Chem. Inf. Comput. Sci., vol. 28, no. 1, pp. 31–36, February 1988.

[5] A. M. Andrianov, G. I. Nikolaev, N. A. Shuldov, I. P. Bosko, A. I. Anischenko, and A. V. Tuzikov, “Application of deep learning and molecular modeling to identify small drug-like compounds as potential HIV-1 entry inhibitors,” Journal of Biomolecular Structure and Dynamics, pp. 1–19, 15 April 2021.

[6] J. Sunseri and D. R. Koes, “Pharmit: interactive exploration of chemical space,” Nucleic Acids Research, vol. 44, no. W1, pp. W442–W448, July 2016.

[7] T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, and S.-H. Jung, “An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy,” J. Med. Chem., vol. 59, no. 14, pp. 6595–6628, July 2016.

[8] D. Schneidman-Duhovny, O. Dror, Y. Inbar, R. Nussinov, and H. J. Wolfson, “Deterministic Pharmacophore Detection via Multiple Flexible Alignment of Drug-Like Molecules,” Journal of Computational Biology, vol. 15, no. 7, pp. 737–754, September 2008.

[9] S. Wang, J. Witek, G. A. Landrum, and S. Riniker, “Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences,” J. Chem. Inf. Model., vol. 60, no. 4, pp. 2044–2058, April 2020.

[10] J. Gasteiger and M. Marsili, “A new model for calculating atomic charges in molecules,” Tetrahedron Letters, vol. 19, no. 34, pp. 3181–3184, Januray 1978.

[11] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, “UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations,” J. Am. Chem. Soc., vol. 114, no. 25, pp. 10024–10035, December 1992.

[12] N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, “Open Babel: An open chemical toolbox,” Journal of Cheminformatics, vol. 3, no. 1, p. 33, October 2011.

[13] T. A. Halgren, “Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94,” Journal of Computational Chemistry, vol. 17, no. 5–6, pp. 490–519, 1996.

[14] P. Tosco, N. Stiefl, and G. Landrum, “Bringing the MMFF force field to the RDKit: implementation and validation,” Journal of Cheminformatics, vol. 6, no. 1, p. 37, July 2014.

[15] A. Alhossary, S. D. Handoko, Y. Mu, and C.-K. Kwoh, “Fast, accurate, and reliable molecular docking with QuickVina 2,” Bioinformatics, vol. 31, no. 13, pp. 2214–2216, July 2015.

[16] O. Trott and A. J. Olson, “AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading,” Journal of Computational Chemistry, vol. 31, no. 2, pp. 455–461, 2010.

[17] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, November 1997.

[18] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” arXiv:1412.6980 [cs], January 2017, Accessed: July 14, 2021. [Online]. Available: http://arxiv.org/abs/1412.6980.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Employing of RL Technology to Develop an Adaptive Motion Controller for a Line Follower Robot

Tatyana Kim
Open Laboratory of Artificial Intelligence
and Robotics
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
tatyana_kim92@mail.ru

Ryhor Prakapovich
Open Laboratory of Artificial Intelligence
and Robotics
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
rprakapovich@robotics.by

Abstract. **The article is focused on the development process of an adaptive motion controller for a line follower robot. The controller learning process took place on the basis of the digital twin of the mobile robot using reinforcement learning technology. The digital twin and the reinforcement learning algorithm were implemented in MATLAB/Simulink. The Twin–Delayed Deep Deterministic Policy Gradient Agents method was used as a learning algorithm. The reward function was taken to minimize the distance between the center of the robot and the middle of the nearest section of the color–contrast line, as well as the difference between the angle of the robot position and the tangent to the current section of the line.**

Keywords: **Reinforcement Learning (RL), MATLAB/ Simulink, Twin-Delayed Deep Deterministic Policy Gradient Agents (TD3), control system, digital twin**

I. INTRODUCTION

Reinforcement learning (RL) is one of the machine learning methods for solving control problems in complex technical systems that cannot or can be problematically described in an analytical form.

The RL method is based on the implementation of the process of maximizing a certain reward (reward) signal when enumerating various behaviors of the studied systems – Agents. The Agent learns to perform those actions that can bring him the greatest reward. In the most interesting and important cases, the Agent's actions can affect not only the local reward received immediately, but also the situation as a whole [1]. Forming a long–term reward is a rather difficult process, since a correctly formed reward will bring the best result and shorten the training time (the better the learning process is).

At present, in addition to classical robotic manipulators, mobile robots (MR) in the form of robotic carts are in high demand in production. As a rule, at the lower level of MR control, PID controllers are most often used [2, 3]. The PID controller allows to adjust the control action of the actuators in such a way as to achieve the required values of the objective function as quickly as possible. Sometimes the selection of coefficients is a rather long process, which does not always lead to success, since there is a chance of overshoot [4, 5]. In closed–loop control systems (CS), the controller uses status observations to improve performance and correct random noise and errors. Engineers use this feedback, as well as the object of control (OC) and the Environment, to design the controller according to the requirements of the system. This concept is easy to put into words, but it will be difficult to implement, since the model can be highly nonlinear or have large spaces of states and actions. There is also a problem related to the fact that for each line type it is required to find the corresponding PID values. This problem can be quickly solved by using RL.

The object of the research is the RoboCake training MR with a differential drive, which is supposed to move along the color contrast line. Its CS is a classic servo drive, including a color contrast line midpoint sensor and a PID controller that controls the angular speed of the wheels. Previously, experiments were carried out on the automatic tuning of the specified PID controller using the software module MATLAB "Interactively Estimate Plant Parameters from Response Data" and the application of genetic algorithms (GA) [6].

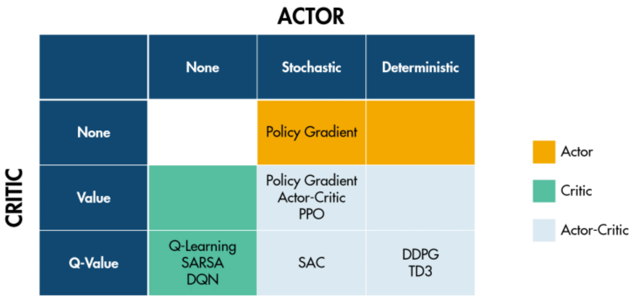
The aim of this work is to implement automatic learning of a PID controller using the RL method and compare its effectiveness with other machine learning methods.

II. SUMMARY OF THE PROBLEM

In RL, the object of research is studied, as a result of which an artificial neural network (ANN) is generated that can simulate the desired object. The most important issue in RL remains the generation of the training sample for the specified ANN. There are

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

several types of Actor–Critic learning algorithms (Fig. 1). Actor–Critical Agents use either a stochastic Actor or a deterministic Actor with a value Critic or a Q–value Critic [7].



ACTOR

CRITIC	None	Stochastic	Deterministic
None		Policy Gradient	
Value		Policy Gradient Actor-Critic PPO	
Q-Value	Q-Learning SARSA DQN	SAC	DDPG TD3

🟧 Actor
🟩 Critic
⬜ Actor-Critic

Fig. 1. Learning algorithms for Agent–Actor–Critic combined with stochastic–deterministic Actor and value–Q–value Critic [7]

As the investigated model, the work uses the DT [8] MR RoboCake, but developed in the MATLAB / Simulink packages. Following the RL methodology, in order to achieve this goal, it is required to describe the software Agent that needs to develop a policy for managing the educational institution. The specified MR is required to move along an elliptical curve (color–contrast line) with the maximum possible speed and the minimum deviation from its center. The term "policy" means a mapping that selects the appropriate actions of the OS on the corresponding changes in the Environment [9]. In the process of training, the Agent uses the following data: the readings of the sensor of the middle of the line (consisting of 3 light sensors), the distance from the center of the MR to the center of the color–contrast line, as well as the angle between the normal of the nearest section of the line and the direction of movement of the MR itself. The result of the observation is the selection of the angular speeds of rotation of the 2 MR wheels. In order for an Agent to form actions correctly, he needs to train repeatedly and for the training to be fruitful, for each action he must be encouraged (rewarded) or fined. Also, stopping criteria are used to reduce the learning time. Each episode of the learning process can stop if: 1) the simulation exceeds the time allotted for movement along a full ellipse; 2) the robot has exceeded the distance to the center of the ellipse line. During each episode, the Agent chooses an action (forms a policy), after the end, he updates his parameters based on the actions and receives the maximum reward. This process continues until the Agent learns to move correctly along the line with the specified conditions [9]. When developing this algorithm, the Reinforcement Learning Toolbox library was used, which provides a policy and a reward function using deep neural networks (DNNs) [10].

III. LEARNING PROCESS

Before starting the learning process, it is required to create a virtual Environment for the functioning of the MR and an interface for interacting with it. Next, you should configure the Agent module from the "Reinforcement Learning Toolbox" library [10]. To implement the Agent, the TD3 algorithm was chosen, which is characterized by the fact that in addition to the Agent, which offers specific actions of the Actor on certain indications of the sensory system, two new entities are also used – two Critics that form a long–term reward. Next, the Agent is configured and formed. The final stage is training and verification of the results of the trained Agent.

After a positive completion of the learning process, using the *getLearnableParameters()* command, weights are extracted from the trained ANN, which are the desired coefficients of the PID controller.

A. Reward function

The reward process is an important step in RL, as it affects the performance of the Agent in relation to the goal that CS MR seeks to achieve. A correctly selected reward signal forces the Agent to move in the right direction with a minimum deviation from the line and maximize the total reward received by the Agent over a long period of time, stimulating him (Agent) for long–term rewards. The reward is formed in the form of a scalar signal that is received by the Agent and generated by the Environment.

"Two criteria are taken into account as a reward for a committed action. First, the distance (*r*) from the center of the robot (*x*,*y*) to the nearest unvisited point (x_0 y_0) of the ellipse located on the line is calculated using the formula for calculating the distance between two points (1). The closer the Agent, the more reward he will receive, the further, the less reward" [11]. Secondly, calculate the angle (φ) between the tangent to the ellipse at the point (x_0 y_0), where the robot should be, and the robot guide using formula (2). The smaller the angle (φ), the more the Agent will receive a reward (Fig. 2).

$$r = \sqrt{\left(x - \mathrm{x}_0\right)^2 + \left(y - \mathrm{y}_0\right)^2} \tag{1}$$

$$\varphi = \arctan\left(\left|\frac{-b^2 * sinsin\,(\alpha) * R}{a^2 coscos\,(\alpha) * R}\right|\right) \tag{2}$$

where b is the semi–minor axis, a is the semi–major axis, α is the angle between the radius (R) and the semi–major axis (*a*), R is the radius of the ellipse [12].

Since our reward signal consists of 2 signals combined into a scalar, then we determine that the distance between the line and the robot is of paramount importance, and the resulting angle (φ) we reduce the influence to get the reward.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

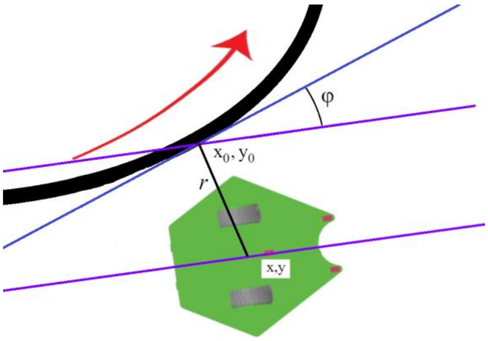


Fig. 2. The direction of movement of the robot, r is the minimum distance to the nearest point on the elliptic curve

B. *Implementation of the Environment*

The Environment is a DT MR implemented in the MATLAB application package [9], which generates the following states for the Agent:

– the position of the robot;

– indication from the sensor.

C. *Agent implementation*

The Agent is an CS that studies the CO and receives the following states from the Environment:

– observation, which includes:

– the previous formed action;

– the difference between the current state of the sensor and the desired value;

– the previous generated award;

– the current reward for the performed action;

– stopping criteria.

The Agent interacting with the Environment "implements the following basic functions necessary for the work of RL:

– list of available actions;

– handling status and awards from the Environment;

– obtaining the accumulated experience, presented in the form of a neural network (NN) for the Actor and the Critic" [11].

The list of available actions includes the angular speed from the left and right wheels, where the constant speed is 12 rad/s with a wheel radius of 0.025 cm.

The Agent receives readings from 3 sensors, such as the encoders of the right and left wheels and 1 light contact sensor, which determine the speed of the wheels and the location of the robot, as well as the previous action that the Agent formed.

The Agent receives the current reward when it minimizes the distance between the nearest point on the ellipse and the robot's location, while also minimizing the slope between the tangent to the ellipse and the robot's rail. The reward function for the RL Agent was defined as negative, since the RL Agent maximizes this reward, thereby minimizing the error.

The conditions are the criteria for stopping, formed in such a way as to shorten the training time. If the robot has exceeded the specified value (20 cm), then training starts over.

The learning process was built on the basis of TD3. This TD3 algorithm is the next version of the DDPG (Deep Deterministic Policy Gradient) algorithm, which is more reliable, increases the stability [13] of learning, and "Eliminates function approximation errors in methods of criticizing actors" [14]. The exceptional nature of this algorithm is that it combines 3 main algorithms for RL, such as Double Deep Q–Learning [15], Policy Gradient [16] and Actor–Critic [17].

This algorithm is based on an Agent–Critic and an Agent–Actor, which use a Critic with a Q–value and a deterministic Actor, respectively.

"The advantage of this method is that the TD3 Agent approximates the long–term reward, taking into account the observation and action, using the 2 presented Critics. When constructing neural networks for Actor, radial descent optimization can lead to negative weights. To avoid negative weights, we replaced the normal *fullConnectedLayer()* with *fullConnectedPILayer()*. This layer ensures that the weights are positive" [18]. After training, we extract these weights for our PID controller and apply them to the test model.

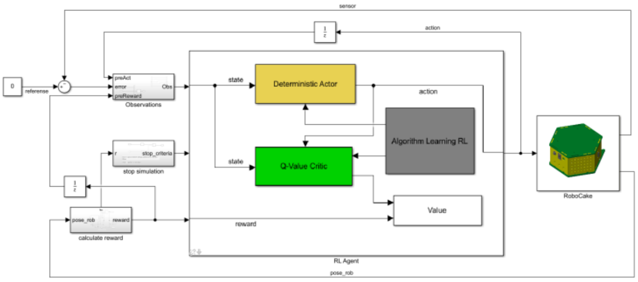


Fig. 3. Architecture of Reinforcement Learning based on DT in the MATLAB modeling Environment

D. *Learning Algorithm*

At this stage, we are ready to train the Robot, this is a rather long procedure, and in this case the training process took 2.15 hours and as a result, everything that the CS setup specialist had to do to set the correct architecture and calculate the CS parameters, all this did RL.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

IV. THE RESULTS OBTAINED, THEIR SIGNIFICANCE AND COMPARISON WITH PREVIOUS WORK

In the current research work, a software Agent was developed using the TD3 algorithm for a simulated DT of a 2–wheeled robot moving along an elliptical curve in the MATLAB application package. A simulation Environment was implemented and a reward function was developed. Comparing the results obtained earlier by the GA method [6], with the new results obtained by the RL method, we can conclude that the latter method reduces the number of episodes and training time by several times (Table I).

TABLE I. LEARNING OUTCOMES OF TD3 AND GA METHOD

PID controller	***P***	I	D	***Learning Algorithm***	***Training time (hours)***	***Number of episodes***
1	1.3593	1.4066		TD3	0.45	100
2	3.676	-0.159	-4.665		2.15	500
3	0.0903	0.0085	0.7691		2.90	800
4	-4.627	0.15	0.3011	GA	5.55	2000
5	-1	0	-0.0621		4.15	1500
6	-0.4164	-0.5453	-0.1568		6.94	2500

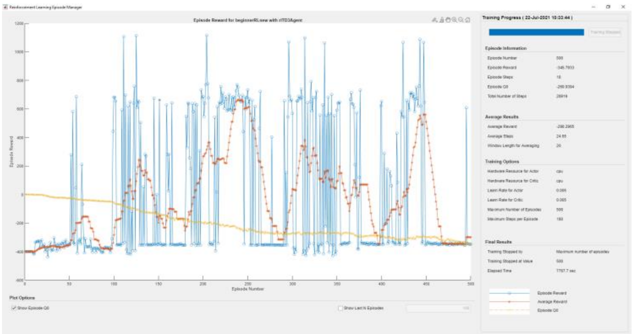


Fig. 4. Result of robot training after 2.15 hours, Episode Reward – the reward received by the robot for each episode, Average Reward – the average reward for window length for Averaging equal to 20, Episode Q0 – a critical assessment of the long–term reward for each episode

This graph shows that the highest reward is 1113, the average is 715, where the robot can drive an elliptical curve in 8 seconds minimizing the distance to the curve.

V. CONCLUSION

In this article, using the RL methods, we have chosen the optimal values of the PID controller that controls the movement of the 2–wheel MR along the color contrast line. For this, a virtual environment for the functioning of the MR was implemented, a software Agent was configured, a reward function was developed, training was implemented using the TD3 method using a deterministic Actor and a Q–value Critic. As a result, the tuned PID controller allows the MR to accurately move along the curved color–contrast line at a speed of 0.66 m /s.

REFERENCES

[1] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, London, England, The MIT Press, 2014, pp. 13.

[2] A. V. Philippov, M. A. Kosolapov, I. A. Maslov, and G. I. Tarasova, “Automated tuning of the PID controller for the control object of the tracking system using the MATLAB Simulink software package”, Science, technology and education, no. 12 (18), 2015, pp. 53–59 (in Russian).

[3] W. S. Levine, “PID Control,” in The Control Handbook, Ed. Piscataway, NJ, IEEE Press, 1996, pp. 198–209.

[4] F. G. Martins, Tuning PID controllers using the ITAE criterion, January 2005 International Journal of Engineering Education 21(5). Intern. J. Engng Ed. Vol. 21, no. 5, pp. 867–873, 2005 0949–149X/91 Printed in Great Britain.

[5] J. G. Ziegler, N. B. Nichols, “Optimum settings for automatic controllers,” Trans. ASME, 1942, vol. 64, pp. 759–768,

[6] T. Yu. Kim and G. A. Prokopovich, “Optimization of the parameters of the PID controller of the educational mobile robot control system using genetic algorithms”, unpublished (in Russian).

[7] Reinforcement Learning Agents (Release 2021a). [Online]. Available:https://www.mathworks.com/help/reinforcement-learning/ug/create-agents-for-reinforcement-learning.html.

[8] T. Yu. Kim, “Development of a digital twin of a mobile robot for research and educational purposes based on MATLAB / Simulink”, XVIII International Conference of Young Scientists "Youth in Science – 2.0'21", submitted for publication (in Russian), in press.

[9] What Is Reinforcement Learning? (Release 2020b), [Online]. Available:https:// www.mathworks.com/help/reinforcement–learning/ug/what–is–reinforcement–learning.html.

[10] Reinforcement Learning Toolbox™ User's Guide 2021a, [Online]. Available: https://www.mathworks.com/help/pdf_doc/reinforcement–learning/rl_ug.pdf.

[11] M. P. Mayorov, “Reinforcement Learning Algorithm for Solving a Robot Motion Problem”, master’s thesis, South Ural State University, National Research University, Chelyabinsk, Russian, 2019. [Online]. Available: https://dspace.susu.ru/xmlui/bitstream/handle/0001.74/29488/2019_222_majorovmp.pdf?sequence=1 (in Russian).

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

[12] “Ellipse”. [Online]. Available: https://ru.wikipedia.org/wiki/Эллипс.

[13] S. Fujimoto, H.–H. Meger, D. Meger, Addressing Function Approximation Error in Actor–Critic Methods, Cornell University, Oct., 2018. [Online]. Available: https://arxiv.org/pdf/1802.09477.pdf.

[14] S. Fujimoto, H. Herke, D. Meger, Addressing Function Approximation Error in Actor–Critic Methods, Cornell University, Oct., 2018. [Online]. Available: https://arxiv.org/abs/1802.09477.

[15] H. Hasselt, A. Guez, and D. Silver, Deep reinforcement learning with double Q–learning. In AAAI, pp. 2094–2100, 2016.

[16] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, Deterministic policy gradient algorithms, In ICML, 2014.

[17] R. S. Sutton, D. A. McAllester, S. P. Singh, and Y. Mansour, Policy gradient methods for reinforcement learning with function approximation, in: Advances in Neural Information Processing Systems, 2000, pp. 1057–1063.

[18] Tune PI Controller using Reinforcement Learning (Release 2021a), [Online]. Available: https://www.mathworks.com/help/reinforcement-learning/ug/tune-pi-controller-using-td3.html.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Nuclei Detection Based on Single-point Labels

Valery Malyshau
Biomedical Image Analysis Department
United Institute of Informatics Problems of NAS of Belarus
Minsk, Belarus
malyshevalery@gmail.com

***Abstract.* Whole-slide image analysis is a long-lasting and laborious process. There are many ways of automatic analysis for histological images. The nuclei detection and classification is one of the most common and medically meaningful medical information-rich methods. However, sometimes the goal of nuclei detection is not to provide detailed information for the medical professionals but to be used for further aggregation. In such cases, nuclei segmentation exceeds requirements and takes extra resources during the data annotation. Keeping this in mind we optimized the existing state-of-art method for nuclei segmentation and classification to work with nucleus centers as input data. Combined with novel optimization technique and neural network activation function it resulted in the algorithm with has improved performance, easier training process and uses input data that is faster to produce.**

***Keywords:* nuclei detection, neural networks, histology, cancer**

I. INTRODUCTION

Whole-slide images (WSIs) are large images showing tissue morphology. Such images are the «gold standard» for cancer diagnosis. However large dimensions that can reach 100000px makes manual analysis an arduous task so histopathologists focus on important areas rather than analyzing every corner of an image [1]. Automatic analysis may reduce the workload but the outcome of algorithms must still be supervised by specialists. An ideal result of the automatic analysis would be the clinical outcome but intermediate results for example nuclei segmentation and classification produce significant insights into the data that can be used by specialists or other algorithms.

Nuclear detection is an essential task that arises during WSI analysis. It helps to quantify WSI for clinical and research reasons [2]. Regularly, detection is a union of segmentation and classification subtasks. In this work, we prioritize nuclear localization without specific boundary selection what helped us to build a more efficient algorithm that requires nuclei centers annotation. Such annotations require less time to obtain in comparison with conventional nuclei boundaries.

The nuclei detection problem is a regular object detection problem so a range of object detection methods were applied to it [3, 4] (F1 score equals 0.50 with classification and 0.94 for nuclei segmentation). However, relatively to WSIs and especially nuclei detection these methods do not show great performance comparing to other methods developed specifically for nuclei detection [5–7] (F1 for nuclei detection and classification falls in range of 0.7-0.86 depending on the method and dataset). There are several reasons for that i.e. large resolution of WSIs, data heterogeneity [8], many nuclei presented even on a small region of an image when current object detection methods work better with a relatively small number of various sized objects.

In this work, we use HoVer-Net [5] as a baseline for our model and evaluation procedure. This deep-learning model produces state-of-art results by combining U-Net shaped architecture with predicting vertical and horizontal maps for nuclei to split nearly located ones and then infer segmentation masks and classes for each presented nuclei.

As mentioned above WSI analysis is a laborious process, so is annotation and data preparation for deep learning. Therefore, we considered using points in nuclei centers as input data rather than complete boundaries around each nucleus. So the data preparation phase may be much faster and cheaper. In some works, instead of nuclei boundaries made by pathologists an algorithm was used to do that [9]. However, such methods can introduce an additional margin of error due to natural inaccuracies in machine learning algorithms. So we propose a method that works without considering nuclei segmentation. Even if in some cases segmentation is required [10], for example when nuclei shape features are the key for predicting an outcome, there are also cases when nuclei location and class would be enough [11] i.e. calculating amount of malignant nuclei on WSI. In addition to developing a deep-learning-based algorithm for processing point-annotated nuclei, we showed improved performance using novel advances among optimization procedures and activation functions for our deep-learning model and HoVer-Net.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The main contributions of this paper are listed below:

- Empirical evidence of performance improvement when novel optimization procedure and activation function are applied to the current state-of-art method for nuclei segmentation and classification.
- Efficient approach for nuclei detection using nucleus center maps scaled by recent advances in deep learning [12].

II. RELATED WORK

The main work which motivated us to work on optimizations and more effective use of medical annotations was the HoVer-Net deep learning architecture [5]. This model outperforms classical approaches from object detection like Mask-RCNN [13] or simplified detection procedures using segmentation networks such as U-Net [14] as well as their combinations [3]. Also, HoVer-Net shows surpassing accuracy in comparison with other deep learning models specified for nuclei detection [5–7]. Regarding the specifics of the deep-learning model, HoVer-Net can be seen as an improvement over typical U-Net with the residual and dense linkage between layers and several branches for classification and segmentation so an increased number of parameters and branches with specified purposes led to improved performance.

As we decided to work with raw nuclei labels in the center of nuclei we rejected models which goal was to create a mask for each nuclei or object in the case of general object detection algorithms. However, this work [15] proposes using predicted centers of objects as anchors for further segmentation by pyramid-like neural networks. Considering we worked with histology data with nuclei of the same size and images depicting the same magnification in borders of a single dataset we did not use feature pyramids and followed the HoVer-Net pattern with residual convolution with dense deconvolution in U-Net shape for predicting nuclei centers like in manner similar to the one described in [12].

III. MATERIALS

To provide relevant evidence that our method is as accurate as the original HoVer-Net we utilized the same subset of nuclei detection datasets. They included one classification dataset called CoNSeP [5] and several segmentation datasets which are CPM15 and CPM17 [16] TNBC [7] and Kumar [17]. We cut images from all datasets into square regions of the same size which we conveniently call tiles in this paper. Every tile was of size 256x256 pixels and they were extracted from original images with 128-pixel step. Detailed information on datasets is shown in Table 1. Also, we used the union of CPM datasets for the sake of comparing them with the HoVer-Net paper. Datasets CPM15 and TNBC which included train test split so we could compare our results with ones provided for the original paper. In other cases, we created a random train test split which stayed the same during the experimental procedure. The same is applied for K-Fold evaluation in which case folds for experiments on one dataset using different methods were the same. Considering the small number of nuclei of miscellaneous class so the train test split affects the accuracy of this class a lot, we removed it from the CoNSeP dataset and in our iteration of HoVer-Net model training we kept this fact in mind.

TABLE I. DATASET INFORMATION

Dataset	N Images	N Nuclei	N tiles
CPM15	15	2 905	306
CPM17	32	7 570	337
Kumar	30	21 623	1470
TNBC	50	4 056	450
CoNSeP	41	24 319	2009

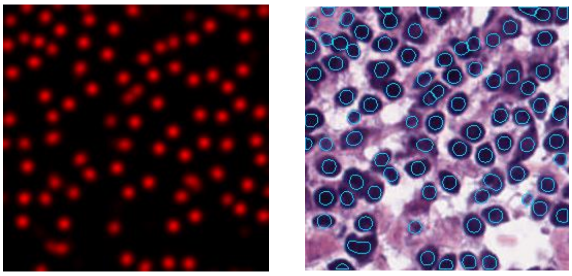

a) b)

Fig. 1. Predicted a) nucleus centers map and b) resulting nucleus central areas

IV. METHOD

The algorithm for nuclei detection consists of two major parts. Firstly, we predict nuclei location using an adapted deep learning model. The deep learning model aims to predict nuclei locations as smoothed points in the nuclei centers. Ground truth data were single points blurred with a Gaussian kernel. Each channel on the resulting pseudo-image was the outcome for a single nuclei class. We trained the model with MSE loss.

Secondly, we ran a watershed algorithm on masks obtained from the deep learning model so the nuclei are split and the final result is obtained by taking the center of mass for each selected region. The example of nuclei center map and result of detection algorithm are presented in the Fig. 1.

Regarding the model, it was built with U-Net shaped architecture where the encoder is constructed from the residual blocks (2 convolutions with activations and residual link at the end) and the decoder

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

is made from dense blocks (3 convolutions with activations and acquired channels are attached to main features). The overview of the model is presented in Fig. 2. The whole algorithm can be viewed as a modification of a deep watershed [18] which uses a watershed algorithm to predict nuclei centers and separate them instead of regular segmentation masks.

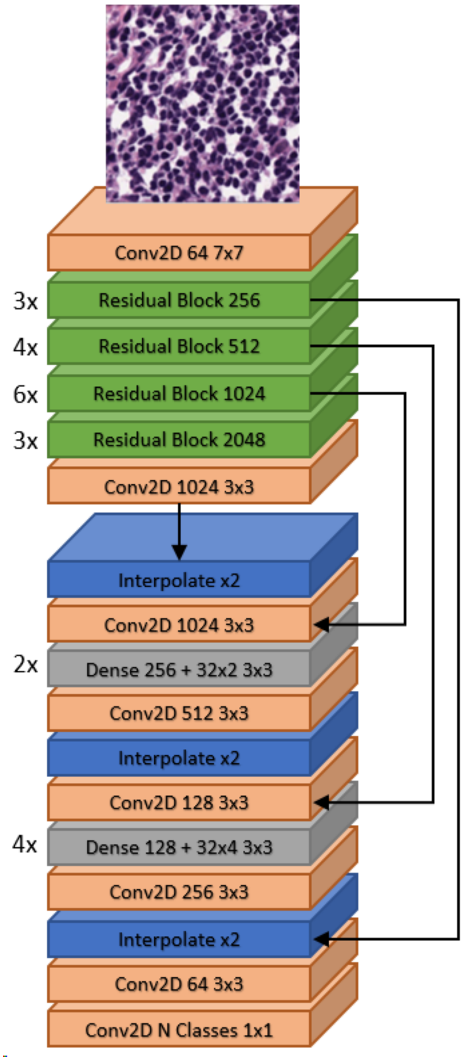


Fig. 2. Proposed Neural Network Architecture

Considering we used batch normalization between convolutional layers we applied novel optimization algorithm AdamP [19]. In addition, we changed conventional ReLU activation to Mish [20]. Together these two advancements improved the accuracy of our deep learning model and HoVer-Net trained for comparison in the same conditions.

V. RESULTS

To perform a thorough evaluation, we trained and tested both our algorithm and HoVer-Net with novel optimization technique and activation function using train test splits recommended by data providers. In several datasets, such splits were not provided so we created a random split and fixed it so the same data was used for training of both methods. On the other hand, we performed K-fold cross-validation with 5 folds to be able to obtain averaged scores considering relatively small sizes of datasets.

For evaluation, we chose F1 scores as our method do not work with segmentation masks so panoptic quality [21], Dice score and Aggregated Jaccard Index which composes detection and segmentation performance are not suitable in our case. Scores for comparison are presented in Table II.

Generally, both our model and HoVer-Net shows similar performance with some fracture of fluctuation.

VI. CONCLUSION

We showed that nuclei can be detected without the segmentation step which regularly is executed in either manual way or by another algorithm or deep learning model. Such an approach does not lose performance while being faster in training and inference. So if the desired result does not include nuclei segmentation it can be avoided while gaining benefits from the proposed model. Additionally, we tested novel optimization and activation function on histology data showing that their performance is higher than traditional methods when applied to histology nuclei detection data.

TABLE II. NEURAL NETWORKS TESTING SCORES

Method		Segmentation					CoNSeP (Classification)				
		CPM15	CPM17	All CPM	Kumar	TNBC	Epithelium	Inflammatory	Spindle	Mean	Detection
HoVer-Net paper		-	0.854	0.774	0.770	0.743	0.635	0.631	0.566	0.565	0.748
K-Fold	HoVer-Net	**0.882**	0.892	**0.896**	0.864	0.878	**0.766**	**0.781**	**0.694**	**0.747**	**0.815**
	Our	0.868	**0.895**	0.893	**0.875**	**0.879**	0.727	0.775	0.658	0.720	0.793
Train/Test	HoVer-Net	-	**0.870**	-	**0.829**	-	0.699	0.677	**0.616**	0.664	0.772
	Our	-	0.862	-	0.812	-	**0.713**	**0.716**	0.614	**0.681**	**0.773**

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

REFERENCES

[1] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, “Histopathological image analysis: a review,” IEEE reviews in biomedical engineering, vol. 2, pp. 147–171, October 2009.

[2] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, “Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential,” IEEE reviews in biomedical engineering, vol. 7, pp. 97–114, December 2013.

[3] A. O. Vuola, S. U. Akram, and J. Kannala, “Mask-RCNN and U-net ensembled for nuclei segmentation,” IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 208–212, April 2019.

[4] H. Narotamo, J. M. Sanches, and M. Silveira, “Combining deep learning with handcrafted features for cell nuclei segmentation,” 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1428–1431, July 2020.

[5] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, “Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images,” Medical Image Analysis, vol. 58, pp. 101563., December 2019.

[6] S. E. A. Raza, L. Cheung, M. Shaban, S. Graham, D. Epstein, S. Pelen-garis, M. Khan, and N. M. Rajpoot, “Micro-Net: a unified model forsegmentation of various objects in microscopy images,” ArXiv e-prints, p. arXiv:1804.08145, April 2018.

[7] P. Naylor, M. La´e, F. Reyal, and T. Walter, “Segmentation of nuclei in histopathology images by deep regression of the distance map,” IEEE Transactions on Medical Imaging, vol. 38, pp. 448–459, August 2018.

[8] Y. Xue, and N. Ray, “Cell detection in microscopy images with deep convolutional neural network and compressed sensing,” arXiv preprint, p. arXiv:1708.03307, August 2017, unpublished.

[9] L. Yang, R. P. Ghosh, J. M. Franklin, S. Chen, C. You, R. Narayan, et al., “NuSeT: a deep learning tool for reliably separating and analyzing crowded cells,” PLoS computational biology, vol. 16, p. e1008193, September 2020.

[10] Q. D. Vu, S. Graham, T. Kurc, M. N. N. To, M. Shaban, T. Qaiser, et al., “Methods for segmentation and classification of digital microscopy tissue images,” Frontiers in bioengineering and biotechnology, vol. 7, p. 53, April 2019.

[11] A. Kumar, and M. Prateek, “Localization of nuclei in breast cancer using whole slide imaging system supported by morphological features and shape formulas,” Cancer Management and Research, vol. 12, p. 4573, 2020.

[12] H. Höfener, A. Homeyer, N. Weiss, J. Molin, C. F. Lundström, H. K. Hahn, “Deep learning nuclei detection: a simple approach can deliver state-of-the-art results,” Computerized Medical Imaging and Graphics, vol. 70, pp. 43–52, December 2018.

[13] X. Xie, Y. Li, M. Zhang, and L. Shen, “Robust segmentation of nucleus in histopathology images via Mask R-CNN,” International MICCAI Brainlesion Workshop , pp. 428–436, September 2018.

[14] M. Z. Alom, C. Yakopcic, T. M. Taha, and V. K. Asari, “Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net),”. NAECON 2018-IEEE National Aerospace and Electronics Conference, pp. 228–233, July 2018.

[15] H. Chen, and H. Zheng, “Object detection based on center point proposals,” Electronics, vol. 9(12), p. 2075, December 2020.

[16] Q. D. Vu, S. Graham, M. N. N. To, M. Shaban, T. Qaiser, N. A. Koohbanani, S. A. Khurram, et al.,“Methods for segmentation and classification of digital microscopytissue images,” Frontiers in bioengineering and biotechnology, vol. 7, pp. 53., April 2019.

[17] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, “A dataset and a technique for generalized nuclear segmen-tation for computational pathology,” IEEE Transactions on Medical Imaging, vol. 36, pp. 1550–1560, July 2017.

[18] M. Bai, R. Urtasun, “Deep watershed transform for instance segmentation,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5221–5229, 2017.

[19] B. Heo, S. Chun, S. J. Oh, D. Han, S. Yun, G. Kim, G., et al., “AdamP: slowing down the slowdown for momentum optimizers on scale-invariant weights,” Proceedings of the International Conference on Learning Representations (ICLR), pp. 3–7, May 2021.

[20] D. Misra, “Mish: A self regularized non-monotonic neural activation function,” arXiv preprint, p. arXiv:1908.08681, October 2019, unpublished.

[21] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, “Panoptic segmentation,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9404-9413, 2019.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Comparison of Deep Learning Preprocessing Algorithms of Nuclei Segmentation on Fluorescence Immunohistology Images of Cancer Cells

Xu Silun
Dept. of System Analysis and Computer Simulation
Belarusian State University
Minsk, Belarus
xusilun@hotmail.com

Victor Skakun
Dept. of System Analysis and Computer Simulation
Belarusian State University
Minsk, Belarus
skakun@bsu.by

***Abstract.* Immunohistology fluorescence image analysis is an important method for cancer diagnosis. With the widespread application of convolutional neural networks in computer vision, segmentation of images of cancer cells has become an important topic in medical image analysis. Although there are many publications describing the success in application of deep learning models for segmentation of different kind of histology images, the universal algorithm is still not developed. The image preprocessing consisting in splitting images in smaller parts and normalization is important in deep learning especially when the training set is of a limited size. In this study, we compared several approaches to create the training set of a sufficient size while having a limited number of labeled whole slide immunohistology images of cancer cells. Also, we explored different normalization methods.**

***Keywords:* CNN, medical image analysis, image preprocessing, image segmentation, nucleus of cancer cell, U-Net**

I. INTRODUCTION

Convolutional neural networks (CNN) have gradually begun to be applied in the field of image analysis since they achieved a huge breakthrough in the field of handwritten font recognition [1]. The outstanding performance of CNN in the ImageNet competition shows that it has great potential in image analysis fields such as image feature extraction and image classification. At present, CNN has widely been applied also in medical image segmentation. The U-Net [2] architecture, specially developed to segment objects like cell nucleus on biomedical images, is widely used in medical image segmentation. The skip connections introduced in U-Net helps to merge the features of different scales that enhance its performance. For example, Neha Todewale successfully applied U-Net to perform segmentation of mammogram images [3]. Ajinkya Jawale et al. made segmentation of the brain tumor images using U-Net [4]. Adnan Saood et al. realized COVID-19 lung CT image segmentation and comparative analysis using U-Net [5]. Other CNN models, such as VGGNet [6], ResNet [7], FCN [8], Inception [9], all based on ideas of deep learning, were developed in the past ten years for image segmentation. Most of them has been successfully applied also in a field of biomedical image segmentation and classification [10]. We selected U-Net for our study because having a relatively simple architecture it still shows very good performance.

The purpose of our research is to segment cell nuclei on the fluorescence images of cancer tissue slices. Cancer tissue slice, stained with fluorescent agents accordingly to a certain protocol, is observed by the fluorescence confocal microscopy to obtain a three-channel color picture. Then the attempt to segment cells (nuclei and cytoplasm) on these images is made for a subsequent quantitative and qualitative pathological analysis.

The segmentation of objects like nucleus or cytoplasm on biomedical images is a pixelwise binary classification problem. The goal is to assign each pixel either to the class of pixels that forms the nuclei area or not. As a step of image segmentation using neural networks, the data preprocessing has a significant impact on the segmentation results. The main purpose of image preprocessing is to eliminate irrelevant information and enhance the detectability of useful information (represented by the pixels or features in the CNN terms) to the greatest extent, thereby improving the accuracy of cell segmentation [11]. For biomedical images segmentation, preprocessing steps usually require cropping, splitting in smaller parts, regularization, intensity enhancement and normalizing to the range [0, 1] and so on [12]. Among them, due to the uneven distribution of cells in biological tissues, preprocessing steps such as the selection of cell nucleus regions also affects the ability to obtain correct segmentation results.

In the cancer pathology analysis, one usually works with samples of relatively small sizes. High diversity of cancer cases does not allow to get a large set of

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

histological images of desired similarity. Another problem is a labeling of target objects on images. When the task is a segmentation of cells or nuclei on images, the person who does labeling must be an expert in both the cancer diagnosis and in the microscopy. Fortunately, each whole slide immunohistology image contains hundreds of cells (target objects for the segmentation). It opens the perspective for enhancing the training set by splitting the input images into a number of small images (patches). Simple splitting images into not overlapped patches, splitting with overlap (sliding), random extraction of patches are ways to enhance the training set [10].

To fulfil this task one question is naturally arising, what size of a patch is optimal for the segmentation of nuclei of cancer cells of a certain average size? Also, it is interesting to examine different strategies of the patch extraction in order to increase the size of the training set. One has to take into account that it is impossible to increase the size of the training set infinitively by a simple extraction more and more overlapped patches. Even in a case of using abundant data augmentation the segmentation will suffer from overfitting. In addition, it is interesting to know how the image preprocessing, like normalization and standardization, influences the segmentation results.

Therefore, the goal of this study is to find the most effective algorithm of preprocessing of immunohistology fluorescence images in the task of segmentation of nuclei of cancer cells.

II. MATERIALS AND METHODS

In this study we are performing the segmentation of fluorescence images of the breast tumor tissue slices. The images where obtained using Nikon TE200 epi-fluorescent inverted microscope equipped with the Photometrics 300 series CCD camera at 10x magnification and stored in RGB color system. The size of the images is 2048 × 2048 pixels in each of the three channels, the resolution is 0.2 μm / pixel, or 5 μm.

The protein estrogen receptor was used as an cancer indicator [13]. In contrast to healthy cells, a protein cytokeratin appears in the cytoplasm of cancer cells. This protein is labeled with cyanine dye Cy3 and registered in the green color channel of the image. To label all nuclei, the 4,6-diamidino-2-phenylindole dihydrochloride (DAPI) dye was used. Its fluorescence was recorded in the blue channel. The cyanine dye Cy5 (recorded in the red channel of the image) was used to label the estrogen receptor, which is located primarily in the nuclei of cancer cells. Accordingly, two dyes, Cy5 and Cy3, are markers of cancer cells.

Nine experimental images were labeled by experts initially semi automatically using CellProfiler (https://cellprofiler.org/) then manually. Labeled images (ground truth) are the binary images, where pixels of nucleus of cancer cells were set to 1.

TABLE I. LAYER STRUCTURE OF U-NET

Encoder layers	Decoder layers
3x3 Conv+ELU, F=16	2x2 ConvTranspose(S=2), F=128 (V)
Dropout=0.1	3x3 Conv+ELU, F=128
3x3 Conv+ELU, F=16 (I)	Dropout=0.2
MaxPool (S= 2)	3x3 Conv+ELU, F=128
3x3 Conv+ELU, F=32	2x2 ConvTranspose(S=2), F=64 (VI)
Dropout=0.1	3x3 Conv+ELU, F=64
3x3 Conv+ELU, F=32 (II)	Dropout=0.2
MaxPool (S= 2)	3x3 Conv+ELU, F=64
3x3 Conv+ELU, F=64	2x2 ConvTranspose(S=2), F=32 (VII)
Dropout=0.2	3x3 Conv+ELU, F=32
3x3 Conv+ELU, F=64 (III)	Dropout=0.2
MaxPool (S= 2)	3x3 Conv+ELU, F=32
3x3 Conv+ELU, F=128	2x2 ConvTranspose(S=2), F=16 (VIII)
Dropout=0.2	3x3 Conv+ELU, F=16
3x3 Conv+ELU, F=128 (IV)	Dropout=0.1
MaxPool (S= 2)	3x3 Conv+ELU, F=16
Middle layers	1x1 Conv+Sigmoid, F=1
3x3 Conv+ELU, F=256	
Dropout=0.3	
3x3 Conv+ELU, F=256	

Abbreviations in the table: S – stride, F – number of filters. The following layers were concatenated: (I)-(VIII), (II)-(VII), (III)-(VI), (IV)-(V).

The U-Net architecture, specially developed for a segmentation of biomedical images and showed good performance, was selected in our study. It is symmetrical neural net and has five sets of convolution/deconvolution layers, see Table I (in TensorFlow notation).

Our realization of U-Net was based on the architecture that shown nice results in the 2018 Data Science Bowl Kaggle competition and available at GitHub (https://github.com/dubeyakshat07/Cell-Nuclei-Image-Segmentation-using-U-Net). The model has 1,941,105 learning parameters. We used binary cross entropy as the loss function and intersection over union (IOU, known also as Jaccard similarity coefficient) calculated at the threshold 0.5, as the segmentation metric. The smaller the loss value, the better the learning effect. The larger the IOU metric – the better the segmentation (closer to the ground truth).

Taking into account the wide range and scalability of the application, we used a Python language environment to build the U-Net model, and selected the Jupyter Notebook to run and test the code. The main model components were implemented through the Keras library (TensorFlow kernel). Augmentation was performed using the Albumentations library (https://albumentations.ai/) and calculation of the loss function and the evaluation metric was done using the Segmentation models library (https://segmentation-models.readthedocs.io).

III. RESULTS AND DISCUSSIONS

As we have only 9 labeled images (of size 2048 × 2048 px.) the main goal at deep learning is to increase the size of the training set. The rough estimation of the nuclei size of 1250 ± 800 pixels corresponds approximately to the circle with radius of 20 px. It allows us to split the input images into a set of much smaller sizes – patches, thus defining a training set of optimal size. The dimension of the patch size is not trivial to guess because the segmentation results are obviously depended on how close to the ground truth is the predicted border of a nucleus. If we select small patch window the learning process will definitely suffer from the fact that probably most of nuclei will be located not in the center of the window, but over their borders. In opposite case we can get just rough estimation of the nuclei area. The answer to the question how many times an overlapped part of the image can be present in the training set is also not evident. The overfitting makes useless all our work otherwise.

Therefore, to answer these questions we initially performed the segmentation of nuclei at different patch sizes and extraction strategies. At the second step we have studied several methods of preprocessing the patches such as image histogram equalization, standardization and normalization to the range [0, 1].

The U-Net model used in our study is described in details in the Materials and Methods section. We selected a simple augmentation including random 90-degree rotation, horizontal flipping and transposing. Although we utilized simple augmentation, it can effectively prevent overfitting. The intersection over union (at level 0.5) metric was set as a measure of accuracy of the segmentation. Seven images were used for training, one image for validation and one for testing. The same algorithm of preprocessing was applied to both training and validation images. The empty patches (the number of pixels less than 150) were removed from the training set. Number of epochs was set to 30. Earlier stopping callback with patience = 8 was used in the training process.

The first task is to obtain a sufficiently large training set and to avoid overfitting. We started with a patch size of 512 × 512 and decreased it to 64 × 64. The latter case gave much worse results, probably because the most nuclei crossed borders of patches, therefore we did not include it into the analysis. More specifically we studied the following cases: 1) simple splitting into patches of the size 128 × 128 without overlap; 2) splitting into patches of the size 128 × 128 with overlap in 64 px. (1/2 of the patch size); 3) splitting into patches of the size 256 × 256 with overlap in 128 px. (1/2 of the patch size); 4) random cropping 128 × 128 patches 8 times after intermediate splitting into 256 × 256 subimages; 5) splitting into patches of the size 512 × 512 with overlap in 128 px. (1/4 of the patch size). The side lengths of patches are therefore 128, 256, and 512 pixels. At the case 1) we got the training set of 1576 images. A at the case 4) we got the training set of 12064 images.

The values of the loss function and the IOU metric on the test set are summarized in the TABLE II. Other relevant information about the number of patches and the total number of pixels in the training set as well as the batch size, number of steps per epoch is also summarized there. The values of the loss function and IOU metric for 30 CNN epochs, obtained for both training and validation data, are shown in Fig. 1.

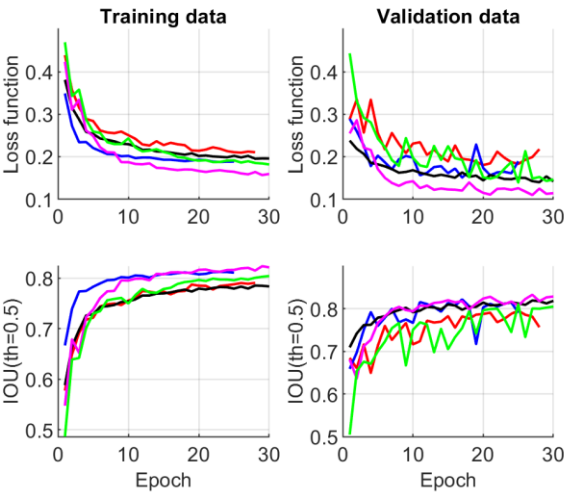


Fig. 1. Values of the loss function and IOU metric for 30 CNN epochs, after cases 1-5 of patch extraction methods. Color coding:| case 1 – red, case 2 – blue, case 3 – green, case 4 – black, case 5 – magenta

In spite of the fact that in the case 5 we got the best metric value, we conclude the case 4 with random extraction of patches of a size 128 × 128 to be the best. This conclusion came from the visual inspection of the segmentation results (data not shown). Here we use a simple but effective method to increase the size of the training set. In this case, patch overlap is useful, but when the number of patches reaches a certain level, then overfitting is occurring. When the number of patches is at the largest, we get the best results. In contrast, the maximum number of pixels is achieved in the case 5. One may come with a conclusion that if augmentation is performed then as larger the training set the better the results despite of the presence of overlapping areas of the images. To prove or reject this conclusion we did an experiment (data not shown) where we selected the case 2 but increased virtually twice (because of augmentation) the training set by selection of larger number of steps per epoch to get the same number of patches in the training process. We got slightly worse value of the IOU metric (= 0.834). Thus, the random extraction of patches is better. We also tried the case with random extraction of four patches from 256 × 256 sub images. The result was worse than at the simplest case 1. Therefore, we proved that the random cropping is a powerful method but only when enough number of patches is extracted to obtain a complete coverage of the image. Also, we came to the conclusion that the patch of the size 128 × 128 px. is optimal for the segmentation of nuclei with the average radius of about 20 px.

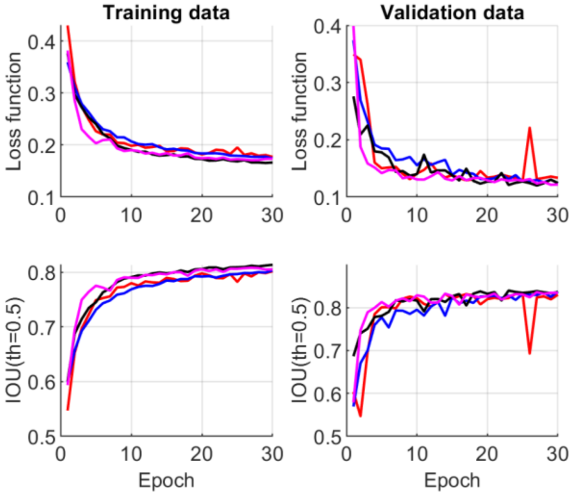


Fig. 2. Values of the loss function and IOU metric for 30 CNN epochs after different methods of patch normalization. Color coding: case 1 – black, case 2 – red, case 3 – magenta, case 4 – blue.

In the second step of our study, we picked out the case 4 of randomly selected patches with the side length of 128 pixels. The following methods of image normalization and standardization were studied: 1) raw data (without normalization); 2) normalization to the range [0, 1], where 1 corresponds to the maximum pixel intensity in the patch; 3) centering and scaling; 4) normalization to the range [0, 1] by initial patch histogram equalization then division to 255.

Surprisingly we obtained very similar results. The best method was the image histogram equalization (see Table III and Fig. 2).

The method of histogram equalization effectively increases the contrast of the cell nucleus, therefore supports its better segmentation. However, differences in the IOU metric are very small. Therefore, in our case, image normalization does not play a key role (see Table III).

IV. CONCLUSION

The preprocessing of input images plays important role in the deep learning. Having only 9 whole slide labeled images of the size 2048 × 2048 pixels we were able to get satisfactory results in the segmentation of nuclei on the fluorescence images of the cancer cells of the breast tumor. We concluded that the random cropping is a powerful method for increasing the training set when the number of patches is high enough to obtain a complete coverage of the input image. The simple splitting of the image with overlap may be sufficient if the augmentation is used. Also, we came to the conclusion that the patch of the size 128 × 128 px. is optimal for the segmentation of nuclei with the average radius of about 20 px.

In our numerical tests the image normalization does not play a key role. Histogram equalization with subsequent normalization to the range [0, 1] demonstrated the best result. The method of histogram equalization effectively increases the contrast of the cell nucleus and thus supports the better segmentation.

TABLE II. RESULTS OF SEGMENTATION AFTER DIFFERENT METHODS OF PATCHES CREATION

Parameters and metrics	**Methods of patches creation**				
	Splitting to 128×128 without overlap	**Splitting to 128×128 with overlap in 64 px**	**Splitting to 256×256 with overlap in 128 px**	**Random cropping 128×128 8 times after intermediate splitting into 256×256 images**	**Splitting to 512×512 with overlap in 128 px**
Total number of patches	1576	5939	1508	12064	1177
Total number of pixels in all patches	25,821,184	97,304,576	98,828,288	197,656,576	308,543,488
Batch_size	64	64	24	32	16
Steps per epoch	25	93	63	377	74

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Parameters and metrics	Methods of patches creation				
	Splitting to 128×128 without overlap	**Splitting to 128×128 with overlap in 64 px**	**Splitting to 256×256 with overlap in 128 px**	**Random cropping 128×128 8 times after intermediate splitting into 256×256 images**	**Splitting to 512×512 with overlap in 128 px**
Loss function	0.222	0.188	0.156	0.156	0.157
IOU (th=0.5)	0.786		0.837	0.837	0.839

TABLE III. RESULTS OF SEGMENTATION AFTER DIFFERENT METHODS OF PATCHES NORMALIZATION

Parameters and metrics	Methods of patches normalization			
	Without normalization	**Division by max value of pixel intensity in a patch**	**Standardization**	**Histogram equalization**
Loss function	0.159	0.154	0.160	0.154
IOU (th=0.5)	0.837	0.838	0.838	0.840

REFERENCES

[1] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," IEEE proc., vol. 86(11), pp. 2278–2324, Nov. 1998.

[2] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," arXiv:1505.04597v1 [cs.CV], May 2015.

[3] N. S. Todewale, "Lesion segmentation from mammogram imagesusing a U-Net deep learning network," IJERT, vol. 9(02), Feb. 2020.

[4] A. Jawale, P. Warole, S. Bhandare, K. Bhat, P. R. Chandre, "Jeevn-Net: Brain tumor segmentation using cascaded U-Net & overall survival prediction," IRJET, vol. 7(01), Jan. 2020.

[5] A. Saood, I. Hatem, "COVID-19 lung CT image segmentationusing deep learning methods: U-Net versus SegNet," BMC Med Imaging, 21:19, 2021.

[6] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556v6 [cs.CV], Apr. 2015.

[7] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," arXiv:1512.03385v1 [cs.CV], Dec. 2015.

[8] J. Long, E. Shelhamer, T. Darrell, “Fully convolutional networks for semantic segmentation,” arXiv:1411.4038v2 [cs.CV], Mar. 2015.

[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, et al, "Going deeper with convolutions," arXiv:1409.4842v1 [cs.CV], Sept. 2014.

[10] Q. D. Vu, S. Graham, T. Kurc, M. N. N. To, M. Shaban et al., "Methods for segmentation and classification of digital microscopy tissue images," Front. Bioeng. Biotechnol, vol. 7, article. 53, April 2019.

[11] S. Dakhare, H. Chowhan, M. B. Chandak, "Combined approach for image segmentation," IJCTT, vol. 11(03), May 2014.

[12] Y. Lisitsa, M. Yatskou, V. Skakun, A. Digris, I. Shinhariev, V. Apanasovich, "Simulation model to study denoising methods," Proc. of the 12th Intern. Conf. PRIP, Minsk [UIIP NASB, 2014, pp. 157–160].

[13] G. G. Chung, M. P. Zerkowski, S. Ghosh, R. L. Camp, D. L. Rimm, "Quantitative analysis of estrogen receptor heterogeneity in breast cancer," Lab Invest, vol. 87, no. 7, pp. 662–669, 2007.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Multiple Human’s Pose Detection Algorithm for Real Time Video

Aliaksandr Leunikau
Belarusian State University
Minsk, Belarus
Alex.levnikov@gmail.com

Alexander Nedzved
FAMCS of
Belarusian State University
Minsk, Belarus
NedzvedA@tut.by
ORCID 0000-0001-6367-5900

Stanislav Sholtanyuk
Belarusian State University
Minsk, Republic of Belarus
ssholtanyuk@bsu.by
ORCID 0000-0003-0266-7135

Alexei Belotserkovsky
Dept. of Intelligent Information Systems
United Institute of Informatics Problems of NAS of Belarus
Minsk, Belarus
alex.belot@gmail.com

***Abstract.* This article describes a realtime algorithm to determine a person’s posture at a certain point in time.**

***Keywords:* human’s pose recognition, virtual human skeleton, object movement analysis, neural networks, image recognition**

I. INTRODUCTION

The pace of development of technology and robotics in the modern world is striding far into the future, but there are still many unsolved problems assigned exclusively to humans. Isolation of familiar faces from the environment, determination of the character and state of a person by non-verbal gestures and facial expressions, determination of emotions, type of activity, and occupation. All of these tasks are considered purely human. But thanks to the non-linear advances in technology, computer vision, machine learning, and artificial intelligence, robotics and almost any electronic device equipped with sufficient resources can learn how to solve these problems.

This research is aimed at solving creative, inherent only to humans, problems. Specifically, tasks related to pattern recognition and positioning of objects in space.

This topic is relevant for many areas of human life, from mass media systems to medicine and security. The algorithm can be used to analyze gestures in sign language translation, analyze people's behavior by security cameras, overlay animations in the film and game industries.

The purpose of this work is to develop a new algorithm for analyzing the positioning of people in space, capable of increasing the efficiency of solving problems of image recognition in images.

II. PROBLEM REVIEW

Assessing a person's positioning is a problem of localizing anatomical key points, later called body parts, which in physical terms are the joints points of the human body, with the exception of the face key points. This problem is mainly focused on finding body parts, combining them into a complete skeleton, and tracking it throughout the video sequence.

Recognizing the poses of several people in an image, especially socially active ones, presents a unique set of challenges:

- Each image can contain an unknown number of people appearing and disappearing at any position and scale.
- Interactions between people cause complex spatial interference, which is caused by occlusions, physical contacts between people, properties of joints, which greatly complicates the unification of individual parts of the body in the limb, and later into the skeleton.
- The complexity of execution increases with the number of people, which imposes significant restrictions on the mode of execution in real-time.

III. SOLUTION REVIEW

The algorithm is a bottom-up method for estimating associations through part affinity fields, a set of two-dimensional vector fields that encode the location and orientation of limbs in an image region.

The algorithm has three main steps:

- Image preprocessing.
- Simultaneous body parts detection and association.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

- Multiple people processing.

A. IMAGE PREPROCESSING

The first step of the algorithm is the analysis of the image by a convolutional network consisting of the first 10 layers of the VGG-19 network [1] and configured directly to generate characteristic maps, which, as a result of post-processing, send a set of these maps to the second step of the algorithm. The input is a color RGB image with size w×h, where w is width and h is height.

B. SIMULTANEOUS BODY PARTS DETECTION AND ASSOCIATION

The neural network simultaneously predicts a set of probability maps $S_j \in R^{w\times h}$, $j \in \{1 \dots J\}$ and a set of two-dimensional compatibility vector fields L, which store the degree of compatibility between limbs. The set $L = (L_1, L_2, \dots, L_c)$ has vector fields C, where $L_c \in R^{w\times h\times 2}$, $c \in \{1 \dots C\}$. Each image position in L_c is encoded by a 2D vector.

The network is divided into two branches: the upper branch, predicts probability maps, and the lower branch, predicts the compatibility fields. Every branch has an iterative predictive architecture. An iteration stage, following Weil's method [2], refines the network predictions with intermediate control after each stage $t \in \{1 \dots T\}$.

At the first stage (t=1), the network creates a set of probability maps $S^1 = p^1(F)$ and a set of compatibility fields of parts $L^1 = b^1(F)$, where p^1 and b^1 are convolutional neural networks. At each stage t, the previous stage predictions S^t and L^t are combined with the initial characteristics F to obtain refinements of the network forecasts:

$$S^t = p^t(F, S^{t-1}, L^{t-1}), \forall t \geq 2,$$

$$L^t = b^t(F, S^{t-1}, L^{t-1}), \forall t \geq 2,$$

where p^t and b^t – convolutional neural networks outputs at stage t

We use an error L^2 between network predictions and known correct maps and fields. Here we spatially weigh the error functions to get rid of the problem that some datasets do not fully label all people

$$f_S^t = \sum_{j=1}^{J} \sum_{p} W(p) * \left\| S_j^t(p) - S_j^*(p) \right\|_2^2 \ ,$$

$$f_L^t = \sum_{c=1}^{C} \sum_{p} W(p) * \| S_c^t(p) - S_c^*(p) \|_2^2,$$

$$f = \sum_{t=1}^{T} (f_S^t + f_L^t),$$

where S_j is the probability map, L_c is the vector compatibility field, W is the binary mask, p is pixel of image. If $W(p) = 0$ there is no characteristic information on the p.

The mask is used to avoid errors in the validity of positive predictions during training. Intermediate control at each stage solves the vanishing gradient problem by periodically replenishing the gradient.

1) *Building Probability Maps to Detect Body Parts*

To evaluate f_s during training, we generate probability maps from two-dimensional key points. Each probability map is a two-dimensional representation of what happens to a specific body part for each pixel. Ideally, if there is one person in the image, only one peak should be present in each probability map if the corresponding part is visible; if more than one person occurs, there must be a peak corresponding to each visible part j for each person.

First, we create individual probability maps, for each person. The x_j is the position of body part j, $j \in R^2$, for person k, $k \in R^2$. The value of the probability map at the point p $\in R^2$ is determined by the formula:

$$S_{j,k}^* = e^{-\frac{\left\| p - x_{j,k} \right\|_2^2}{\sigma^2}},$$

where σ corresponds to control of the concentration of the peak region.

The probability map to be predicted by the network is a collection of individual probability maps for every human's pose:

$$S_j^*(p) = max_k \left(S_{j,k}^*(p) \right),$$

We accept the maximum of all maps instead of the average to better demonstrate the accuracy of the peaks in occlusion. As result the probability maps predict positions of candidates for body parts.

2) *Compatibility Vector Fields Calculation*

A set of detected body parts is processed to form the poses of an unknown number of people. To do this, It is need a reliable measure of compatibility for every pair of detected body parts. One of the possible ways to measure compatibility is to introduce an additional point - the middle between each pair of parts and check its belonging to the limb. However, when people come together, these midpoints are likely to give false associations. These false associations arise from two limitations:

- This method encodes only the position, not the orientation of each limb.
- It reduces the support area of the limb to one point.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

To address these limitations, a new representation, called part compatibility fields, is created. It retains both location information and orientation throughout the support area of the limb. The compatibility (affinity) of the parts is a two-dimensional vector field for each limb. For each pixel in an area belonging to a specific limb, the 2D vector encodes a direction from one part of the limb to another. Each type of limb has a corresponding field of compatibility connecting its body parts into a limb.

For evaluation f_L during training, we determine the vector compatibility field at point on the image p defined by formula:

$$L^*_{c,k}(p) = \begin{cases} v \text{ if } p \text{ lies on the limb } c,k \\ 0 \ otherwise \end{cases}$$

Let $x_{j1,k}$ and $x_{j2,k}$ be the positions of the joints of body parts $j1$ and $j2$ for a person k. If point p lies on a limb, the value of $L^*_{c,k}(p)$ is a unit vector $\boldsymbol{v}$ that points from $j1$ and $j2$ and shows the direction of the limb. The vector is equal to zero for all other points t:

$$\boldsymbol{v} = \frac{(x_{j1,k} - x_{j2,k})}{\left\| x_{j1,k} - x_{j2,k} \right\|_2}.$$

The set of points is defined as the number within the segment, those points p for which inequalities:

$$0 \leq \boldsymbol{v} * (p - x_{j1,k}) \leq l_{c,k},$$

$$\left| v_\perp * (p - x_{j1,2k}) \right| \leq \sigma_i,$$

where σ_i is limb width, the distance in pixels, $l_{c,k}$ is a limb length, $k = \left\| x_{j1,k} - x_{j2,k} \right\|_2$, and $v_\perp$ is a perpendicular to $\boldsymbol{v}$ vector (Fig. 1).

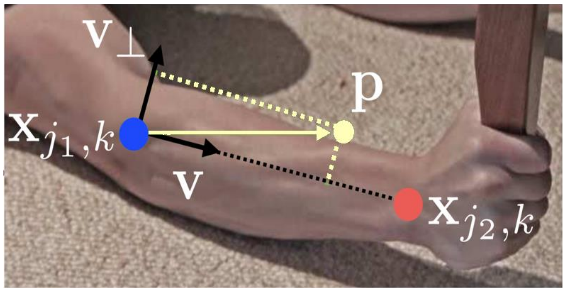


Fig. 1. Posipions of vectors for body part

The general affinity vector field $L^*_c(p)$ averages the fields $L^*_{c,k}(p)$ of all people on the image and is given by formula:

$$L^*_c(p) = \frac{1}{n_c(p)} \sum_k L^*_{c,k}(p),$$

where $n_c(p)$ is the number of nonzero vectors at point p for all k people (i.e., the average value in pixels where individual limbs for coordinate of people overlap).

For testing, we use measurement of the relationship between the identified candidates by calculating the sum:

$$E = \sum_{u=0}^{1} L_c(p(u)) * \frac{d_{j1} - d_{j2}}{\left\| d_{j1} - d_{j2} \right\|_2},$$

were $p(u)$ function that is calculated over the corresponding vector field, along the segment connecting the locations of the candidates:

$$p(u) = (1-u)d_{j1} + u \cdot d_{j2}.$$

In other words, we measure the alignment of the predicted field with the limb, which will be formed by connecting the detected body parts. In particular, for candidates d_{j1}, d_{j2} we project the affinity field of the predicted portion L_c along the line segment to measure the accuracy of their relationship.

In practice, we approximate the integral by sampling and summing uniformly distributed u values.

C. MULTIPLE PEOPLE PROCESSING

At this stage, probability maps and affinity fields are analyzed to output two-dimensional key points for all people in the image.

The algorithm does not perform the maximum suppression on the probability maps to obtain a discrete set of candidates for the next processing steps. Due to the analysis of all people at once or inaccuracy in image processing, several candidates may arise at once for each part of the body. These candidates constitute a large set of possible limbs. Every limb is controlled using the sum E.

The optimal selection problem corresponds to the K-dimensional matching problem, which is known to be NP-hard. This seems to be a rather expensive relaxation that consistently determines high-quality matching of parts in the limb. This is because the pairwise associative evaluation implicitly encodes the global context due to the high sensitivity of the compatibility fields.

First, we get a set of candidates D_J for several people, where $D_j=\{d^m_j$, for $j \in [1,...,J]$, $m \in [1,...,N_j]\}$, N_j is the number of candidates for part j and $d^m_j \in R^2$ is the location of the m-candidate for body part j. These candidates still need to be connected to other body parts from the same person. In other words, we need to find candidate pairs that are actually a connected limb. For this, a variable $z^{m,n}_{j1,j2} \in [0,1]$ is a flag indicating whether the two candidates are connected to each other. The goal is to find the optimal assignment for the set of all possible connections:

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

$$Z = \{z_{j1,j2}^{m,n}: for\ j1, j2\ \in [1, \dots, J], m \in [1, \dots, N_{j1}], n \in [1, \dots, N_{j2}]\}.$$

If we consider one pair of parts j_1, j_2 (for example, the neck and the right thigh) for the limb c, then the search for the optimal match is reduced to the problem of matching bipartite graphs with maximum weight [3]. In this problem, the nodes of the graph are the candidates for detecting a body part D_{j1} and D_{j2} and the edges are all possible connections between pairs of candidates. In addition, each edge has its own weight. A concordance in a bipartite graph is a subset of edges selected in such a way that no two edges separate a node. Our goal is to find a match Z with the maximum weight for the selected edges:

$$max_{Z_c} E_c = max_{Z_c} \sum_{m \in D_{j1}} \sum_{n \in D_{j2}} E_{m,n} z_{j1,j2}^{m,n},$$

$$\forall\ m \in D_{j1}, \sum_{n \in D_{j2}} z_{j1,j2}^{m,n} \ \leq 1 \square$$

$$\forall\ n \in D_{j2}, \sum_{m \in D_{j1}} z_{j1,j2}^{m,n} \ \leq 1 \square$$

Where E_c is the total weight of the correspondence for the part type C, Z_c is the subset Z for the part type C, $E_{m,n}$ is the correspondence between the parts d^m_{j1} and d^n_{j2} .

The equations stipulate that two limbs of the same type do not share one part. We can use the Hungarian algorithm [4] to get the optimal match.

When it comes to finding the whole pose of many people, the Z definition is a K-dimensional matching problem. This problem is NP hard and there are many relaxations. We add two more relaxations to optimize this algorithm:

- The minimum number of edges is chosen to obtain the skeleton of a human pose, rather than using the full graph.
- The matching task is decomposed into a set of bidirectional matching subtasks and independently determines the matching in adjacent tree nodes:

$$max_Z E = \sum_{c=1}^{C} max_{Z_c} E_c \ .$$

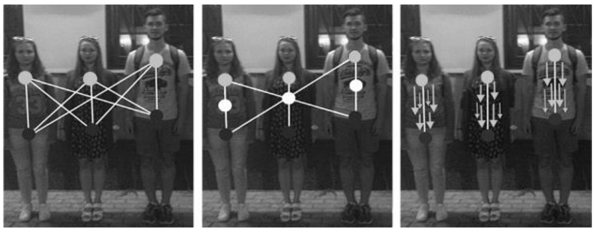


Fig. 2. Joints of body parts for every person on image

Therefore, candidates for limb connection are obtained independently of each other. These candidates are then assembled into limbs and combined into full body poses for multiple people (Fig. 2). This optimization scheme for a tree structure is an order of magnitude faster than optimization for a fully coupled schedule [5, 6].

The result of the algorithm is a full-size virtual human skeleton, displaying the positioning of his anatomical body parts, as well as the position of the whole body.

IV. SOLUTION TESTING AND RESULTS

The algorithm is evaluated according to two criteria:

- MPII multiuser dataset,
- COCO milestone dataset.

These two datasets contain images with different scenarios that contain many real-world problems such as large numbers of people, zooming, occlusion, and contact. Our approach set a new record for the speed of work on the COCO dataset [7] and significantly exceeds the previous modern results on MPII data [8] for several people. The table below shows some of the qualitative results of the algorithm.

A. *Test Results on MPII Multi-Person Dataset*

For comparison on the MPII dataset, we use the toolkit [32] to measure the mean accuracy (mAP) of all body parts based on the PCKh threshold. Figure 3.1 compares mAP performance between our method and other approaches. First, over a subset of 288 test images, and then over the entire MPI test suite. Besides the mAP score, we compare the average processing and optimization times on the image.

For a subset of 288 images, our method outperforms previous modern bottom-up methods by 8.5% mAP. It is noteworthy that our inference time is 6 orders of magnitude less than competing algorithms. Section 3.3 analyzes the runtime in more detail.

For the entire set of MPII tests, our method already surpasses the previous modern algorithms by a large margin, that is, by 13% mAP. Using 3-scale search (x0.7, x1 and x1.3) further increases performance to 75.6% mAP. Comparing mAP with previous bottom-up approaches shows us how effective the representation of functions, PAF, are for linking body parts. Based on a tree structure, our expensive processing method provides better accuracy than a graph optimization formula based on a fully coupled graph structure [32, 33].

B. *Test results on COCO Keypoints Challenge dataset*

The COCO training set consists of over 100,000 people with over 1 million major key points (body parts). The test suite contains subsets of "test-

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

challenge", "test-dev" and "test-standard", which have approximately 20K images.

Algorithm	***AP***	***AP-50***	***AP-75***	***AP-M***	***AP-L***
Current	**60.5**	**83.4**	**66.4**	55.1	**68.1**
G-RMI	59.8	81.0	65.1	**56.7**	66.7
DL-61	53.3	75.1	48.5	55.5	54.8
R4D	49.7	74.3	54.5	45.6	55.6

The COCO score determines the comparability of the recognized points (OKS) and uses the average accuracy (AP) at more than 10 OKS thresholds as the main scoring criterion. OKS is calculated based on the size of the person and the distance between the predicted points and the GT points. Fig. 3.2 shows the results from the best teams for a task. It should be noted that our method has lower accuracy than top-down methods for people with smaller scales (APM) [16]. The reason is that our method has to deal with a much wider range of scales covered by all the people in the image in one frame. Top-down methods, on the other hand, can scale each deterministic region to a larger size and thus reduce the error at smaller scales.

C. Realtime Tests

To analyze the performance of our method, we collect videos from different numbers of people. The original frame size is 1080×1920, which we change to 368×654 during testing to match the GPU memory. Runtime analysis is performed on a laptop with a single NVIDIA GeForce GTX-1080 GPU.

In Fig. 3 we are using person detection and CPM for one person as opposed to top-down approaches, where the execution time is roughly proportional to the number of people in the image. The time it takes to complete our bottom-up approach grows relatively slowly as the number of people increases. The lead time consists of two main parts:

CNN processing time with O (1) complexity, which is constant with different numbers of people

Time of distribution of parts between people, the complexity of which is O (), where n is the number of people.

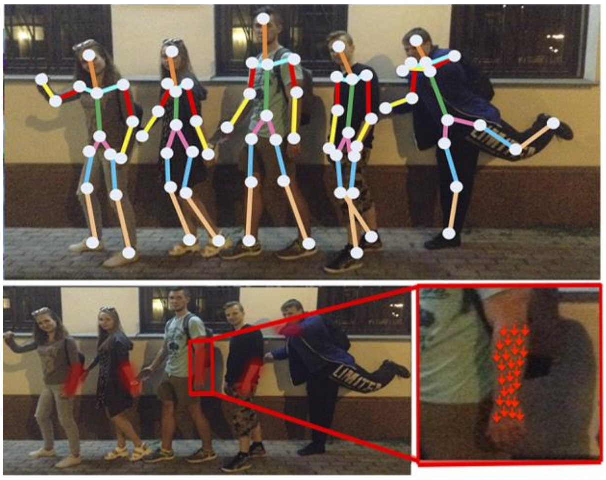


Fig. 3. Results detection of positions of body parts and intermediate vectors contruction

The parsing time does not have a significant impact on the overall execution time, as it is two orders of magnitude less than the processing time of CNN, for example, for 9 people, parsing takes 0.58 ms and CNN takes 99.6 ms. Our method achieved 8.8 frames per second for videos with 19 people.

ACKNOWLEDGMENT

This research was partly supported under BRFFR projects (F20KIGT-006 and F21UKR-001).

REFERENCES

[1] M. Sun, S. Savarese, Articulated part-based model for joint object detection and pose estimation, ICCV, 2011.

[2] D. B. West, Introduction to graph theory, vol. 2, Prentice hall Upper Saddle River, 2001.

[3] Y. Yang, D. Ramanan, Articulated human detection with flexible mixtures of parts, TPAMI, 2013.

[4] A. Newell, K. Yang, J. Deng, Stacked hourglass networks for human pose estimation, ECCV, 2016.

[5] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler, and B. Schiele, Deepcut: Joint subset partition and labeling for multi person pose estimation, CVPR, 2016.

[6] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, B. Schiele, Deepercut: A deeper, stronger, and faster multi-person pose estimation model, ECCV, 2016.

[7] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, C. L. Zitnick, Microsoft COCO: common objects in context, ECCV, 2014.

[8] X. Chen, A. Yuille, Articulated pose estimation by a graphical model with image dependent pairwise relations, NIPS, 2014.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Small Image Training Sets: Exploring the Limits of Conventional and CNN-based Methods

Vassili Kovalev
Biomedical Image Analysis Dept.
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
vassili.kovalev@gmail.com

Abstract. **This work is dedicated to the problem of image classification under the condition of small image datasets. Both traditional and CNN-based methods are examined and compared based on a benchmark image dataset. The dataset consisted of 12000 routine hematoxylin-eosin stained histological images. They represent the biopsy samples of normal tissue and the malignant tumors caused by breast cancer. The commonly-known image analysis methods which make use of color co-occurrence matrices of images converted to an adaptive 32-color space and the limited number of their principal components (PCA) were used as image features. The features were inputted to SVM and Random Forests classifiers. The original image training set was gradually reduced from 8400 to 840 images with the step of 10%. In addition, the very-small sub-samples of 5% (420), 2.5% (210), 1.25% (105), and 1% (84) of original image dataset were also examined. In its turn, the classical CNN was employed that consisted of only 3 convolutional + MaxPooling layers with 16, 32, and 64 filters respectively. This is because the small image training sets were specifically targeted in this particular study. The convolutional part was followed by a fully connected neural network with 512 intermediate nodes. As a result, it was found that traditional methods outperform the CNN-based image classification technique on the training sets comprised of less than 840 images.**

Keywords: **Image Classification, Benchmarking, Convolutional Neural Networks, Histology images**

I. INTRODUCTION

Nowadays, the Convolutional Neural Networks (CNNs) and Deep Learning (DL) techniques are widely used for solving various image processing, segmentation, classification, clustering, and even realistic image generation problems. These methods have demonstrated tremendous promises in different application domains including medical image analysis, classification, and computerized disease diagnosis [1, 2]. However, training of CNNs with recent architectures requires large amounts of professionally labeled medical images of different classes that could be difficult to collect, laborious to label, and costly.

The histopathology image analysis based on light microscopy has long been recognized as a gold standard in cancer diagnosis. Modern digital pathology which includes whole slide imaging (WSI) scanners and automated image analysis solutions provides a more efficient and cost-effective way of handling, visualization, and analysis of the pathology image data [3]. Although the conventional methods of WSI image analysis based on the extraction of color and morphological features are still in use [4], the new DL approaches often demonstrate better performance and higher tolerance to image variability caused by a number of different factors [5, 6]. In [6] authors have isolated, carefully enumerated, and characterized 10 major challenges of AI in digital pathology which we are currently facing. The challenges that are most relevant to the present study include lack of labeled data, pervasive variability, and so-called realism of DL which is associated here with the available computation power.

Presently, it is commonly understood that the classification results always depend on the degree of representativeness of images included in the training set. Therefore, it is highly desirable that these images should be as “representative” as possible for the classification problem we are dealing with. In terms of the feature space, this means that the training image samples should cover well the regions of feature space that could be potentially populated by the image classes we considering. Such a problem is directly relevant to the following two major factors: the size of the training set and, (b) the variability of images inherent to the classes.

In this paper, we are trying to shed light on the problem of small image training sets and image variability on the typical example of histopathological images used for breast cancer diagnosis. Both traditional and CNN-based methods are examined and compared based on common histological image datasets used as benchmarks.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

II. MATERIALS AND METHODS

A. Patients and Whole Slide Images

The whole slide histopathological images acquired from biopsy samples of 90 different patients suspicious for breast cancer were used as the source of image data. These WSIs represent a sub-sample of hematoxylin-eosin stained images of lymph node sections used in the Grand Challenge [5]. The challenge was aimed at discovering the best methods and algorithms for detecting breast cancer metastases. A total of 76 WSIs contained metastases of different sizes whereas the other 14 not presented any pathological changes and were considered to be the norm. An example fragment of a WSI image, as well as the high-resolution picture of its inhomogeneous region, are shown in Fig. 1. It should be noted that histological images of biopsy samples may contain both normal and tumor regions simultaneously what is clearly demonstrated by Fig. 1.

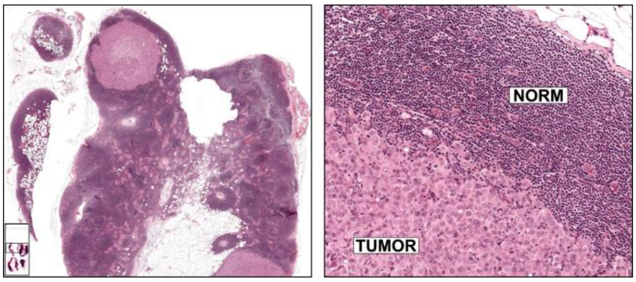


Fig. 1. Example fragment of WSI image
and its characteristic region

B. Image datasets

Original professionally-labeled WSI images were partitioned into non-overlapping image sections (image tiles) of 256×256 pixels in size at the highest resolution level that corresponds to the x40 optical microscope magnification. A total of 12000 tiles including 6000 tiles of the norm and 6000 tiles of tumor were randomly sub-sampled from the resultant set of tiles. Examples of the two image classes are given in Fig. 2.

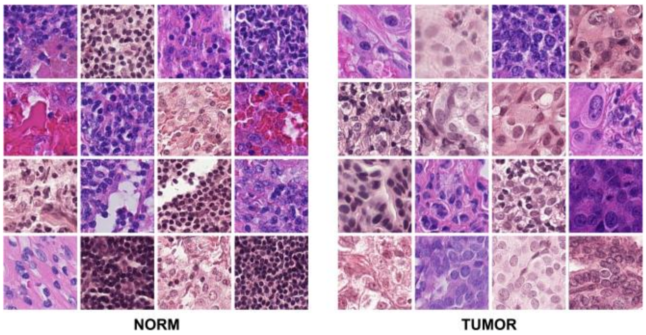


Fig 2. Example image tiles of two classes

The well-balanced train and test image datasets were created following the 70/30 percent proportion. This has resulted in 8400 image tiles included in the training set (4200 tiles of norm plus 4200 tiles representing the tumor) and 3600 images (1800 of norm regions and 1800 of tumor) used for testing. In a similar study [4], we experimentally confirmed that WSI tissue images of each particular patient are holding certain characteristic image patterns (features) which makes them somewhat different from any others. As a result, including image tiles of one single WSI to both training and test sets creates a bias that resulted in an artificial increase of classification accuracy. With this in mind, here, image tiles of any given patient were included in the training or test set only and never in both simultaneously.

For the computational experiments involving the deep learning techniques the image training set was further subdivided into the 5880 training images as such and 2520 validation ones. Again, these particular datasets were well balanced containing exactly 50% of the norm and 50% of images representing tumor regions. The random sub-sampling was preferred on all the occasions where possible.

C. Conventional methods

As usual, the conventional method of binary classification task considered in this study included feature extraction and classification steps. The feature extraction was performed based on color co-occurrence matrices [7]. Given that the hematoxylin-eosin stained histological images are reasonably poor in colors, the original RGB color space was reduced down to the palette of the most common 32 colors. This was accomplished with the help of an adaptive algorithm of reducing color space based on *k*-means clustering as implemented in commonly-known Python PIL library. The inter-pixel spacings were selected to be 1, 2, and 4 pixels. As a result, the co-occurrence image descriptors had the form of 3D arrays of “colors-colors-spacing” type with a dimensionality of 32×32×3.

It is known (and it is very natural) that elements of co-occurrence matrices are highly correlated and therefore they are too redundant to be utilized as image features directly. For this reason, their principal components (PCs) derived with the help of the Principal Component Analysis (PCA) method were used instead. The advantage of such features is that they are compact, linear, and mutually uncorrelated.

The image features were inputted into the Support Vector Machine (SVM) and Random Forests (RF) classifiers. These classifiers were selected because they typically provide competitive results and their software implementations are available broadly. The relatively low computational expenses required by the classifiers allow to subsample given amount of image data from the whole dataset of 8400 training images and repeat the subsampling-training-prediction loop 100 times for obtaining reliable estimates of classification accuracy.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

D. CNN-based methods

In this study, the simple and well-known CNN architecture was employed that consisted of only 3 convolutional + *MaxPooling* layers with 16, 32, and 64 filters respectively (Fig. 3). This is because the assessment of the use of limited image training sets was particularly targeted in this study. In addition, such a decision enables other researchers to easily reproduce results of the computational experiments reported in this paper. The convolutional part of CNN was followed by a fully connected neural network with 512 intermediate nodes.

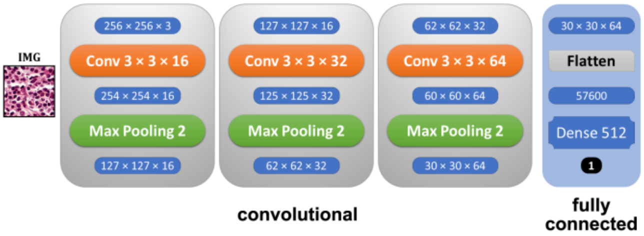


Fig 3. The CNN architecture being employed

Note that despite the simple architecture, under condition of 256×256 pixels of input image size the CNN contains 29,515,809 trainable parameters (weights). It was found that the use of *GlobalMaxPooling* instead of a straightforward *Flattening* of the output of the convolutional part of neural network reduces the number of parameters down to 57,377 with no accountable reduction of the classification accuracy.

E. Experimental arrangements

In order to obtain a relatively complete picture of the influence of training set size on the classification results, the original image training set was gradually reduced from 8400 to 840 images with the step of 10%. In addition, the very-small sub-samples of 5% (420), 2.5% (210), 1.25% (105), and even 1% (84) of original image dataset were also examined where possible. In all the occasions the test set was kept exactly the same and consisted of 3600 images.

The pipeline of computational experiments included the major steps given below.

(a) Initial preparations. They included converting original color RGB images into the reduced paletted representation with 32 colors as well as calculation of color co-occurrence matrices for each of 12000 images.

(b) Creating a data table by way of storing vectorized versions of co-occurrence matrices into 12000 different rows. Performing PCA on the resultant data table for obtaining a concise feature representation of every image involved in the experiments.

(c) Splitting the whole image dataset into the train and test subsets by 70/30 rule. In the case of CNN-based classification (not applicable to conventional SVM and RF) the train set was further split by the same rule into the part used for training as such and the validation.

(d) Cary out the conventional part of classification experiments in 14 steps by way of step-by-step reduction of training set size from the original 8400 images down to 84 ones as described above. At every classification step except for the first one, the training+prediction procedure repeated 100 times on varying image training sets obtained by a random sub-sampling from the original 8400 items. As a result, the total number of training and prediction steps was amounted up to 13*100+1 = 1301. This was to account for the inhomogeneity of original image classes as well as for the variability of images the training set is made of.

(e) The CNN-based experiments were done in a similar manner. However, in this case, one more key parameter came into the way what is the number of training epochs that need to be performed. Also, due to known fluctuations of the training process, the exact measurement of classification accuracy often includes performing a safe, i.e., over-rated number of epochs in order to identify the best one. There are some more control parameters such as random seeds different values of which may lead to slightly different results. These parameters increase potential computational expenses even further. For estimation purposes let us simply suppose that we need only 10 additional exploratory runs due to these factors specific for CNNs. Then the number of repetition loops of type {sub-sampling} {training} {prediction} {adjusting control parameters} increases up to approximately 13,000 what is going beyond the reason.

Thus, in order to make conventional and CNN-based results comparable, at each step of experiments we selected the training dataset that provided the best classification accuracy by SVM classifier and repeat it on exactly the same set of training and test images using CNNs.

III. RESULTS

Results of classification experiments are given below and itemized in the same way as their description presented in the previous section.

(a) Color co-occurrence matrices were computed using a fast algorithm based on indexing arrays that implemented in R language [8]. The elements below the leading diagonal of square-shaped Color-Color slices of resultant 3D arrays were summed up to the corresponding elements situated above the leading diagonal to avoid dependence of results on the rotation and reflection of original images as described in [7].

(b) The image features, i.e., principal components were selected using values of 0.9, 0.95, and 0.98 as thresholds for cumulative variance. These resulted in 21, 221, and 370 components respectively. The value of 0.95 was finally selected as the basic and used in all the computational experiments.

(c) In CNN-based classification with 14 different training set sizes the images for training and validation were selected at random by 70/30 proportion. The amounts of norm and tumor images were kept equivalent in both.

(d) Results of image classification using conventional methods are summarized in Fig. 4. As it can be seen from the top panel of Fig. 4, the difference between the mean accuracy values provided by SVM and RF classifiers is reasonably low with a maximum mutual deviation of about 1%.

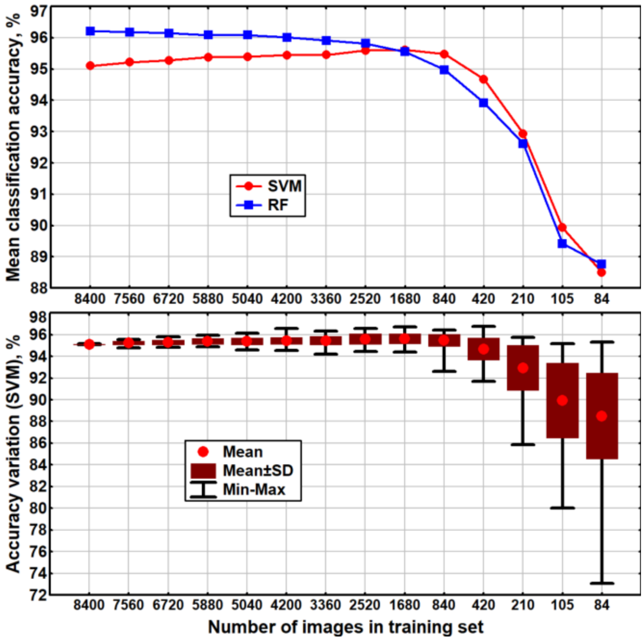


Fig. 4. The mean classification accuracy achieved across 100 replications by SVM and RF methods (top) and its variation in case of SVM (bottom)

Next, the shape of plots suggests that the mean accuracy is keeping almost constant for all training set sizes reduced from 8400 down to 840 images. Then it drops quickly from 95.1% (SVM) and 96.2% (RF) down to 88.5% and 88.8% respectively when reaches the smallest training set of 84 images.

Interestingly, the most fortunate combination of 84 training images among 100 randomly sampled ones for SVM-based classification provided 95.3% of the classification accuracy on the balanced test set consisting of 3600 images. This is slightly better than the worst results on all the repetitions of experiments and all tested training set sizes (see corresponding whiskers of the box-and-whiskers plots of Fig. 4).

The pattern of variability of classification results is somewhat more interesting (see SVM as an example on the bottom panel of Fig. 4). While the training sets remain relatively large, the standard deviation keeping small and ranged from STD=0.172% for 7560 images and going up to STD=0.545% for 840 images. Then it increases significantly and achieves STD=3.946% in the case of 84 images randomly chosen for training. The extreme values ranged even more substantially. For instance, in case of SVM and 84 training images, the classification accuracy varied in 100 repetitions from 73.1% to 95.3%. The described behavior can be explained by the following two reasons:

- the large portions of training images represent better the whole population (general regularity),
- the histological images used in this study are very heterogeneous (see Fig. 2) and vary significantly depending on the patient, biopsy techniques, sample preparation and staining protocols, image acquisition devices used in different hospitals, and some other factors.

(e) Results of CNN-based classification are given in Fig. 5. From a first glance, it becoming clear that results produced by CNN are comparable with the ones obtained using color co-occurrence features followed by SVM and RF (see bars for 8400, 420, and 840 images in the training set). However, once the training set is reduced further, the popular nowadays DL-based approach starts to lose completely against classical methods. This is especially obvious when the CNN results are compared to the ones produced by SVM. For making this fact easier to capture, the bottom panel of Fig. 5 provides two plots that compare results produced by CNN and the maximum accuracy achieved either by SVM or RF classifiers for each step of the experiments.

It is clear that due to the low computational expenses both of them can be comfortably run in parallel and the best result can be taken as the final solution. Note that such results are not surprising at all because it is commonly known that CNNs are hardly usable in the circumstances when only a few tens or hundreds of images are available for training (the use of possible benefits provided by augmentation and other similar techniques should be discussed separately).

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

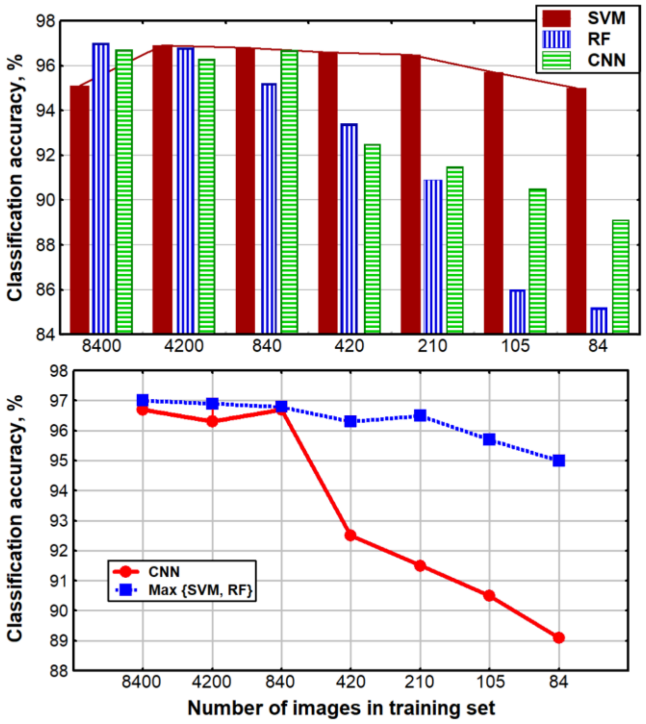


Fig. 5. The best classification accuracy achieved on large (8400–840) and small (420–84) training images by SVM, RF, and CNN methods

VI. CONCLUSIONS

Results reported with this study allow drawing the following conclusions.

(1) Conventional and CNN-based methods produce similar classification accuracy on relatively large training sets (from 840 to 8400 images). However, on smaller training sets containing 84-420 images, conventional methods reliably outperform the results demonstrated by CNN.

(2) Under the condition of the high variability of the content of original images and small training sets the classification results may vary substantially depending on the images used for training. For instance, in this particular study, the classification accuracy varied in a wide range from 73.1% to 95.3%. In the case of conventional methods, this problem can be resolved by multiple re-sampling training images and re-running the training for obtaining a reliable estimate of the accuracy. However, with CNNs such a solution can be not feasible due to much higher computational expenses.

(3) The use of recent large and heavy CNN architectures with small image datasets is questionable. However, a separate investigation is necessary for quantitative assessment.

ACKNOWLEDGMENT

The author wishes to thank his employer for providing all the necessary conditions for conducting this exploratory research work.

REFERENCES

[1] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, R. M. Summers. “A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies with Progress Highlights, and Future Promises,” in Proceedings of the IEEE, vol. 109, no. 5, pp. 820-838, May 2021, doi: 10.1109/JPROC.2021.3054390.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, C. I. Sanchez, “A survey on deep learning in medical image analysis.” Medical Image Analysis, vol. 42, pp. 2017.

[3] F. Aeffner, M.D. Zarella, N. Buchbinder, M. M. Bui, M. R. Goodman et al. “Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology association.” Journal of Pathology Informatics, vol. 10, no. 9, 2019.

[4] M. Veta, J. Y. Heng, N. Stathonikos, E. B. Bejnordi, F. Beca, et al. “Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge,” Medical Image Analysis, vol. 54, pp. 111-121, 2019.

[5] B. E. Bejnordi, M. Veta, P. J. van Diest, B. van Ginneken, N. Karssemeijer, et al. “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” Journal of the American Medical Association, vol. 318, no. 22, pp. 2199-2210, 2017.

[6] H. R. Tizhoosh, L. Pantanowitz. Artificial intelligence and digital pathology: Challenges and opportunities. Journal of Pathology Informatics, vol 9, no. 38, 2018.

[7] V. Kovalev and S. Volmer. ”Color co-occurrence descriptors for querying-by-example,” in International Conference on Multimedia Modelling, Oct. 12-15, Lausanne, Switzerland, IEEE Comp. Society Press, pp. 32-38, 1998.

[8] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Weather Recognition based on Attention Image Search Method

Yue Zhang[1,2], ChengZhang[1,2], ShichaoKan[1,2], YigangCen[1,2], LinnaZhang[3]
1 Institute of Information Science, Beijing Jiaotong University, Beijing, China
2 Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing, China
3 School of Mechanical Engineering, Guizhou University, Guiyang, China
{17112065,19125149,16112062,ygcen}@bjtu.edu.cn, zln770808@163.com

***Abstract*. Weather monitoring plays a vital role in intelligent traffic transportation, and the improvement of weather recognition accuracy can effectively improve driving safety. At present, classification-based and segmentation-based algorithms for weather recognition have achieved good performance, but it is still full of challenges in real applications. On the one hand, the number of classes in public data sets is insufficient, which cannot identify the conditions such as stagnant water and debris flow. On the other hand, the current weather recognition methods have poor generalization ability, the model needs to be retrained when classes are changed. In this paper, we first propose a new multi-traffic weather (MTW) data set for weather recognition, it contains much richer classes. Then, a new weather recognition method based on attention image retrieval (AIR) is proposed to improve the performance of recognition. Compared with the previous methods, our method can obtain better generalization performance.**

***Keywords:* weather recognition, image retrieval, attention**

I. INTRODUCTION

Weather recognition in natural scenes plays an important role in the field of intelligent traffic. Especially bad weather will not only greatly weaken the efficiency of transportation, but also directly cause traffic accidents, endangering people's lives and property. Different from general object recognition [1], [2], weather recognition is to recognize the entire image and requires an understanding of complex phenomena, such as light and reflection on the surface of the object. The current system relies on either a series of expensive sensors or on manual assistance to identify weather conditions. Since the weather condition varies locally from place to place in the same region, expensive sensors and a lack of human assistance limit the availability of local weather measurements.

Recently, weather recognition method based on computer vision has become more and more popular. The traditional weather image recognition method uses histogram features for representation [3], [4] and applies support vector machine (SVM) method for classification [5], [6]. With the development of deep learning, the accuracy of weather recognition based on deep learning has also been greatly improved, Cewu *et al.* [6] combined the data-driven CNN feature and well-chosen weather independent features to train a latent SVM classifier. This method is insensitive to global intensity transfer. Elhoseiny [7] focused on studying the feature spaces in the weather classification task. Lin *et al.* [8] proposed a region selection and concurrency model (RSCM) to help discover regional properties and concurrency for multi-class weather recognition. However, these methods need to retrain the model when adding or deleting some classes for different regions. Besides, the recognition accuracy for specific scenarios needs to be further improved to meet the requirements of actual traffic road weather recognition.

Existing datasets (e.g. TWI [6] and MWD [8]) include different scenes and can not identify some conditions of traffic jams caused by bad weather (i.e. stagnant water and debris flow). Thus, we propose a finer traffic scene weather data set called the multiple traffic weather (MTW) data set. Compared with the previous data sets, our data set contains more classes (11 classes) with 2,444 images, including sunny (norm), rainy (heavy rain, light rain), snowy (heavy snow, light snow), and haze (mist and dense fog), others(flood, debris flow, stgnant water, and landslide).

In order to meet the requirements of real scenarios, in addition to proposing more comprehensive data sets, it is necessary to build a more accurate and flexible recognition algorithm. The image recognition based on image retrieval method has been proved to be very effective in face recognition [2] and person re-identification [10]. Image retrieval method can effectively calculate the similarity between query image features and image features of the prior images, and further obtain output results according to the label of the nearest prior image. Concretely, we propose a weather recognition method based on attention image retrieval (AIR), which combines the ResNet network structure and the attention mechanism as the encoder to learn the features of each class. Then the classification loss and metric loss are employed to update the weight of the encoder. In the testing phase, we adopt the encoder to extract features and compare them with the features of the prior images to determine the weather class of the input image. Compared with the previous methods based on classification or segmentation, our method not only has a better recognition rate, but also has a stronger generalization ability such that new categories can be added without retraining the

---

This work was supported by the Fundamental Research Funds for the Central Universities under Grant 2021QY002, in part by the National Natural Science Foundation of China under Grant 61872034, 62062021 and 62011530042, in part by the Beijing Municipal Natural Science Foundation under Grant 4202055, in part by the Natural Science Foundation of Guizhou Province under Grant [2019]1064. (Corresponding author: Yigang Cen.)

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

model.

The main contributions of this paper are as follows. We propose a new open-source traffic road weather data set called the multi-traffic weather (MTW), which contains richer and finer classes (11 classes).We propose a new weather recognition method based on attention image retrieval (AIR), and demonstrate its superiority on the public weather data set and our MTW data set.

II. WEATHER RECOGNITION DATA SET

In this section, we first revisit some existed weather recognition data sets, Martin *et al.* [3] developed a database with three classes, i.e. clear, light rain, and heavy rain. Then Cewu *et al.* [6] proposed two-classes weather images (TWI) data set, which includes 5,000 sunny and 5,000 cloudy images. Lin *et al.* [8] constructed a multiclass weather data set (MWD) with 65,000 images from 6 common weather categories, i.e., sunny, cloudy, rainy, snowy, haze, and thunder. In addition, Villarreal *et al.* [9] proposed a rain, fog, snow (RFS) data set collected online that depict scenes of three weather conditions, and each condition contains 1100 images. A comparison of weather images from different data sets shows that current data sets are limited in two ways: (1) The number of weather class is not enough. Moreover, the classification is not detailed enough. For example, dense fog and mist have different effects on traffic, which should be identified and given different warnings. (2) Only some common weather categories are included, which ignores road safety problems (i.e. stagnant water and debris flow) caused by bad weather.

In order to reduce the impact of bad weather on traffic, we construct a new open-source data set called the multi-traffic weather (MTW) data set. In our daily life, common weather conditions include sunny (norm), rainy (heavy rain, light rain), snowy (heavy snow, light snow), and haze (mist, dense fog). Common natural disasters that affect traffic include flood, stagnant water, debris flow, and landslide. By using the web crawler, we collected a total of 11 categories and 500 images for each class. After our screening, there were finally $3,254$ images left, including $2,444$ images for the training set, and 810 images for the test set. As shown in Table. I, we split the data set into the training set and testing set for each class. The ratio of the training set and the testing set is about $3:1$.

III. METHODOLOGY

A. Overall Network Structure

We propose a novel weather recognition framework based on the attention image retrieval method (AIR). In the training phase, as shown in Fig. 1, it includes an encoder and two losses. The images are input into the backbone network to extract the features. Due to the large difference in data scenes collected, the network needs to extract some features required to exist in a specific class. Therefore, before the global average pooling (GAP) of the benchmark network, we proposed a weather attention (WA) module to make the network pay more attention to the features of the weather. Then the obtained features are divided into two branches, one branch takes the classification loss, namely the cross-entropy loss. Another branch takes the metric loss. Here, we introduce the multi-similarity loss, which can effectively learn the relationship between positive and negative examples such that the network can learn robust features. In the test phase, the features obtained by the benchmark network are used to calculate the similarity with the prior images, and the class of the image is determined according to the class of its nearest neighbor in prior images.

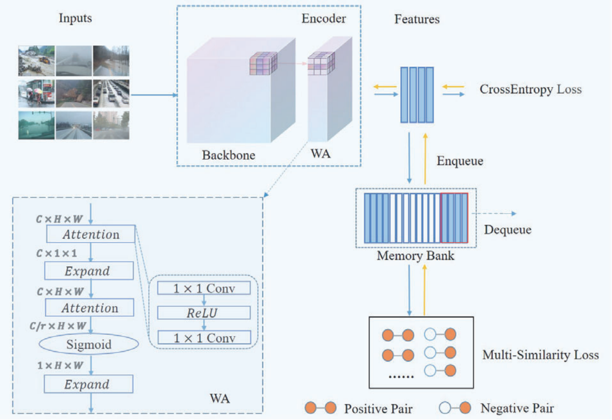


Fig. 1. The overall architecture of the training network, it includes an encoder and two loss functions

B. Weather Attention Module

In this work, in order to make the network focused on weather features, we propose a weather attention (WA) module, which considers channel attention (global feature) and spatial attention (local features) information. Specifically, our channel attention adopt an 1×1Conv-ReLU-1×1 Conv operation as the SENet [11]. The channel attention maps are dotted with the feature maps, their results are input into 1×1Conv-ReLU-1×1 Conv to reduce the channel dimensions. In Fig. 1, according to the enhancement of channel attention and spatial attention, the network can filter out some weather independent features that can improve the accuracy of weather recognition.

C. Loss Functions

As shown in Fig. 1, we extract feature maps from the backbone and WA module, then a softmax function is used to predict the classes of the input image features. The cross-entropy loss is employed to compute loss between predict label $\hat{y}$ and real label y, it is computed as:

$$L_{cls} = -\left[y log\hat{y} + (1-y)\, log\, (1-\hat{y})\right]. \tag{1}$$

In addition, we introduce a pair-based metric loss function, the multi-similarity (MS) loss [12], to realize more efficient sample training through two iterations of sampling and weighting. For each sample pair, we need to consider not only the self-similarity of the sample pair itself, but also its relative similarity with other sample pairs. Therefore,

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

TABLE I
PROPOSED MULTI-TRAFFIC WEATHER (MTW) DATA SET, IT MAINLY INCLUDES FIVE CATEGORIES, INCLUDING RAIN, SNOW, HAZE, NORM, AND OTHERS. OTHERS REFER TO CATEGORIES THAT ARE NOT COMMONLY USED. IN ADDITION TO THE NORM CLASS, EACH CLASS CAN CONTINUE TO BE FINELY DIVIDED INTO DIFFERENT CLASSES.

	snow		haze		rain		others				norm
class	light snow	heavy snow	mist	dense fog	light rain	heavy rain	stgnant water	debris flow	flood	landslide	norm
train	146	333	136	34	269	134	436	154	355	149	398
test	48	111	44	11	89	44	144	51	118	50	100

TABLE II
WEATHER IDENTIFICATION RESULTS BASED ON DIFFERENT BENCHMARK NETWORK STRUCTURES ON THE PROPOSED MTW DATA SETS.

Methods		Accuracy(%)
Backbone	ResNet18	75.2
	ResNet34	75.6
	ResNet50	**76.5**
	ResNet101	76.3
Component analysis	baseline	76.5
	baseline+MS	77.0
	baseline+MS+Attention	**77.5**
Our AIR method	classification	77.5
	image search	**79.9**

important sample pairs can be adopted and weighted more efficiently and accurately. The MS loss is formulated as:

$$L_{MS}=\frac{1}{m}\sum_{i=1}^{m}\left\{\frac{1}{\alpha}log\left[1+\sum_{k\in P_i}e^{-\alpha(S_{ik}-\lambda)}\right]\right. \\ \left.+\frac{1}{\beta}log\left[1+\sum_{k\in N_i}e^{\beta(S_{ik}-\lambda)}\right]\right\} \qquad (2)$$

P_i reprsents a positive sample, and N_i represents a negative sample. m represents the sample number. λ, α, β are fixed hyper-parameters. S_{ik} represents the similarity of two features. Specifically, $S_{ik} :=< f(x_i;\theta), f(x_k;\theta) >$, where $< \cdot >$ denotes dot product, which results in an $m\times m$ similarity matrix S with the element at (i,k) is S_{ik}.

IV. EXPERIMENTAL RESULTS

A. Implementation Details and Evaluation Protocol

We perform our experiment on one 1080Ti GPU and PyTorch 1.1.0 platform. We adopt an ImageNet-pre-trained ResNet-X as the backbone, Adam is used with a learning rate 0.000003, and weight decays $5e^{-4}$. Each training mini-batch contains 10 images of 5 instances. For the k nearest neighbor (KNN), we choose k=1, which means the label of the nearest neighbor is the class of the tested image.

B. Comparison With the State-of-the-Art Methods

In this paper, for our proposed MTW data set, we first conduct some experiments on different backbone networks. The results are shown in Table. II. The recognition accuracy of the mentioned data set increases as the network deepens. But on the ResNet50 network framework, the result achieves the highest value of 76.5%. So for subsequent improvements, we will adopt ResNet50 as our baseline.

TABLE III
PERFORMANCE COMPARISON WITH THE STATE-OF-THE-ART METHODS ON TWI DATA SET.

Method		Accuracy (%)
Yan *et al.* [13]	ISNN(2009)	27.2
Roser *et al.* [14]	IV (2008)	28.5
Lalonde *et al.* [15]	IJCV (2012)	41.8
Lu *et al.* [6]	TPAMI (2017)	55.3
Elhoseiny *et al.* [7]	ICIP (2015)	92.9
SE-ResNet101 [11]	CVPR (2018)	94.5
CBAM-ResNet101 [16]	ECCV(2018)	94.6
BnInception	-	89.8
ResNet18	-	93.8
ResNet34	-	94.6
ResNet50	-	**94.7**
Our AIR method	-	**95.6**

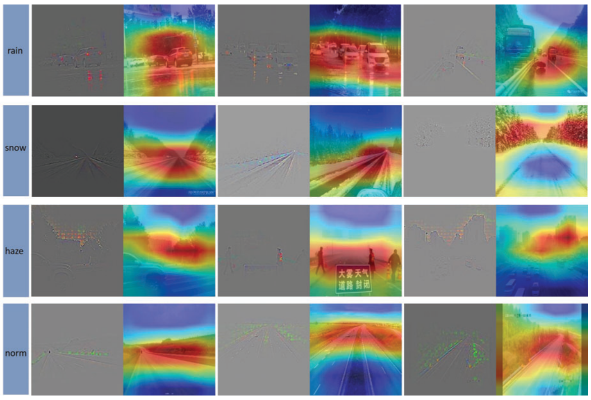


Fig. 2. Heat maps of attention of the proposed models in the proposed common weather categories. Each row represents a class, for each image, we show the GAP map (left) and Cam ++ heat map (right), respectively

Then we conduct a series of experiments and analyze the effectiveness of our proposed method in Table. II. Here, the backbone network is ResNet50, and the cross-entropy loss function is used to calculate the classification loss. The recognition accuracy rate is 76.5%. When different modules are added, the accuracy will be improved correspondingly. From Table. II, we can see that search recognition accuracy reaches 79.9%, which is improved 2.4% compared with the classification.

In order to verify the effectiveness of our proposed method, we test it on the TWI data set and compare it with some state-of-the-art recognition methods, the results are shown in Table. III. Here, we choose the highest results of the previous methods to compare with our result. We also conducted

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

TABLE IV
THE PROPOSED MTW DATA SET IS DIVIDED INTO DIFFERENT SUBSETS, AND THE SUBSETS ARE TRAINED TO TEST THE ACCURACY, RESPECTIVELY.

dataset	the number of the class	classes	Recall@1(%)	Accuracy(%)
CC	4	rain snow haze norm	93.5	93.1
CF	7	heavyrain lightrain heavysnow lightsnow densefog mist norm	85.9	87.0
UF	5	debris-flow flood landslide stgnant-water norm	86.6	87.8

some experiments on different ResNet networks, including two networks with attention mechanisms (i.e. SE and CBAM). From Table. III, we can see that the ResNet101 network structure combines with different attention mechanisms have achieved good performance. And ResNet50 achieved a better classification result than other ResNet networks. Our AIR method achieves the highest accuracy, which is superior to the current best results of about 0.9%. It demonstrates the effectiveness of the proposed AIR method.

C. Discussion

In addition, the proposed MTW data set is divided into different subsets: common coarse classification (CC), common fine classification (CF), and uncommon fine classification (UF), and the subsets are trained to test the accuracy, respectively. As shown in the Table.IV, in the CC subset, there exists four classes: norm, rain, snow, and haze. Experimental results show that the model can still get considerable performance for classes that have never been trained. In addition, we draw the GAP (global average pooling) map and Grad-CAM++ map of common weather classes images, respectively. As shown in Fig. 2, the first row shows images in the rain class, and the model focuses on the foreground and where the road reflects light. The second row shows the images in the snow class, for which the model focuses on the road-side position. In the third row and the fourth row, the model focuses on the upper part of the road, the attention region includes the sky. This demonstrates that our model will pay attention to different information, so as to distinguish different categories.

V. CONCLUSIONS

In this paper, we first proposed a multi-traffic weather data set, which mainly concentrates on traffic road scenarios and contains richer and finer classes. It enables new experiments both for training better models and as a new benchmark. Then, we proposed a new weather recognition method based on the attention image retrieval (AIR) method. It effectively improves the accuracy of weather recognition, and can be widely used in different weather recognition scenarios. In other words, it is well extensible and does not require retraining the model when adding or subtracting some classes. Our future work will continuously improving and supplementing our data set. In addition, we need to further explore the solution of some images containing multiple classes.

REFERENCES

[1] Kan, S., Cen, L., Zheng, X., Cen, Y., Zhu, Z., Wang, H. : A supervised learning to index model for approximate nearest neighbor image retrieval. Signal Processing: Image Communication, 78, 494–502 (2019).

[2] Li, S., Liu, B., Chen, D., Chu, Q., Yuan, L., Yu, N. : Density-Aware Graph for Deep Semi-Supervised Visual Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13400-13409 (2020).

[3] Roser, M., Moosmann, F. : Classification of weather situations on single color images. In 2008 IEEE Intelligent Vehicles Symposium, pp. 798-803 (2008, June). IEEE.

[4] Yan, X., Luo, Y., Zheng, X. : Weather recognition based on images captured by vision system in vehicle. In International Symposium on Neural Networks,pp. 390-398 (2009, May). Springer, Berlin, Heidelberg.

[5] Chen, Z., Yang, F., Lindner, A., Barrenetxea, G., Vetterli, M. : How is the weather: Automatic inference from images. In 2012 19th IEEE International Conference on Image Processing pp. 1853-1856 (2012, September). IEEE.

[6] Lu, C., Lin, D., Jia, J., Tang, C. K. : Two-class weather classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3718-3725 (2014).

[7] Elhoseiny, M., Huang, S., Elgammal, A. : Weather classification with deep convolutional neural networks. In 2015 IEEE International Conference on Image Processing (ICIP), pp. 3349-3353 (2015, September). IEEE.

[8] Lin, D., Lu, C., Huang, H., Jia, J. : RSCM: Region selection and concurrency model for multi-class weather recognition. IEEE Transactions on Image Processing, **26**(9), 4154-4167 (2017).

[9] Guerra, J. C. V., Khanam, Z., Ehsan, S., Stolkin, R., McDonald-Maier, K. : Weather Classification: A new multi-class data set, data augmentation approach and comprehensive evaluations of Convolutional Neural Networks. In 2018 NASA/ESA Conference on Adaptive Hardware and Systems (AHS), pp. 305-310 (2018, August). IEEE.

[10] Zhang, Y., Jin, Y., Chen, J., Kan, S., Cen, Y., Cao, Q. : PGAN: Part-based nondirect coupling embedded GAN for person reidentification. IEEE MultiMedia, **27**(3), 23-33 (2020).

[11] Hu, J., Shen, L., Sun, G. : Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132-7141 (2018).

[12] Wang, X., Han, X., Huang, W., Dong, D., Scott, M. R. : Multi-similarity loss with general pair weighting for deep metric learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5022-5030 (2019).

[13] Yan, X., Luo, Y., Zheng, X. : Weather recognition based on images captured by vision system in vehicle. In International Symposium on Neural Networks, pp. 390-398 (2009, May). Springer, Berlin, Heidelberg.

[14] Roser, M., Moosmann, F. : Classification of weather situations on single color images. In 2008 IEEE Intelligent Vehicles Symposium, pp. 798-803 (2008, June). IEEE.

[15] Lalonde, J. F., Efros, A. A., Narasimhan, S. G. : Estimating the natural illumination conditions from a single outdoor image. International Journal of Computer Vision, 98(2), 123-145 (2012).

[16] Woo, S., Park, J., Lee, J. Y., Kweon, I. S. : Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), pp. 3-19 (2018).

[17] Chattopadhay, A., Sarkar, A., Howlader, P., Balasubramanian, V. N.: Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) pp. 839-847 (2018, March). IEEE.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Automation of the Study of Radiologically Isolated Syndrome in Multiple Sclerosis

Ivan Kosik
FAMCS of
Belarusian State University
LICT of Belarusian State
Medical University
Minsk, Belarus
ivankosik91@gmail.com

Alexander Nedzved
FAMCS of
Belarusian State University
UIIP of NAS Belarus
Minsk, Belarus
NedzvedA@tut.by
ORCID 0000-0001-6367-5900

Ryhor Karapetsian
Laboratory of Information and
Computer Technology of
Belarusian State Medical
University
Minsk, Belarus
greg.itlab@gmail.com

Vera Yashina
Federal Research Center
“Computer Science and
Control” of RAS
Moscow, Russia
werayashina@gmail.com

Igor Gurevich
Federal Research Center
“Computer Science and
Control” of RAS
Moscow, Russia
igourevi@ccas.ru

Abstract. **In this paper the UNet 3+ model is used for detection regions of multiple sclerosis on radiological images. For increase quality the specific image preprocessing improves quality of dataset and results of detection. The proposed solution for the automatic identification of pathological areas using artificial neural networks has significantly increased the speed of analyzing the state of the pathological pattern.**

Keywords: **medical image analysis, UNet 3+, regions detection, segmentation, dataset preprocessing**

I. INTRODUCTION

Multiple sclerosis is a serious disease of the central nervous system that leads to disability among people, including young people of working age. Insufficient knowledge of the pathogenesis of this disease and an increase in the frequency of its occurrence require the intensification of studies of this pathology [1].

The most important symptom of multiple sclerosis is a focal lesion of the central nervous system caused by auto aggression against myelin proteins in the brain and spinal cord. In this case sites of demyelination occur and sclerotic plaques appear in the small veins of the brain, in the cerebellum, in the spinal cord, in the optic and other cranial nerves. They are an important diagnostic feature of multiple sclerosis.

Magnetic resonance imaging (MRI) is the most effective tool for visualization of demyelination sites nowadays. However, for an objective description of the state and dynamics of the pathological process, visualization of foci on MRI sections should be supplemented with data on the size, intensity and localization [2].

Detection of lesions is the first step towards obtaining additional quantitative information from MRI images. In most cases, their boundaries are blurred and have poor contrast against the background of the brain tissue. Therefore, the detection of pathological sites is the most time consuming and at the same time the most important stage. The accuracy of the diagnostics largely depends on the accuracy of the detection. Meanwhile, in medical practice, segmentation of foci is carried out either by manual methods (contouring), or by semi-automatic methods (for example, by algorithms of area growth or "smart brush"). The diagnostician must first visually assess the information content of the image area in terms of the level of brightness and localization, and then carry out the recognition procedure. Considering that the radiologist has to analyze at least 120 slices (50 + 40 + 30 in three orthogonal projections), it is easy to understand how difficult this work is. This affects the accuracy of the results, especially when examining multifocal patterns [3].

The use of advanced methods of volumetric reconstruction of MRI images, based on fully automatic segmentation of informative objects by neural networks, contributes to an increase in the productivity and accuracy of the study.

II. PREPARATION OF TRAINING IMAGES

We used MRI-series of 50 patients obtained in different modes (T1, T2, Flair, etc.) for training the neural network. In addition, data augmentation was

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

used. It is the technique for generation additional training data from the initially available initial set of images [4,5]. In our case, the initial training set consisted of 455 3D-series. It corresponds to different modes and different studies of 50 patients of the 9th City Clinical Hospital of Minsk.

For the augmentation the Albumentations library was used. We chose the following parameters for augmentation:

- rotations up to 20 degrees;
- shift;
- scaling;
- horizontal reflection;
- vertical reflection;
- sharpening;
- spectrum change;
- optical distortion.

To increase the volume of initial data, we used methods based on geometric and brightness variability for informative objects in the original images. Methods of the first type increas the data volume by reorienting and scaling the available input images. This allowed the network to learn invariance to this kind of distortion, even if the distortion was absent in the original images.

III. DEFINITION OF CNN MODEL

For the most complete automation of the procedure for segmentation of pathological areas on MRI scans, a selection of the architecture of a convolutional neural network (CNN) and a strategy for its training was performed, which would allow obtaining a high return with a limited initial sample. The learning strategy is based on the most complete extraction and effective generalization of meaningful information concentrated in the initial data. The effect is achieved due to the automatic generation of additional data from this data. The network architecture consists of a tapered section for capturing context, and a symmetrical expanding section for more accurate localization and contouring of objects. This network organization is called U-Net 3+ [6] like as Fig 1.

The UNet 3+ gives simplified overviews of UNet, UNet++. Compared with UNet and UNet++, UNet 3+ combines the multi-scale features by re-designing skip connections as well as utilizing a full-scale deep supervision, which provides fewer parameters but yields a more accurate position-aware and boundary-enhanced segmentation map [7, 8].

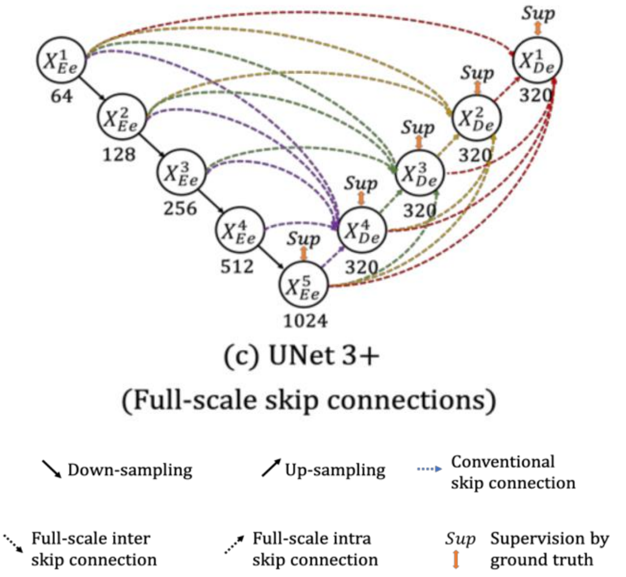


Fig .1. The model of UNet 3+ [1]. The depth of each node is presented below the circle

IV. NEURAL NETWORK TRAINING

The series were divided into training set and validation one in a ratio of 75% to 25%, respectively. The formation of sets of training and validation images by series allows you to include all MRI slices of the series in only one sample (training or validation). This is justified by the fact that similar regions belonging to neighboring slices will lead to overfitting if they get in different samples. An additional set of test images for a separate small group of patients is formed during the testing phase of the final version of the neural network.

The next step in the training procedure is to obtain training images (MRI slices) from the 3D series, divided into training set and validation one. Sections without pathological regions are not used in training, since preliminary tests showed that training on all sections gives worse results. Examples of the formation of training sets are given below. It is clear that the number of series selected for the same MRI scan mode exceeds the number of unique patients, since several series (taken at different times) can belong to the same patient.

Two neural networks were trained: one on MRI series obtained in the T2 TSE scan mode; the second on all images (all modes were used). The amount of training data is shown in Table I.

TABLE I. AMOUNT OF TRAINING DATA FOR EACH MODEL

Model	**Number of 3D-series**		**Number of slices**	
	Training	***Validation***	***Training***	***Validation***
Mode T2 TSE	119	40	1500	506
All MRI-modes	341	114	4789	1452

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

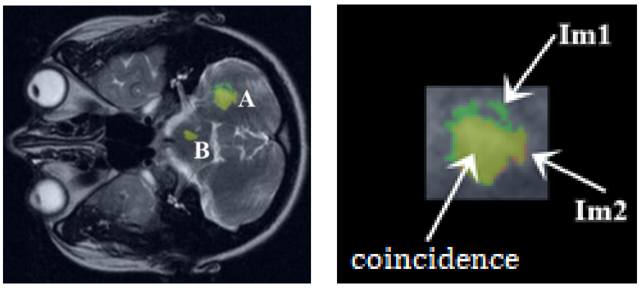
V. QUALITY ASSESSMENT OF NETWORK

To assess the quality, the Intersection Over Union metric (or Jaccard index) was used, given by the formula (1): (1):

$$IoU = \frac{TP}{TP+FP+FN} = \frac{Im_1 \cap Im_2}{Im_1 \cup Im_2} = \frac{Im_1 \cap Im_2}{Im_1 + Im_2 - Im_1 \cap Im_2}, \quad (1),$$

where *TP* (true positive) is selection of a pixel that actually belongs to the focus of demyelination; *FP* (false positive) is false selection of a pixel that does not belong to the focus; *FN* (false negative) is false marked pixel as not belonging to the focus; *Im1* is an area identified by an expert as a focus of demyelination; *Im2* is area identified by the neural network as a focus of demyelination.

The metric can take values from 0 to 1. It assesses how closely the area of the real focus Im1 (selected interactively by the expert) matches the area Im2, segmented automatically by the neural network. The higher the metric value, the greater the coincidence, and, therefore, the more reliable the neural network model works. In Figure 1, the same focus is highlighted by an expert (green) and a neural network (brown).



a) b)

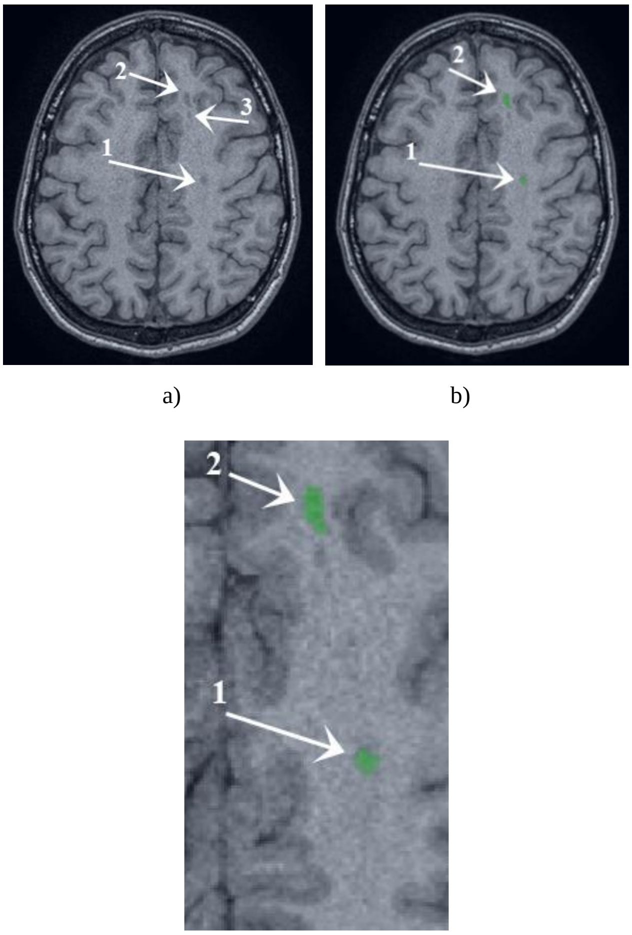
Fig. 2. Jind metric demonstration: total image of MRI-section, with highlighted foci A and B (a); fragment of the MRI image with focus A (b)

There is an overlap of areas of green and brown in Figure 1b. In Figure 1b, the overlapping areas of green and brown correspond to the area selected by both the expert and the neural network.

VI. DISCUSSION AND CONCLUSION

The best accuracy was obtained for the trained model with the input image size 352x352 and the batch-size 9. It is equal to 0.62 (IoU, Jaccard index). combination of the binary cross-entropy and the Sørensen loss function (Dice loss). This model was trained for 100 epochs using a GeForce GTX 1080 Ti graphics card. The training lasted 4 hours.

Examples of the results of segmentation of areas of destruction of the myelin sheath are shown in Fig. 3, 4. The color of individual areas corresponds to the color in Fig. 2: green corresponds to the selection of the focus by the expert, brown to the selection of the focus by the neural network. The overlapping areas of green and brown correspond to the area selected by both the expert and the neural network.



a) b)

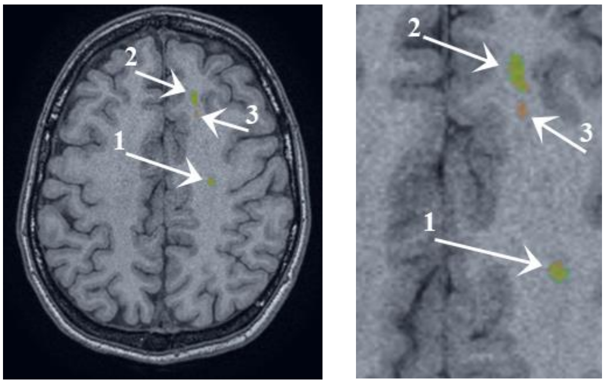
c)

Fig. 3. Myelin sheath lesions: original T1W image (a); areas identified by an expert (b); part of an enlarged image (c)

Fig. 4 shows that, the network found all the affected areas. For two of them, the highlighting coincided with the expert's opinion, one is highlighted contrary to the expert's opinion. In such cases, additional validation is required.

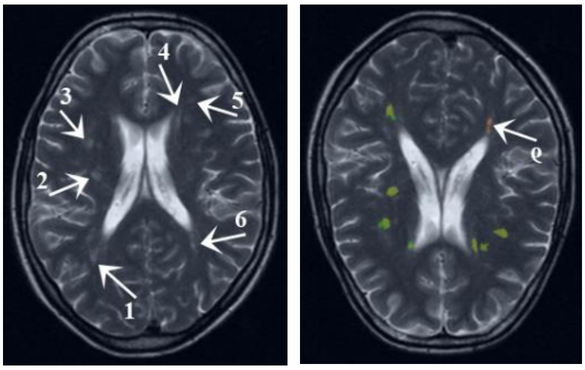
In Fig. 5, the network also highlighted all the foci (the areas highlighted by the expert and the network changed color when superpositioned). In addition, the network highlighted a problem area, which the expert did not mark as a focus of demyelination (it remained brown). In this case, additional validation of the results is required.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.



a) b)

Fig. 4. Myelin sheath lesions: areas identified by neural network (a); a enlarged part of the image with the result of comparing the lesions identified by the expert and the network (b)



a) b)

Fig. 5. Myelin sheath lesions: marked on the original T2W image (a); overlay of results of the selection by the expert and the neural network (b). The arrow indicates area 6, identified as an outbreak only by the network

In Fig. 6, the neural network did not identify area 3 as a focus of demyelination, but at the same time found areas 1, 2, 4, which for some reason were not noted by the expert. Additional analysis of the results is required, but neural network extraction looks more promising.

In general, it can be seen that the network successfully detects problem areas on MRI-scans obtained in different modes. The most accurate results are obtained with T2 TSE images. It should be noted that the sizes of the areas selected by the network do not always coincide exactly with the sizes of the same areas identified by the expert. This is due to the ambiguity of the focus boundaries, and the result of network segmentation is not always less accurate, especially when there are many lesions on different sections. Working under severe stress, the expert gets tired and over time his attention weakens.

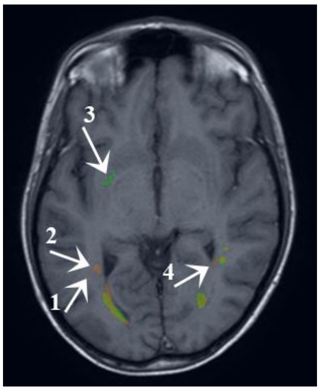


Fig. 6. Myelin sheath lesions: the result of the superposition of the results of the selection of foci by the expert and the neural network. The arrow indicates the areas of non-coincidence

The proposed solution for the automatic identification of pathological areas using artificial neural networks has significantly increased the speed of analyzing the state of the pathological pattern. Traditional manual isolation of demyelination lesions requires at least 65 minutes per patient, semi-automatic isolation takes about 23 minutes, whilea automatic segmentation is performed within 1 minute. In addition, automatic segmentation using a neural network allows to get hundred-percent repeatability of the analysis results and significantly reduce the workload for neurologists and radiologists.

ACKNOWLEDGMENT

This research was funded by RFBR and BRFBR, project 20-57-00025/BRFFI F20R-134 and partly supported under the project BRFFI F20KIGT-006.

REFERENCES

[1] C. M. Poser, V. V. Brinar, "Diagnostic criteria for multiple sclerosis: an historical review". Clinical Neurology and Neurosurgery, June 2004, 106 (3), pp. 147–58.

[2] Massimo Filippi and Maria A. Rocca MR Imaging of Multiple Sclerosis. Radiology, 2011, 259:3, pp.659-681.

[3] C. Shorten, T. M. Khoshgoftaar, A survey on Image Data Augmentation for Deep Learning, J Big Data 6, 2019, pp. 60.

[4] J. Nalepa, M. Marcinkiewicz, M. Kawulok, Data Augmentation for Brain-Tumor Segmentation: A Review. Frontiers in Computational Neuroscience, no. 83, 2019, pp. 1662-5188.

[5] Ankur Biswas, Paritosh Bhattacharya, Santi P. Maity & Rita Banik Data Augmentation for Improved Brain Tumor Segmentation, IETE J. of Research, 2021, pp. 0377-2063.

[6] Huang, Huimin, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Y. Iwamoto, Xianhua Han, Yen-Wei Chen and Jian Wu, “UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation.” ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp.1055-1059.

[7] J. Duan and X. Liu, "Online Monitoring of Green Pellet Size Distribution in Haze-Degraded Images Based on VGG16-LU-Net and Haze Judgment," in IEEE Transactions on Instrumentation and Measurement, vol. 70, 2021, pp. 1-16.

S. -T. Tran, C. -H. Cheng and D. -G. Liu, "A Multiple Layer U-Net, Un-Net, for Liver and Liver Tumor Segmentation in CT," in IEEE Access, vol. 9, 2021, pp. 3752-3764.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

DSDNet Neural Network for Shadow Detection from Urban Satellite Images

Oleg Naidovich
Belarusian State University
Minsk, Belarus
o.naidovich@gmail.com

Alexander Nedzved
Belarusian State University
Minsk, Belarus
nedzveda@gmail.com

Shiping Ye
Zhejiang Shuren University
Hangzhou, China
zjsruysp@163.com
ORCID 0000-0002-9771-7168

***Abstract.* Shadow detection is one of the fundamental and challenging tasks in the field of computer vision and image processing. The increase of computing power has enabled many deep learning approaches to solve this problem. In this article we consider a DSDNet neural network in order to detect shadows on the base of texture analysis of the shadow area and bright area of the urban area.**

***Keywords*: shadow detection, DSDNet, deep neural networks, segmentation, Satellite image**

I. INTRODUCTION

Shadow is an illumination phenomenon, which is caused by the occlusion of light by some object, resulting in color and intensity changes in the local surfaces. Knowing where the shadow is allows us to infer, for example, scene geometry, lighting direction, and camera parameters. However, the presence of a shadow can degrade the performance of many fundamental computer vision tasks such as semantic segmentation, object detection and visual tracking. Consequently, the shadow detection problem has been studied for many years and presents a severe problem among computer vision tasks.

A shadow appears when an object partially or completely obscures a direct light source. The structure of the shadow is strongly dependent on the features of the object such as geometry and height. That is why it is more efficient to conduct research on structured objects like buildings. Generally, shadows are classified into attached shadows and cast shadows.

- Cast shadows arise when a light source is obstructed by a part of the same or another object. Such shadows appear on knowledge with a flat roof (see Fig. 1).
- Attached shadows are the shadows that form on the objects themselves. They arise when the angle between a surface normal and a light source direction is obtuse. They usually appear on buildings with hipped roofs (see Fig. 2).

Knowing where the shadow is, allows us to infer, for example, scene geometry, lighting direction, and camera parameters. Moreover, deleting shadows from the image can be used to detect objects such as buildings, trees, roads, etc. In addition, after removing shadows the objects will be displayed more evidently, hence they can be correctly recognized.

The presence of a shadow can degrade the performance of many fundamental computer vision tasks such as semantic segmentation, object detection, and visual tracking. The main problem caused by shadows is the loss of information in the image. It can lead to distortion of various parameters derived from pixels, so areas cannot be interpreted correctly.

The shadow detection problem has been studied for many years and it is a major concern among computer vision tasks. There are a lot of methods for detecting shadows in images. The purpose of this paper is to figure out the workflow of the neural network DSDNet and compare the results of its work to the other methods.

II. OTHER APPROACHES

There are two groups of methods in detection shadows on images: traditional based methods and deep learning based methods.

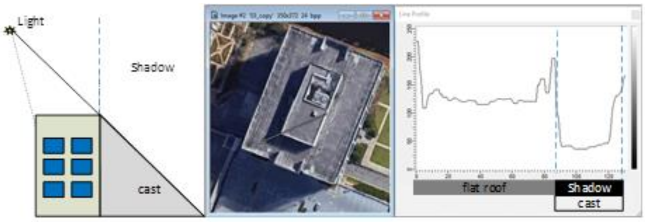


Fig. 1. Representation of the building with flat roofs

Fig. 2. Representation of the building with hipped roofs

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

A. *Traditional methods*

Traditionally, unique image shadow detection methods are based on exploitation of physical models of color and illumination. Other traditional methods use hand-crafted features which are based on marked shadow images. First of all, such methods describe image regions by feature descriptors and then categorize the regions into shadow and non-shadow regions. Such feature descriptors, for example, color, edges, and texture are frequently used in decision trees or SVM algorithms.

However, in satellite images, there are often non-shadow regions that emerge like shadows and are detected as shadows. And there are also shadow regions that emerge like non-shadow patterns and are detected as non-shadows as well. That is why these traditional predicated methods, which are based on color chromaticity and illumination, or use hand-craft features like illumination cues and color cannot deal with complex cases.

B. *Deep learning methods*

Due to the growth of computing power, it became possible to apply deep learning techniques to computer vision tasks. Recent state-of-the-art neural networks can be learned to detect shadows, which achieve significant performance improvements over the traditional ones.

For example, convolutional neural network is demonstrated to be a very powerful tool to learn features for detecting shadows, especially when large data is available. Also the generative adversarial networks (GANs) and recurrent neural networks (RNNs) adopted to detect shadows. They are based on capturing contextual information and exploring spatial context of the image. In general, the task of training any neural network is reduced to the task of minimizing the loss function by adjusting the parameters using the gradient descent method. Below it is presented a Distraction-aware Shadow Detection Network (DSDNet), which can be called a deep CNN.

III. DSDNET APPROACH

For shadow detection it was used convolutional neural network DSDNet which is proposed in [1] and shown in Fig. 3. In order to construct a multiscale network, it was used RexNet-101 as a backbone network. Backbone features run through each scale (i.e. conv conv1, conv2_x, conv3_x, conv4_x, conv5_x) and get into DS module. At each scale, an encoder converts the backbone features to image features. Each of DS modules will take as input an image features and produce DS features, which catch the distraction semantics. In the end, the DS features are concatenated from top to down and finally sent to a fusion layer. This size provides on the output one feature map acquisition. Finally this map is followed by a sigmoid activation function to output a binary shadow map as the final output.

A. *Distraction-aware Shadow (DS) module*

The DS module is used in order to learn semantic features of the distraction regions and concatenate the distraction features with the input image features to produce distraction-aware features, which are used for shadow segmentation. As the input data, it used image features (size: H*W*32), which were produced by a backbone network and Encoder layer (Fig. 3). The output data is DS features with the same size. DS module is made of a FP sub-module and a FN sub-module, and operations to combine different features.

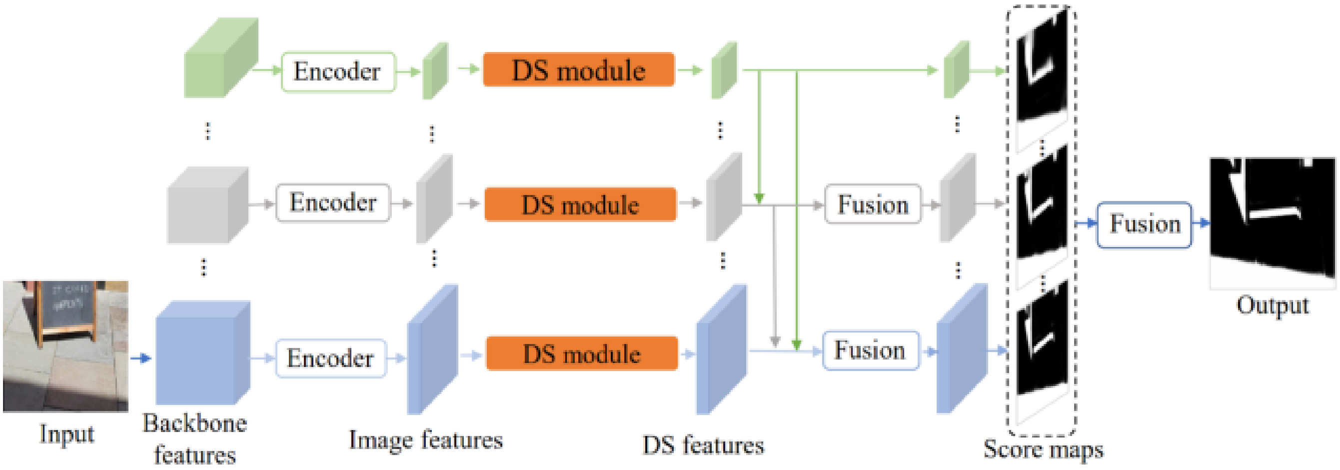


Fig. 3. Architectures of Distraction-aware Shadow Detection Network (DSDNet)

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

In short, image features are passed through FN sub-module in order to produce FN-masked image features, which are concatenated with Image features to produce FN-augmented features. Then, FP sub-module takes as input Image features and receives FN-augmented features and gets FP-aware image features, which are subtracted from FN-augmented features to get desired distraction-aware features.

B. False negative sub-module

FN sub-module is used in order to learn FN features and FN-masked features, which are used to augment the input image features. This module is designed to enable the network to focus on possible FN regions, which would help the network better discriminate the FN regions, whose visual appearances are highly variable from general shadows. Firstly, it employs a feature extractor on the image features to extract the FN features. It used the FN features for FN prediction, by estimating a soft binary map indicating the possible FN locations on the input image. It was necessary as we want to force the FN features to capture the semantics to recognize potential FN regions. Secondly, the FN features are concatenated with the image features and fed into an attention block to produce a mask. Thirdly, a masked image presentation is obtained by multiplying image features with duplicated soft mask element-wise. To enhance the feature activations on FN regions, received features are added to image features to produce FN-augmented image features.

C. False positive sub-module

The **FP sub-module** is used in order to learn the FP features which are used to enhance the FN-augmented features (Fig. 4). Similar to the previous scheme, it was used a feature extractor layer to get false positive features. Then we concatenate the received feature map from the FN sub-module and feed it into the Conv block in order to generate FP-aware image features. Finally, we subtract obtained FP features from FN features to weed out the negative effect of FP features on detection. This operation would make the network less sensitive to potential false positive distraction.

D. Convolutional Layers

The convolutional layers used in our network, except those listed below, are all followed by a Batch Normalization layer and ReLu activation function.

- *Encoder* consists of 2 convolutional layers, which consist of 32 filters with 3x3 kernels.
- Encoder consists of 2 convolutional layers, which consist of 32 filters with 3x3 kernels. *Fusion layer* consists of one convolutional layer with one filter with kernet size = 1 x 1.
- *The feature extractor* in FN Sub-module and FP Sub-module consists of 2 convolutional layers with 32 kernels of size 3 × 3.
- *The attention block* has one convolutional layer with 64 kernels of size 3×3, followed by a sigmoid activation function.
- *The Conv block* in FP Sub-module constituted of the 3 convolutional layers have 64 filters each, with kernel size = 1 × 1, 3 × 3 and 1 × 1, then it is followed by another 3 convolutional layers with 64 filters in the first layer and 32 filters in the others with the same kermel size.

The architecture of the DSDNet neural network is implemented in Python programming language using PyTorch library.

IV. EXPERIMENT

In training neural network it used three public datasets, where all images were reduced to a size of 320x320. In order to expand the training sample, all pictures were augmented by random horizontal flipping. The following sets of images were used as training data: SBU [2, 3] (4089 pictures for training, 638 for testing), UCF [4] (135 pictures for training, 110 for testing) and ISTD [5] (1870 pictures for training, 540 for the test). The neural network DSDNet for detecting shadows from satellite images allows us to reveal the semantics of the image due to the DS module. This module significantly increases the accuracy of shadow segmentation by double verification of geometric features, specifying their belonging to the shadow.

To evaluate the results BER metric was used (1):

$$BER = 1 - \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right), \qquad (1)$$

where TP, TN, FP, FN - denote the numbers of true positives, true negatives, false positive and false negative shadow pixels, respectively. BER is an effective metric to calculate the efficiency on the class imbalance results, and that is why it is extensively used for shadow evaluation. A lower score indicates a better performance.

DSDNet method was compared with many state-of-the-art shadow detection methods like DSCNet [6], scGAN [7], BDRAR [8] and ST-CGAN [9]. Table 1 presents the results of a quantitative comparison of the neural networks which were presented above. It shows that DSD network has the best scores on all test datasets. Fig. 5 shows visual results of the considered neural network workflow on satellite images.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

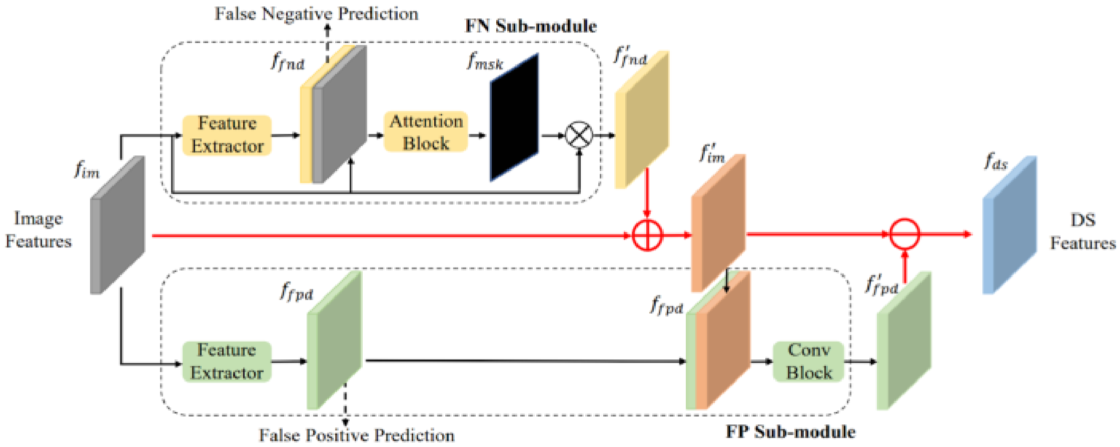


Fig.4. Architecture of the Distraction-aware Shadow module (DS module)

Fig.5. The result of DSDNet work. The white part is the shadow pixels. The black part is non shadow pixels

TABLE I. QUANTUTATUVE SHADOW DETECTION RESULTS

Methods	**BER**		
	SBU	**UCF**	**ISTD**
DSDNet	**3.45**	**7.59**	**2.17**
DSCNet	5.59	10.54	3.42
scGAN	9.04	11.52	4.70
BDRAR	3.64	7.81	2.69
ST-CGAN	8.14	11.23	3.85

V. CONCLUSION

The basic difference of buildings patterns is geometric features on satellite images. It allows us to describe buildings as structured objects. The DS module is an additional clarification of their belonging, thus the accuracy of determining the shadow increases. The DS module augments input image features with explicitly learned distraction features by a specific fusion strategy to produce distraction-aware features for robust shadow detection. However, there are many cases where the network gives a wrong result. It can happen either on weak shadow images where the shadows have very similar brightness to the background or images with an extremely dark background, where the shadows are almost blended into the background. Due to a small number of labeled satellite images, the neural network was trained on public datasets, which contain not only satellite images, but also other cases. If the neural network is trained only on satellite images, the accuracy will be strongly increased. The creation of new datasets will be a further development of DSDNet to improve its efficiency in order to solve the issue of shadow segmentation from satellite images.

REFERENCES

[1] Q. Zheng, X. Qiao, Y. Cao and R. W. H. Lau, "Distraction-Aware Shadow Detection," 2019 IEEE/CVF Conf. on Comp. Vis. and Pattern Recogn. (CVPR), 2019, pp. 5162-5171.

[2] T. F. Y. Vicente, L. Hou, C.-P. Yu, M. Hoai, D. Samaras, “Large-scale training of shadow detectors with noisily-annotated shadow examples,” in European Conference on Computer Vision, 2016, pp. 816–832.

[3] T. F. Y. Vicente, M. Hoai, and D. Samaras, “Noisy label recovery for shadow detection in unfamiliar domains,” in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3783–3792.

[4] J. Zhu, K. G. Samuel, S. Z. Masood, and M. F. Tappen, “Learning to recognize shadows in monochromatic natural images,” in IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 223–230.

[5] J. Wang, X. Li, and J. Yang, “Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal,” in IEEE Conf. on Computer Vision and Pattern Recognition, 2018, pp. 1788–1797.d

[6] X. Hu, C.-W. Fu, L. Zhu, J. Qin, P.-A. Heng, "Direction-Aware Spatial Context Features for Shadow Detection and Removal," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 11, pp. 2795-2808.

[7] V. Nguyen, T. F. Y. Vicente, M. Zhao, M. Hoai, D. Samaras, “Shadow detection with conditional generative adversarial networks,” in IEEE International Conference on Computer Vision, 2017, pp. 4510–4518

[8] L. Zhu, Z. Deng, X. Hu, C.-W. Fu, X. Xu, J. Qin, and P.-A. Heng, “Bidirectional Feature Pyramid Network with Recurrent Attention Residual Modules for Shadow Detection,” in IEEE International Conference on Computer Vision, 2018.

[9] J. Wang, X. Li, J. Yang, "Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal," 2018 IEEE/CVF Conf. on Comp. Vis. and Pattern Recogn., 2018, pp. 1788-1797.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Identifying the Presence of Covid-19 on X-ray Medical Images Using a Neural Network

Alexander Usatov
FAMCS of
Belarusian State University
Minsk, Belarus
ausatov@icloud.com

Ivan Kosik
FAMCS of Belarusian State University
Belarusian State Medical University
Minsk, Belarus
ivankosik91@gmail.com

Abstract. **In this paper, we propose a neural network implementation and its improvements for identifying the presence of covid-19 on X-ray medical images using a neural network.**

Keywords: **covid-19, covid, neural network, convolutional neural network X-ray medical images**

I. INTRODUCTION

Now, there is a big problem in the world with the workload of doctors due to the Covid-19 pandemic. Viral pneumonia is a common complication of influenza-like illnesses and is a complication of SARS-COV-2[1].Viral pneumonia may clear up on its own; however, when severe, it can be life-threatening. Viruses are generally not as common a cause of CAP as some bacteria. However, as well as being a primary pathogen, viruses can be a co-pathogen with bacteria, particularly in those with severe illness requiring admission to ICU and in ventilator-associated pneumonia.A huge number of people need medical care, so the problem of unloading medical personnel using machine learning technologies is urgent. This article explores the possibility of classifying X-ray images using machine learning and deep learning approaches [2]. Since the massive victory of deep convolutional neural network in the 2012 ImageNet competition, the field of deep learning has experienced a huge increase in popularity, and a significant number of different neural network architectures have emerged [3]. For image processing, convolutional neural networks are used, as a rule since they are able to use the local relationship of image pixels. The learning time of such a network becomes a common problem. This article will provide an example of a neural network designed to detect the signs of Covid-19 on chest X-rays and will look at an attempt to shorten the training time of this neural network. X-rays of the chest will be fed to the input of the neural network, an example of such images of a person without Covid-19 and a patient with a diagnosis of Covid-19 can be seen in Fig. 1 [3].

II. DATASET PROPERTIES

In this research we use a public open dataset of chest X-ray and CT images of patients which are positive or suspected of COVID-19 or other viral and bacterial pneumonias (MERS, SARS, and ARDS.) [4]. Data will be collected from public sources as well as through indirect collection from hospitals and physicians. All images and data will be released publicly in this GitHub repo (https://github.com/ieee8023/covid-chestxray-dataset). Dataset is represented as collection X-Ray images. Those images have size 255x255.Every image is supported by metadata about patients including viral, bacterial, fungal, lipoid, aspiration, sex, age, finding, RT_PCR_positive, survival, intubated, intubation present, went_icu and other.

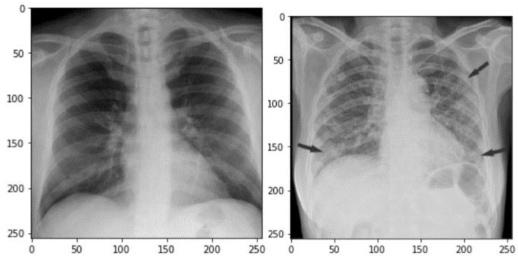


Fig. 1. X-ray of a patient with coronavirus (right) and without (left)

We try modify this information by definition of regions with condensed information using bounding box like as Fig. 2.

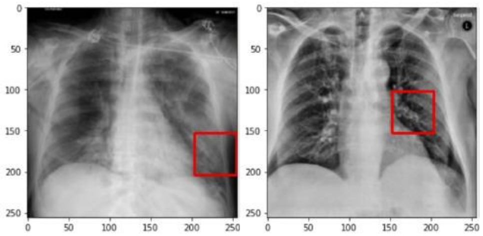


Fig. 2. X-ray images with marked pathology by bounding box

Also, this dataset is supported by information map with features positions (Fig. 3).

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

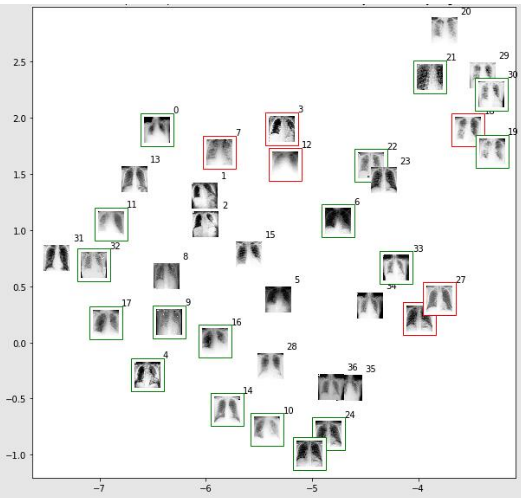


Fig. 3. The map of feature space of pretrained network on COVID-19 X-ray, where true case is marked as green and false case is marked as red

III. NEURAL NETWORK

As basic methods we try to use fully convolutional network for semantic segmentation. It allows to compared with classification and detection tasks, segmentation is a much more difficult task and spend image classification like as [5].

For object detection is used classification within an image with bounding boxes of pathology regions. That means we also need to know the class, position and size of each object. Classify the pathology class for each pixel within an image. That means there is a label for each pixel.

To avoid the “vanishing gradient” problem that sometimes occurs with the usage of the popular ReLU activation function, the ELU activation function has been used instead. Fig. 5 shows the difference between ReLU and ELU, Fig. 6 shows the architecture of the used neural network:

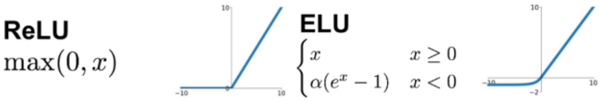


Fig. 5. The difference between ReLU and ELU

IV. IMPROVEMENTS

One of the possible and the most basic simplification of the model is to reduce the dimension of the input data by converting the image to black and white instead of the usual network by three times. It is important to understand that this decreases the amount of incoming data only for the first convolutional layer, since the number of output channels is specified by another parameter.

Let's check the difference in training time for this network [3] with the number of input channels equal to 1 and 3. With three input channels, the neural network has been training for 5 minutes 55 seconds, and with one, the neural network has been training for 5 minutes 33 seconds. The training was carried out on an Nvidia 1650 TI GPU. An increase in the number of input layers from 1 to 3 increases the training time by almost 7% and does not give a noticeable increase in quality, which indicates the usefulness of the proposed approach.

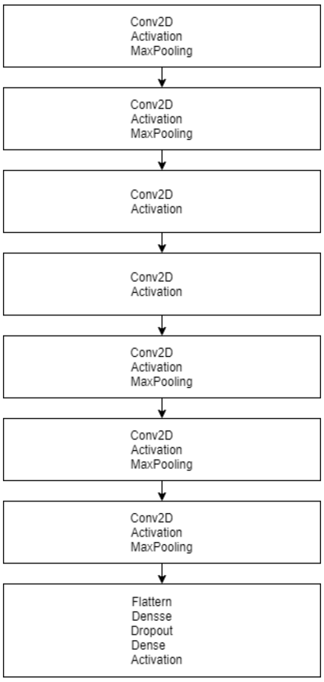
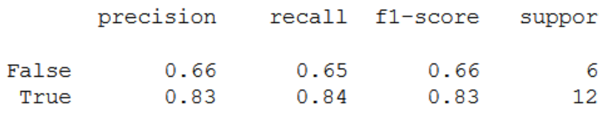


Fig. 6. Architecture of an applied neural network

V. QUALITY METRICS & RESULTS

In diagnostic tasks, positive examples are more important than negative ones, therefore, priority was given to positive examples. The confusion matrix on the test dataset is on Fig. 7:



```
          precision    recall  f1-score   suppor

   False       0.66      0.65      0.66        6
    True       0.83      0.84      0.83       12
```

Fig. 7. Confusion matrix

The discrete ROC curves of these approaches are plotted in Fig. 8. From this picture, it is clearly that proposed model of neural network consistently achieves the impressive performance in terms of the discrete ROC curve.

The Fig. 9 shows an example of an input image and a map of the cumulative intensity of convolutions after layer 19 for this image.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

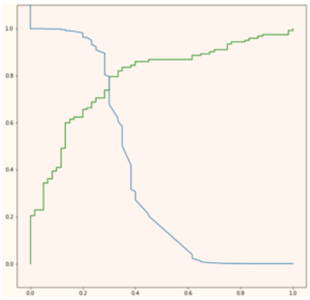


Fig. 8. Architecture of an applied neural network

The black areas correspond to the areas that contribute the least to the result [3].

After smoothing the weighted sum of the image shown in the Fig.6, the following picture (Fig. 9) is obtained. Some of the brighter areas are located in other spots than the location with the Covid-19 signs. This is due to the presence of colors with greater brightness and to the fact that this is not the last layer of the network, so there is reason to suppose that the neural network has learned to process such structures Fig. 10.

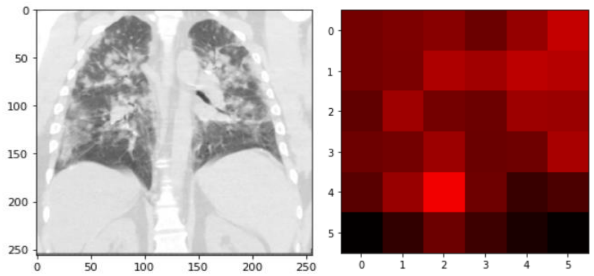


Fig. 9. An example of affected lungs (left), a brightness map at the output after layer 19 (right)

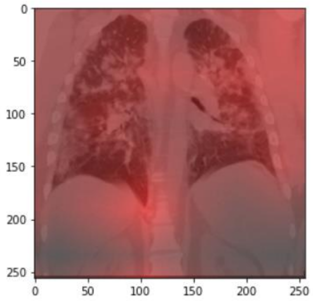


Fig. 10. Smoothed luma map after layer 19

VI. CONCLUSION AND FURTHER RESEARCH

The augmentation is helpful in a small database. However, it is also been found that augmenting a large number of images from a small set of images influences model accuracy because of the addition of noise during model training. On the other hand, a very small number of images did not produce the best result for any model due to the under fitting problems in the pre-trained model. The dataset is downloaded from an open-source online repository. The promising results obtained using CNN models suggest that Chest X-Ray can be useful for early detection of the Coronavirus as compared to the time-consuming pathological test or costly CT-Scan [6].

It is possible to use a more complex neural network structure. There are 2 options here: train the neural network from scratch or try to fine-tune some already existing model, for example, YOLOv4 [7], since it shows impressive results in other tasks. The variant with training a large neural network from scratch takes a lot of time and, probably, a significant increase in the training sample, to avoid memorizing the training sample by the neural network of many training parameters and an insufficient volume of the training sample.

To build a brightness map, you can use a weighted convolution value that will consider the weight in fully connected layers. This will allow to determine which areas of the original image make the greatest contribution more accurately to the result.

The question of the need to use data augmentation remains open, there are reasons that there is no sense in this: the X-ray images are similar to each other, since people are in the same postures during the X-ray examination. Also, all images are usually taken from the same distance and angle.

ACKNOWLEDGMENT

This research was partly supported under the project BRFFI F20KIGT-006.

REFERENCES

[1] Sagnelli C., Celia B., Monari C., Cirillo S., De Angelis G., Bianco A., Coppola N. Management of SARS-CoV-2 pneumonia. J. Med Virol. 2021 Mar; 93(3):1276-1287.

[2] A. M. Недзьведь, С. В. Абламейко, Анализ изображений для решения задач медицинской диагностики, Минск: ОИПИ НАН Беларуси, 2012. – 240 с.

[3] A. A. Alshbishiri, M. A. Marghalani, H. A. Khan, R. G. Ahmad, M. A. Alqarni and M. M. Khan, "Adenoid segmentation in X-ray images using U-Net," 2021 National Computing Colleges Conference (NCCC), 2021, pp. 1–6.

[4] CoronaHack -Chest X-Ray-Dataset [electronic resource]. Access mode: https://www.kaggle.com/praveengovi/coronahack-chest-xraydataset - Date of access: 13.04.2021.

[5] Kawahara J., Hamarneh G. Fully Convolutional Neural Networks to Detect Clinical Dermoscopic Features. IEEE J. Biomed Health Inform. 2019 Mar; 23(2), pp. 578–585.

[6] Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, U. Rajendra .Acharya Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med. 2020 Jun, pp. 121.

[7] Y. Cai et al., "YOLOv4-5D: An Effective and Efficient Object Detector for utonomous Driving," in IEEE Transactions on Instrumentation and Measurement, vol. 70, 2021, pp. 1–13.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Superpixel Clustering for Detection of Binary Object Hierarchy Using Modernized Classical Clustering Methods

Mikhail Kharinov
St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS)
St. Petersburg, Russia
khar@iias.spb.su

Abstract. **The paper proposes the simplest non-algorithmic definition of superpixels as image elements, which itself determines the algorithm for their calculation. A system of three classical methods of image approaching by piecewise constant approximations by means of iterative clustering of image pixels is considered: Ward's clustering, split-and-merge method and K-means method. The modernization of these methods is suggested for reduction of the approximation error E (total squared error) to the achievable minimum values for a fixed cluster numbers g in the current approximation. Advanced versions of the classical methods for reducing of the approximation error E are combined in so-called standard model for detecting of binary hierarchy of objects in the image by means of iterative superpixel clustering. In this paper the advanced versions of mentioned methods are presented and the standard model of binary hierarchy of objects in the image is briefly described.**

Keywords: **digital image, pixel clustering, piecewise constant approximation, total squared error, minimization, superpixel hierarchy, object detection**

I. INTRODUCTION

Modern low-level computer vision and object detection in an image is rather a highly desirable project than an established science about image elements and how to organize and classify these elements to effectively detect objects in a scene. In order to make object detection at the initial stage of image processing a science, first of all, there is a lack of formal definition of image elements, definition of the hierarchy of objects in the image, as well as a system of methods for their effective calculation.

This paper provides a brief overview of the mathematical model, which proposes the definition of the elements that make up the image. The model provides the detection of hierarchically structured objects using modernized methods of classical cluster analysis. Since the model is based on the classical cluster analysis, it is called the standard model.

II. SUPERPIXELS

Although there are no generally accepted definition for image elements, the term has already been coined for them. Wanted image elements are called "superpixels".

The subject of superpixels [1] becomes especially relevant due to the increase in image resolution and the need to reduce the computational complexity of image processing by replacing operations with pixels by operations with superpixels. Usually superpixels are conceptualized as elements of objects or enlarged pixels, in particular, as image segments matching to the boundaries between objects. Grouping pixels is a very good method to reduce computational complexity at the initial stage of image processing. But this does not neglect the problem of justifying how to use the enlarged pixels in the best way.

Superpixels are defined as the maximal sets of pixels that implies preliminary construction of the initial $1, 2, ..., g_1$ series of optimal piecewise constant image approximations in bottom up merging and top down splitting techniques.

The idea of superpixels and approximating of an image by a hierarchy of superpixels is illustrated in Fig. 1.

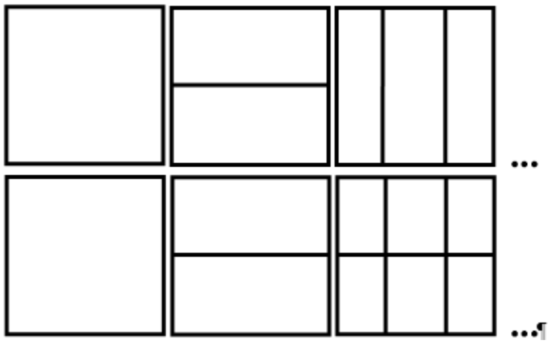


Fig. 1. The sequence of optimal image partitions (above) and the hierarchy of superpixel partitions (bottom).

If the optimal image approximations with $1, 2, 3, ..., g_1$ pixel clusters are determined by the sequence of partitions shown in the upper row of pictures, then the hierarchy of image approximations by superpixels is determined by the hierarchical sequence of partitions into 1, 2, 6..., s superpixels shown in the bottom row of pictures, where $g_1 \leq s \leq (g_1)$!and the exclamation mark "!" indicates a factorial.

Fig. 1 clearly explains that superpixels preserve the boundaries between pixel clusters that are disappeared in the current optimal approximation of the image. The effect of the disap-

---

This work was performed within the framework of the theme 0060-2019-0011 (Fundamentals and technologies of big data for sociocyberphysical systems).

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

pearance of boundaries between pixel clusters is caused by non-hierarchical optimal approximations. It just means that, starting from the optimal approximation, the computer does not "see" the sharp boundaries between the objects. Therefore, to detect objects, it is preferable to use superpixel clustering with a value of error E exceeding the minimum possible one for a given number of superpixels.

III. EXAMPLE OF SUPERPIXELS

Fig.2 presents the optimal and hierarchical superpixel approximations for standard "Lena" image.

In the left column the optimal image approximations with the number of tones from one to nine are demonstrated. The corresponding superpixel approximations of the image with 1, 2, 4, 7, 11, 16, 18, 24 and 28 tones are placed side by side in the right column. Note that the first and the second rows show the pairs of the same approximations as in Fig. 1. Under careful examination, it is not difficult to notice the differences in the approximations being compared, which are effortlessly manifested numerically.

Fig.3 describes the total sequence of 216 optimal approximations calculated for source integer pixel values without an initial pixel enlargement.

The upper graph in Fig.3 shows an increase in the number of clusters $s(g) \sim \sqrt{E(g)}$ in the superpixel image approximation accompanied with an increase in the number of clusters g in the current optimal approximation and a concomitant increment of the number g of reproducible optimal approximations. The dashed line on the graph corresponds to the case of a hierarchy of optimal approximations, when the superpixel approximations coincide with the optimal ones. A significant deviation of the curve $s(g)$ from the dashed straight line indicates for the current optimal approximation that there is a lack of data about sharp objects in optimal approximations with fewer clusters (Fig.1). The bottom graph in Fig.3 shows the dependence of the standard deviation σ on the number of clusters counted along the abscissa on a logarithmic scale. It expresses the approaching of a non-hierarchical sequence of optimal image approximations by the hierarchy of superpixel approximations. The dependence of the standard deviation σ on the number of clusters $s(g)$ in the superpixel approximation is shown by a black curve, and the dependence $\sigma(g)$ on the number of clusters g in the current optimal image approximation is shown in gray.

The hierarchy of superpixels is much more convenient for computing objects than a non-hierarchical set of clusters of optimal approximations. At the same time, the sequence of optimal approximations has a remarkable advantage, expressed in the fact that it is described by the convex dependence of the approximation errors E_g on the cluster numbers g:

$$E_g \leq \frac{E_{g-1} + E_{g+1}}{2}, g = 2, 3, ...,$$

which implies proper ordinate $\sigma \rightarrow E \sim \sigma^2$ transformation for the bottom graph in Fig.3.

Fig. 2. Sequences of optimal (left) and superpixel (right) approximations of "Lena" image.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

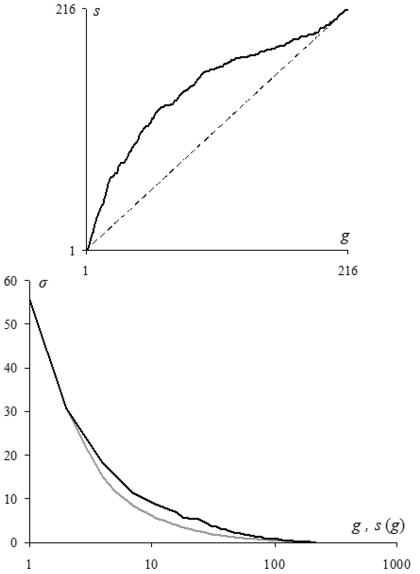


Fig. 3. The numbers of superpixels s depending on the number of clusters g (above). The curves $\sigma(g)$ for optimal (above) and $\sigma(s)$ for hierarchical superpixel image approximations (bottom).

IV. SYSTEM OF THREE CLUSTERING METHODS

To get both advantages of optimal and hierarchical image approximations, it is quite attractive to construct a binary hierarchy of approximations, which is described by the convex dependence of the approximation error E on the number of clusters g. To do so, it seems reasonable to start with a sufficient number of superpixels and construct a target hierarchy that satisfies the convexity condition. In this case, the corresponding resulted curve will pass between the black and gray curves in the bottom graph of Fig.3.

At first glance, the optimal approximations for the above grayscale image, can be obtained using the multithreshold Otsu method [2], [3]. However, due to the exponential growth in the complexity of calculations, it is impossible to obtain even two dozens optimal approximations using this method. The problem is solved (Fig.2, Fig.3) through the combined use of another several algorithms of E minimization. These turned out to be three classical methods of cluster analysis, namely, Ward's pixel clustering, split-and-merge method, and K-means method [4]. An attractive circumstance is that these methods are suitable not only for grayscale, but also for color images. But, on the other hand, for the most effective application to images, they require some modernization.

Original Ward's clustering [5], [6] provides the generation of hierarchical image approximations described by the convex dependence of the approximation error E on the cluster number g. To reduce computational complexity, some seed segments or clusters of close pixels are commonly instead of the original pixels [6]. However, due to high computational complexity, original Ward's method is rarely used. But its main drawback is the unstable clustering result, which varies depending on the number of enlarged pixel and on the heuristic enlarged pixel themselves.

In order not to deal with the case-dependent convex sequence $E(g)$ and, at the same time, speed up the processing, Ward's pixel clustering is performed within the parts of the image. Since the computational complexity of the Ward's method is quadratically depends on the number N of pixels in the image, it decreases just as fast when the image is processed by parts. According to our estimate [7], for recursively repeated processing by parts, the computational complexity is drop as $N^2 \to N^{\frac{4}{3}} \to N^{\frac{16}{15}} \to N^{\frac{256}{255}} \to etc$, depending on the recursion step.

In a single processing by parts, first the image pixels are divided into g_0 clusters, and for each cluster its own approximation hierarchy is constructed in the merge mode as for separate image. Then the hierarchies are combined into one, by reordering of the cluster mergings and hierarchy constructing is completed by the original Ward's method.

If the provided input partition of pixels into g_0 clusters coincides with the optimal one, then the above procedure will lead to a convex dependence $E(g)$. Otherwise, the convexity may be violated at g_0 value. To avoid violation of the convexity property at g_0 clusters, it is enough to process the input approximation by so called CI-method that is modernized split-and-merge reducing of the approximation error E without changing the number of clusters.

CI-method [7] locally supports the convexity condition and provides for a given image approximation a real minimization of the approximation error E under an unchanged cluster number g_0. It corrects the convexity violation at the cluster number g_0 and ensures the separability of Ward's clustering by the image parts. In this case, Ward's clustering by the image parts is simply recalculated into the conventional Ward's clustering without modifying of pixel clusters obtained for the image parts.

The commonly used K-means method, developed and adapted a long time ago for calculations by means of arithmometer, is replaced in the standard model by the significantly modernized K-meanless method [8], which, presupposes the preliminary approximation error E reduction to the vicinity of the minimum achievable values by recursive Ward's method and CI method.

Unlike K-means, versions of K-meanless method for estimating of the error increment either use the values of the target functional E [8] or the exact formula [9]. Thus, the modernization K-means→K-meanless consists, first of all, in a more accurate analytical criterion for reclassifying the sets of pixels from cluster to cluster, as well as, in minimizing E by reclassification not only individual pixels, but pixel sets from the entire hierarchy of the cluster parts of various sizes.

The modernized three methods for minimizing of the approximation error E all and sundry are distinguished from conventional ones by the use of the entire hierarchy of image clusters. In the case of CI-method and K-method for improving of the quality of a separate image approximation, the entire hierarchy of clusters is corrected through the iterative alternate application of three mentioned methods.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

V. BINARY OBJECT HIERARCHIES

In the standard model, all pixels of the image are divided among themselves into pixels of various objects, which are relegated, for example, to background or to other types of objects. It is supposed that objects make up a binary hierarchy produced by the image partition into g_0 basic objects while other objects are represented as parts or unions of basic objects.

A characteristic feature of the standard model is that the basic objects and the binary hierarchy of objects, germinated by them, are detected ambiguously. Depending on the number of objects g_0, as well as a concrete task, it is set during the process of tuning up of the software system for the best approximation of target objects by means of a minimum number of pixel clusters. For a given number of basic objects g_0, the optimal approximation of the image in g_0 colors with the minimum achievable approximation error $E \sim \sigma^2$ or the standard deviation σ is considered as the best.

VI. STANDARD MODEL OF OBJECT DETECTION

The declared standard model for detecting objects in a digital image actually combines two equivalent models, namely, a meaningful one, which describes the detection of objects by means of sets of pixels of an image matrix, and a computational model, in which the considered pixel clusters are described by networked graphs that support high-speed construction and transformation of pixel clusters, as well as storing and calculating of the required features without repeating the calculations once performed.

The standard model, supporting the reversible computation [10], [11] to describe an ambiguous image as superposition of $g_0 = 1, 2, ...N$ object hierarchies, consists of:

- the concept of an object, binary hierarchy of objects, and the definition of superpixels, from which the system of necessary algorithms, model parameters, data structure and methods of accelerating of calculations are deduced;
- three above modernized method of classical cluster analysis as well as the method of filtering of the objects according to the established threshold [7];
- the data structure of algebraic multilayer network [7], in terms of which the definitions for objects and superpixels are reformulated and high-speed algorithms are performed.

In the standard model, the following settings are provided for detecting objects in the image:

- the number g_0 of basic objects identified with the pixel clusters of the optimal image approximation in g_0 colors, which is contained in the target hierarchy of image approximations;
- the number s of superpixels or the corresponding number g_1 of available optimal image approximations constituted of superpixels without any distortion;
- threshold parameter of heterogeneity $H_{threshold}$ for heterogeneity $\left|H \equiv \frac{dE}{dg}\right|$ itself, determined as the absolute value of the derivative of the approximation error E with respect to the number g of pixel clusters [7].

VII. CONCLUSION

Thus, within the framework of the standard model, the problem of binary hierarchy of objects detecting in an ambiguous image has been formulated and practically solved. Formally, the solution to this problem is expressed as an approximation of a “convex sequence” of optimal piecewise-constant image approximations by means of the same “convex binary hierarchical sequence” of image approximations, which contains an optimal approximation of an image in g_0 colors.

The peculiarity of just this paper is that for the invariant representation of the image regardless of the scale, as well as for the high-speed object detection without loss of accuracy, the definition of image elements (superpixels) is suggested, and an example of their calculation for the standard “Lena” image is given. In the future, it is planned to organize a database of optimal approximations and corresponding hierarchies of superpixels for “Lena”, as well as for other standard images, and place the database in the public domain together with program texts and their executable modules used to generate superpixels and optimal image approximations from database.

The experience of experimental research of the standard model shows that it expands the area of effective application of classical cluster analysis by solving problems of processing images of various content [12].

REFERENCES

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Süsstrunk “SLIC superpixels compared to state-of-the-art superpixel methods,” IEEE transactions on pattern analysis and machine intelligence, 34(11), pp. 2274–2282, 2012.

[2] N. Otsu “A threshold selection method from gray-level histograms,” IEEE transactions on systems, man, and cybernetics 9(1), pp. 62–66, 1979.

[3] P.S. Liao , T.S. Chen, P.C. Chung “A fast algorithm for multilevel thresholding,” J. Inf. Sci. Eng. 17(5), pp. 713–727, 2001.

[4] I.D. Mandel Cluster Analysis, Moscow: Finance and Statistics, 176 pp., 1988.

[5] J.H. Jr.Ward “Hierarchical grouping to optimize an objective function,” Am. Stat. Assoc. 58(Issue 301), pp. 236–244, 1963.

[6] T. N. Tran, R. Wehrens, L. M. C. Buydens “SpaRef: a clustering algorithm for multispectral images,” Analytica Chimica Acta 490 (1–2), pp. 303–312, 2003.

[7] M.V. Kharinov, A.N. Buslavsky “Object Detection in Color Image,” In: Proc. of the 14th Intern. Conf. on Pattern Recognition and Information Processing (PRIP’2019), Minsk, 21–23 May, pp. 43–47, 2019, doi:10.13140/RG.2.2.28493.28640.

[8] S.D. Dvoenko “Meanless k-means as k-meanless clustering with the bi-partial approach,” In: Proc. of the 12th Intern. Conf. on Pattern Recognition and Information Processing (PRIP’2014), Minsk, Belarus, pp. 50–54, 2014.

[9] M.V. Kharinov “Reclassification formula that provides to surpass K–means method,” arXiv preprint, arXiv: 1209.6204v1, 10 pp., 2012.

[10] T. Toffoli “Reversible computing,” In International Colloquium on Automata, Languages, and Programming, Springer Berlin Heidelberg. pp. 632–644, 1980.

[11] M.V. Kharinov “Reversible merging of structured clusters of pixels,” Computer graphics and vision (Graphicon’2016), Proc. of the 26th intern. conf. on computer graphics and vision: Moscow, MSU, Sept. 19 - 23 2016, Nizhny Novgorod, pp. 298–302, 2016.

[12] I.G. Khanykov, V.A. Nenashev “The Application of the High-Speed Pixel Clustering Method in Combining Multi-Angle Images Obtained from Airborne Optical-Location Systems,” 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), IEEE 2020, pp.1–8, doi:10.1109/WECONF48837.2020.9131157.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

An Iterative Error Patterns Library Formation Method for the Decoding of Product Codes

Xunhuan Ren
Dept. of Infocommunication Technologies
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
renxunhuan@bsuir.by

Jun Ma
Dept. of Infocommunication Technologies
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
majun1313@hotmail.com

Valery Konopelko
Dept. of Infocommunication Technologies
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
volos@bsuir.by

Victor Tsviatkou
Dept. of Infocommunication Technologies
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
vtsvet@bsuir.by

***Abstract.* Product codes are preferred in high data rate wireless communication systems to achieve good error performance. In this paper, the problem of two-dimensional syndrome-norm decoding of product codes based on a library of error patterns is considered. In product coding, sequence code is first transformed into a code matrix, and then the row and column check code are calculated. In the decoder, the error position of the two-dimensional can be obtained by the operations that first calculate the syndromes and norms, then match with the error patterns in the existing library. The error pattern library is stored in the memory and generated by the subset of the error pattern. This paper proposed a mathematical model for fast generating a library based on the iterative expansion of the error patterns, which makes it possible to shorten the computational complexity in comparison with the known approaches.**

***Keywords:* product codes, error-correcting coding, norm, syndromic-norm decoding, library of error patterns**

I. INTRODUCTION

Error correcting codes are used in many places, wherever there is the possibility of errors during transmission [1]. In Error control coding(ECC), parity check bits are calculated based on the input data. The input data and parity check bits are transmitted across a noisy channel. In the receiver, an ECC decoder is used to detect or correct the errors induced during the transmission. The number of parity bits depends upon the number of information bits. At present, the most successful coding schemes are turbo codes and low-density parity-check codes, since their excellent capability, closely to the Shannon limit. Under some specific requirement (typically, code-rates near to the unity and low error rates required), product codes may turn into competitive. Product codes [2, 3], which can be easily realized by concatenating simple component codes, have a good protection capability against both random and burst errors. Product codes, whose component codes are Hamming or extended Hamming product codes (BCH product codes), BCH product codes can be constructed to improve the error correction capability, but a more complex decoding process is required. To reduce the decoding complexity of BCH product codes based on the three-stage scheme[4], a syndrome-norm method based on library matching of error patterns is proposed in [5]. The syndrome-norm method is a decoding method. It can fast match the target pattern from the pattern library by using the norm, which is calculated based on the result of the syndrome vector, can uniquely determine the base pattern and the corresponding set of error patterns and then applied a predefine correction method to correct the errors. The library of error patterns consists of some basic patterns that delegate a set of error patterns with a common characteristic. If one pattern can be transformed into another pattern by exchanging internal two-row or two-column, then we consider that these two patterns share the same characteristics. The decoding ability of this method is wholly dependent on the error pattern library. Therefore, one of the limitations of this method is that the formation of the matching library is very time-consuming, so it is tough to form a pattern with a lot of errors.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

In the recent ten years, two library formation methods have been proposed to overcome the problem mentioned above. To simplify the problem, all the patterns are referred as matrices, and in each matrix, each error position is marked as one. Then, it assumed that the maximum number of errors in a certain pattern is not above the square root of the total number of elements of it. A method based on rank has been proposed in [6]. This method first enumerates all the possibilities of error patterns in a given size. This pattern set is named as original error pattern set. Then it compares all these patterns to selects some patterns with distinct feature vector to adding to the library. The feature vector used in this method including the rank of the matrix. Another formation method is based on sub-classified, which proposed in [7]. It tends to first category the original patterns set into several subset according to some criterion then conduct the comparing them in a smaller set rather than directly compare them with each other. This method saves more time when comparing with the rank method. However, the sub-classified method still needs to generate all possible error patterns as the original pattern, which may cost a lot of time. As a result, it is necessary to develop a new library formation method that is not needed to list all the error patterns. In this paper, an iterative formation method has been proposed.

II. BASIC CONCEPTS

A. Product codes

Product codes are widely used to correct errors in data transmission and storage systems. They are formed in the verification codes $C_1(n_1,k_1,d_1)$ and $C_2(n_2,k_2,d_2)$ for the rows and columns of the source code matrix $n_1 \times n_2$, respectively, where n is the length of the code; k is the number of information symbols. The random-error-detecting and random-error-correcting capabilities of code are determined by its minimum distance $d_{\min}$, if the component codes C_1 and C_2 have minimum Hamming distances d_1 and d_2, accordingly, then the minimum Hamming distance of the two-dimensional product codes C_{pc} is the product d_1 and d_2 ($d_{\text{pc}} = d_1 \times d_2$), and at the same time $d_{\text{pc}} \geq 2t+1$, t is the multiplicity of corrected errors [8], which greatly increases the error correction capability. Concept of two-dimensional product (iterated) codes shows as Fig. 1.

The simplest two-dimensional product codes are single parity check (SPC) product codes, assured to correct only one error by inverting the intersection bit in the erroneous row and column [8, 9].

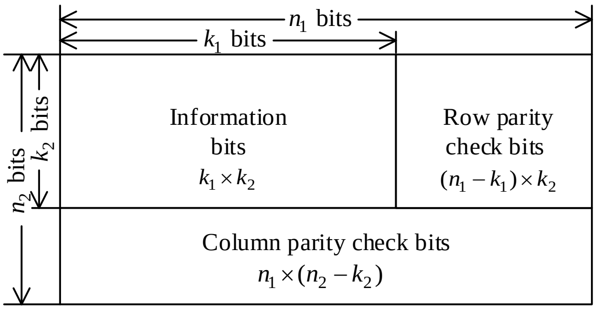


Fig. 1. Concept of two-dimensional product codes

Multidimensional SPC product codes can be constructed to improve the error correction capability, but a more complex decoding process is demanded [10].

B. Syndrome-Norm Decoding For Product codes

In the scheme of syndrome-norm decoding of product codes [4] , the syndromes and norms of rows and columns are first calculated. Then, these two parameters can match with a unique error pattern in the library. But before match operation, it is useful to shorten the search area by calculating the rough number of the error. After the matching operation, we will correct the error by a specified corrected method. The diagram of the whole decoding process is presented in Fig. 2.

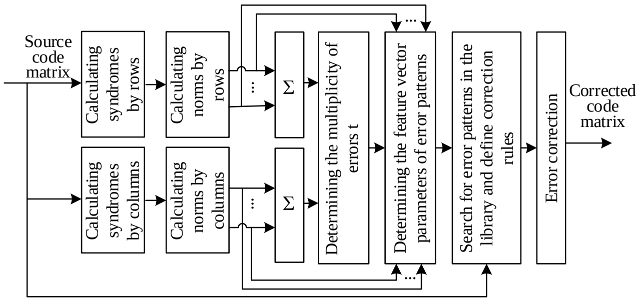


Fig. 2. Diagram of syndromic-norm decoding of product codes

C. Mathematical Condition for Library

The matching library consists of distinct selected error patterns. The selected error pattern is represented by $\{M_R(n)\}_{(n=\overline{1,N(t)})}$, which should satisfy the following mathematical condition:

$$M_R(n) = \left\| m_R(n,i,j) \right\|_{(i=\overline{1,t},\, j=\overline{1,t})}, \tag{1}$$

$$\sum_{i=1}^{t} \sum_{j=1}^{t} m_R(n,i,j) = t , \tag{2}$$

$$\neg\exists n_1 \neg\exists n_2 \left(M_R(n_1)\right) = f_{RT}\left(f_{CT}\left(M_R(n_2), j_c\right), i_r\right) \quad \text{when } j_c = \overline{1,T},\ i_r = \overline{1,T} \tag{3}$$

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

where $N(t)$ – total number of all selected error patterns; f_{RT} – functions for exchange two rows; f_{CT} – functions for exchange two columns; T – total number of possible exchanges in row or column, $T = C_t^2$; i_r , j_c – index of exchange of rows or columns.

III. PROPOSED METHOD

The proposed method is based on the fact that all the patterns with *n* errors can be obtained by extending one error from the patterns with $n-1$ errors. So there is only require to select the representative pattern from a small set rather than all the patterns. Therefore, the proposed method is an iterative method. The mathematic model of the proposed method is presented in following.

$$\begin{aligned}&\left\{M_R(n,t)\right\}_{n=\overline{1,N(t)}} = \left\{M_x(n_x,t)\right\}_{n_x=\overline{1,N_X(t)}}\\&\cup\left\{M_x(n_x,t)^T\right\}_{n_x(t)=\overline{1,N_x(t)}}\end{aligned} \tag{4}$$

$$\begin{aligned}&\left\{M_x(n_x,t)\right\}_{n_x=\overline{1,N_X(t)}}\\&= f_M\left(\left\{M_d(n_d,t), p_{Rank}(n_d,t), P_{RC}(n_d,t), P_{WS}(n_d,t)\right\}_{n_d=\overline{1,N_d(t)}}\right),\\&\left\{M_d(n_d,t)\right\}_{n_d=\overline{1,N_d(t)}} = f_{add}\left(\left\{M_x(n_x,t-1)\right\}_{n_x=\overline{1,N_X(t-1)}}\right),\\&N_d(t) = N_X(t-1)\cdot(t^2-t+1),\\&M_x(n_x,1) = m_x(n_x,1) = 1 \text{ when } N_x(1) = 1,\\&\neg\exists n_1\neg\exists n_2\left(M_R(n_1) = \left(M_R(n_2)^T\right)\right) \text{ when } n_1 = \overline{1,N(t)},\\&n_2 = \overline{1,N(t)},\end{aligned} \tag{5}$$

where $M_x(n_x,t)$ – base error patterns at the *t*-th iteration, $M_x(n_x,t) = \left\|m_x(n_x,\mathrm{t},i,j)\right\|_{\left(i=\overline{1,t},\, j=\overline{1,t}\right)}$; $M_d(n_d,t)$ – extended error patterns that derived from the base error patterns; f_{add} – the function that expanding the error pattern and adding one single element.

$$P_{RC} = f_{sort}\left(p_{RE}, p_{CE}\right). \tag{6}$$

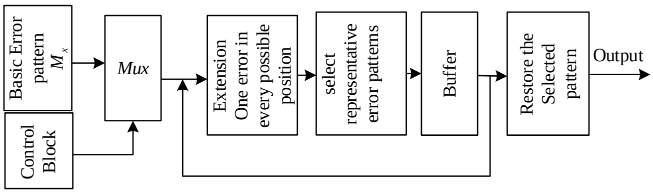


Fig. 3. Block diagram of error pattern generation method based |on the iterative extension

Fig. 3 shows the block diagram of the proposed method, it can be seen that a new set of error patterns can be obtained based on the former set of error patterns, go through a series of expanding operations, and selecting operations. By contrast with other generating methods, the search range of representative error patterns of the proposed is less, which may reduce the time for creating the whole library.

IV. EXPERIMENT AND RESULTS

The comparison of the computational speed is conducted on the basis of the implementation of three different formation methods. The platform of the experiment environment in Matlab under the windows 10 operating system. The average execution time for generating the error patterns library used by the different methods is summarized in Table I.

TABLE I. AVERAGE TIME FOR GENERATING ERROR PATTERN LIBRARIES FOR DIFFERENT METHOD

Methods for generating error patterns	**Number of the errors t**				
	2	***3***	***4***	***5***	***6***
Rank	<1 s	<1 s	<1 s	7 s	10 m
Sub-class	<1 s	<1 s	<1 s	5 s	8 m
Iterative	<1 s	<1 s	<1 s	<1 s	<1 s

We can see from the table. 1 that when the number of errors is below 5, all three methods can generate a corresponding library within one second, however, when the number of errors is surpassed 5, there are emerge some differences. For the rank method, it will take 7 seconds to erect an error pattern library when *t* is 5, and it will take 10 minutes to create another library when *t* is 6. The sub-classified method costs lower time by comparing with the rank method, but it also needs 8 minutes to form a library when *t* is 6. The proposed method has the best result. For all different numbers of *t*, the time spent by the proposed method to construct the same libraries that build by the former two methods is all within one second. It is convinced that the proposed method has a better performance in terms of processing speed when compare with the rank method and subclassified method.

V. CONCLUSION AND FUTURE WORK

A model for forming a library of error patterns based on iterative expansion of the error patterns for the syndrome-norm decoding of iterative codes are proposed in this paper. The proposed method differs from the known rank method and subclassified method by it adopts an iterative forming method to build a library of error patterns. By iterative adding extra one error to the patterns that generated in the former iteration and eliminating redundant error patterns and retaining those representative patterns that have distinct feature vectors. The feature vectors are obtained according to the formula (6). The experiments has proved that the proposed method is a very efficient method, which requires less time to build a same

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

library that formed by the rank method and by the subclassified method because the proposed method cancel to enumerate all the permutation of the possible error situation. The result of the proposed method can be applied in the syndrome-norm method, which is a decoding method based on the library matching.

In the future, we would like to analysis these patterns in the library, and for each pattern, we will construct a corresponding correction method. Then conduct a comprehensive comparison with the famous method. Another interesting direction is explore the performance of the syndrome-norm method.

REFERENCES

[1] S. Lin and D. J. Costello, “Error Control Coding Fundamentals and Applications Second Edition.”

[2] R. M. Pyndiah, “Near-optimum decoding of product codes:block turbo codes,” IEEE Transactions on Communications, vol. 46, no. 8, pp. 1003–1010, 1998.

[3] F. Chiaraluce and R. Garello, “Extended Hamming product codes analytical performance evaluation for low error rate applications,” IEEE Transactions on Wireless Communications, vol. 3, no. 6, pp. 2353–2361, 2004.

[4] B. Fu and P. Ampadu, On hamming product codes with type-II hybrid ARQ for on-chip interconnects. IEEE Trans Circuits Syst I, Reg Papers, vol. 56, no. 9, pp. 2042–2054, 2009.

[5] V. A. Lipnitski, V. K. Konopelko, The theory of syndrome norms in the permutation decoding action unjammable codes. Doklady BGUIR, vol.1, no. 1, pp. 146–157, 2000.

[6] V. K. Konopelko, A. V. Lipnitski and N. V. Spichekova, Point pattern classification and classical number partition problem. Doklady BGUIR, vol. 51, no. 5, pp. 112–117, 2010.

[7] O. G. Smolyakova, V. K. Konopelko Classification vectors of errors at two-dimensional coding of the information. Doklady BGUIR, vol. 37, no. 7, pp. 19–28, 2008.

[8] P. Elias, "Error-free Coding," in Transactions of the IRE Professional Group on Information Theory, vol. 4, no. 4, pp. 29–37, September 1954, doi: 10.1109/TIT.1954.1057464.

[9] S. Lin and D. J. Costello, Error Control Coding, Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 2004.

[10] L. Ping, S. Chan, and K. L. Yeung, “Iterative decoding of multi-dimensional concatenated single parity check codes,” in Proceedings of the IEEE International Conference on Communications (ICC ’98), vol. 1, pp. 131–135, Atlanta, Ga, USA, June 1998.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Innovative Technique for Computing Shortest Length and/or Equal Tails Confidence Intervals in Reliability and Safety under Parametric Uncertainty

Nicholas A. Nechval
BVEF Research Institute
University of Latvia
Riga, Latvia
nechval@telenet.lv

Gundars Berzins
BVEF Research Institute
University of Latvia
Riga, Latvia
gundars.berzins@lu.lv

Konstantin N. Nechval
Aviation Department
Transport and Telecommun. Institute
Riga, Latvia
konstan@tsi.lv

Max Moldovan
Biometry Hub, Faculty of Science
University of Adelaide
Adelaide, Australia
max.moldovan@adelaide.edu.au

Vadims Danovics
BVEF Research Institute
University of Latvia
Riga, Latvia
vadims.danovics@lu.lv

Irina Bausova
BVEF Research Institute
University of Latvia
Riga, Latvia
irina.bausova@lu.lv

Abstract. **A confidence interval is a range of values that provides the user with useful information about how accurately a statistic estimates a parameter. In the present paper, a new simple computation technique is proposed for simultaneous constructing and comparing confidence intervals of shortest-length and equal tails. This unified computation technique provides intervals in several situations that previously required separate analysis using more advanced methods and tables for numerical solutions. In contrast to the Bayesian approach, the proposed approach does not depend on the choice of priors and is a novelty in the theory of statistical decisions. It allows one to exclude nuisance parameters from the problem using the technique of invariant statistical embedding and averaging in terms of pivotal quantities (ISE & APQ) and quantile functions. It should be noted that the well-known classical approach to constructing confidence intervals of the shortest length considers at least three versions of possible solutions and is in need of information about the forms of probability distributions of pivotal quantities in order to determine an adequate version of the correct solution. The proposed technique does not need such information. It automatically recognizes an adequate version of the correct solution. To illustrate this technique, numerical examples are given.**

Keywords**: parametric uncertainty, pivotal quantity, confidence intervals of shortest length and/or equal tails**

I. INTRODUCTION

To make a statistical inference in many problems under parametric uncertainty, the experimenter is interested in constructing a confidence interval that contains the true (unknown) value of the parameter (say, λ) with a given probability [1, 2]. If we are given a random sample $\mathbf{Y}=(Y_1,Y_2,\ ...,Y_n)$ from a density $f_\lambda(y)$ and a pivotal quantity $V(S(\mathbf{Y}),\lambda)$ (which is developed from either maximum likelihood estimate or sufficient statistic $S(\mathbf{Y})$), whose distribution does not depend on λ, then confidence interval for a single unknown parameter λ is often derived by using a pivotal quantity $V(S(\mathbf{Y}),\lambda)$.

II. CLASSICAL APPROACH TO CONSTRUCTING SHORTEST LENGTH CONFIDENCE INTERVALS

It is assumed that the length of the statistical confidence interval is given by

$$L(q_1,q_2\mid S(\mathbf{Y}))\propto L(q_1,q_2)=\int_{q_1}^{q_2}\zeta(\tau)d\tau. \tag{1}$$

The length of the expected confidence interval is given by

$$E_\lambda\{L(q_1,q_2\mid S(\mathbf{Y}))\}\propto L(q_1,q_2)=\int_{q_1}^{q_2}\zeta(\tau)d\tau. \tag{2}$$

In order to find the 100(1−α)% statistical (or expected) shortest-length confidence interval for λ, we should find a pair of decision variables q_1 and q_2 such that $L(q_1,q_2)$ is minimum.

A. Problem Statement in Terms of the Pivot V and Decision Variables q_1 and q_2

Minimize

$$L(q_1,q_2)=\int_{q_1}^{q_2}\zeta(\tau)d\tau \tag{3}$$

subject to

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

$$\int_{q_1}^{q_2} q(v)dv = Q(q_2) - Q(q_1) = 1-\alpha. \qquad (4)$$

B. *Classical Analytical Approach to Solution of the Problem*

Differentiating $L(q_1, q_2)$ with respect to q_1, we get

$$\frac{dL(q_1,q_2)}{dq_1} = \frac{d}{dq_1}\int_{q_1}^{q_2} \zeta(\tau)d\tau = \zeta(q_2)\frac{dq_2}{dq_1} - \zeta(q_1)\,. \qquad (5)$$

From (4) we find the derivative of q_2 with respect to q_1 as follows:

$$\frac{d}{dq_1}\int_{q_1}^{q_2} q(v)dv = \frac{d}{dq_1}(1-\alpha) \qquad (6)$$

that is

$$q(q_2)\frac{dq_2}{dq_1} - q(q_1) = 0. \qquad (7)$$

Thus, we have

$$\frac{dq_2}{dq_1} = \frac{q(q_1)}{q(q_2)}. \qquad (8)$$

Substituting this into (5), we obtain

$$\frac{dL(q_1,q_2)}{dq_1} = \zeta(q_2)\frac{q(q_1)}{q(q_2)} - \zeta(q_1). \qquad (9)$$

Now consider the following three versions of possible decision making.

C. *Version 1 of Possible Decision Making*

$$\frac{dL(q_1,q_2)}{dq_1} = \zeta(q_2)\frac{q(q_1)}{q(q_2)} - \zeta(q_1) = 0. \qquad (10)$$

Then the optimal analytical solution of the problem is given by

$$\frac{q(q_1)}{\zeta(q_1)} = \frac{q(q_2)}{\zeta(q_2)}. \qquad (11)$$

D. *Version 2 of Possible Decision Making*

$$\frac{dL(q_1,q_2)}{dq_1} = \zeta(q_2)\frac{q(q_1)}{q(q_2)} - \zeta(q_1) < 0. \qquad (12)$$

It follows from (4) that

$$1-\alpha < Q(q_2) \le 1. \qquad (13)$$

Then the optimal analytical solution of the problem is given by

$$q_2 = \arg[Q(q_2)=1], \quad q_1 = \arg[Q(q_1)=\alpha]. \qquad (14)$$

E. *Version 3 of Possible Decision Making*

Let us assume that

$$Q(q_2) = 1-\bar{Q}(q_2), \quad Q(q_1) = 1-\bar{Q}(q_1). \qquad (15)$$

It follows from (4) that

$$\int_{q_1}^{q_2} q(v)dv = Q(q_2) - Q(q_1) = \bar{Q}(q_1) - \bar{Q}(q_2) = 1-\alpha. \qquad (16)$$

It follows from (16) that

$$\frac{d}{dq_1}\int_{q_1}^{q_2} q(v)dv = \frac{d}{dq_1}\left(\bar{Q}(q_1) - \bar{Q}(q_2)\right) = \frac{d}{dq_1}(1-\alpha) \qquad (17)$$

that is

$$\bar{Q}'(q_1) - \bar{Q}'(q_2)\frac{dq_2}{dq_1} = 0. \qquad (18)$$

Thus, we have

$$\frac{dq_2}{dq_1} = \frac{\bar{Q}'(q_1)}{\bar{Q}'(q_2)}. \qquad (19)$$

Substituting this into (5), we obtain

$$\frac{dL(q_1,q_2)}{dq_1} = \zeta(q_2)\frac{\bar{Q}'(q_1)}{\bar{Q}'(q_2)} - \zeta(q_1) > 0. \qquad (20)$$

It follows from (16) that

$$1-\alpha < \bar{Q}(q_1) \le 1, \qquad (21)$$

Then the optimal analytical solution of the problem is given by

$$q_1 = \arg[\bar{Q}(q_1)=1], \quad q_2 = \arg[\bar{Q}(q_2)=\alpha]. \qquad (22)$$

III. NEW SIMPLE COMPUTATION TECHNIQUE OF CONSTRUCTING CONFIDENCE INTERVALS

It is assumed that the length of the confidence interval is given by (3). The proposed approach is based on the use of the numerical values of the quantile functions q_1 and q_2. In order to find the 100(1−α)% shortest-length confidence interval for λ, we should find a pair of numerical values of the quantile functions q_1 and q_2 such that $L(q_1, q_2)$ is minimum.

A. *Problem Statement in Terms of the Decision Variable p (Probability) and Quantile Functions q_1 and q_2*

Minimize

$$L^2(q_1,q_2) = \left(\int_{q_1}^{q_2} \zeta(\tau)d\tau\right)^2, \qquad (23)$$

where the quantile function q_1 is given by

$$q_1 = Q^{-1}(p), \qquad (24)$$

the quantile function q_2 is given by

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

$$q_2 = Q^{-1}(1-\alpha+p), \tag{25}$$

subject to

$$0 \le p \le \alpha. \tag{26}$$

The decision variable to be determined is p (probability).

If $p = \alpha/2$ we have the confidence interval of equal tails.

B. New Simple Computation Method for Numerical Solution of the Problem

The optimal numerical values of the quantile functions q_1 and q_2, which minimize $L(q_1,q_2)$, can be obtain from (23)-(26) using computer software "Solver".

IV. NUMERICAL EXAMPLE 1

A. Classical Analytical Approach: The Problem Statement in Terms of the Pivot V and Decision Variables q_1 and q_2

Minimize

$$L(q_1,q_2) = \int_{q_1}^{q_2} \zeta(\tau)d\tau = \int_{q_1}^{q_2} \tau^{-2}d\tau = q_1^{-1} - q_2^{-1}. \tag{27}$$

subject to

$$\int_{q_1}^{q_2} q(v)dv = Q(q_2) - Q(q_1) = q_2^n - q_1^n = 1-\alpha, \tag{28}$$

where

$$q(v) = nv^{n-1}, \quad 0 < v < 1, \tag{29}$$

$$Q(u) = \int_0^u q(v)dv = u^n. \tag{30}$$

B. Analytical Solution of the Problem

$$\frac{dL(q_1,q_2)}{dq_1} = \zeta(q_2)\frac{d(q_2)}{d(q_1)} - \zeta(q_1) = \zeta(q_2)\frac{q(q_1)}{q(q_2)} - \zeta(q_1) =$$

$$= q_2^{-2}\frac{nq_1^{n-1}}{nq_2^{n-1}} - q_1^{-2} = \frac{q_1^{n-1}}{q_2^{n+1}} - \frac{1}{q_1^2} = \frac{q_1^{n+1} - q_2^{n+1}}{q_1^2 q_2^{n+1}} < 0, \tag{31}$$

where $q_2 > q_1$. It follows from (28) that

$$1-\alpha < Q(q_2) \le 1 \quad (\text{or } (1-\alpha)^{1/n} < q_2 \le 1). \tag{32}$$

For this example, the version 2 is an adequate version of possible decision making.

C. Optimal Analytical Results

It follows from the above that the minimum of $L(q_1,q_2)$ occurs at $q_2 = 1$, $q_1 = \alpha^{1/n}$. Thus, the shortest length $(1-\alpha)$ - confidence interval is given by

$$L(q_1,q_2) = \frac{1}{\alpha^{1/n}} - 1. \tag{33}$$

If, say, n=3, α=0.1, then $L(q_1,q_2) = 1.154435$. Note that $L(q_1,q_2)$ goes to 0 as $n \to \infty$.

D. New Simple Computation Method: the Problem Statement in Terms of the Decision Variable p (Probability) and Quantile Functions q_1 and q_2

Minimize

$$L^2(q_1,q_2) = \left(\int_{q_1}^{q_2} \zeta(\tau)d\tau\right)^2 = \left(\int_{q_1}^{q_2} \tau^{-2}d\tau\right)^2 =$$

$$= \left(q_1^{-1} - q_2^{-1}\right)^2 = \left(\frac{1}{p^{1/n}} - \frac{1}{(1-\alpha+p)^{1/n}}\right)^2, \tag{34}$$

where the p–quantile function q_1 of V is given by

$$q_1 = Q^{-1}(p) = p^{1/n} \tag{35}$$

and the $(1-\alpha+p)$ -quantile function q_2 of V is given by

$$q_2 = Q^{-1}(1-\alpha+p) = (1-\alpha+p)^{1/n}, \tag{36}$$

subject to

$$0 \le p \le \alpha. \tag{37}$$

E. Numerical Solutions

The optimal numerical solution minimizing $L(q_1,q_2)$ can be obtained using the computer software "Solver". If, for example, n=3, α=0.1, then the optimal numerical solution is given by

$$p = 0.1, \quad q_1 = 0.464159, \quad q_2 = 1 \tag{38}$$

with the 100(1−α)% shortest-length confidence interval

$$L(q_1,q_2) = 1.154435. \tag{39}$$

The 100(1−α)% equal tails confidence interval is given by

$$L(q_1,q_2 \mid p = \alpha/2) = 1.697173 \tag{40}$$

with

$$p = 0.05, \quad q_1 = 0.368403, \quad q_2 = 0.983047572. \tag{41}$$

F. Inference

The proposed method correctly recognized the adequate version 2 of possible decision making and gave accurate numerical results.

G. Relative Efficiency

The relative efficiency of $L(q_1,q_2 \mid p = \alpha/2)$ as compared with $L(q_1,q_2)$ is given by

$$\text{rel.eff.}_L\left\{ L(q_1,q_2 \mid p = a/2), L(q_1,q_2) \right\} =$$

$$= \frac{L(q_1,q_2)}{L(q_1,q_2 \mid p = a/2)} = \frac{1.154435}{1.697173} = 0.680211. \tag{42}$$

V. NUMERICAL EXAMPLE 2

A. Classical Analytical Approach: The Problem Statement in Terms of the Pivot V and Decision Variables q_1 and q_2

Minimize

$$L(q_1,q_2)=\int_{q_1}^{q_2}\zeta(\tau)d\tau=\int_{q_1}^{q_2}\tau^{-2}d\tau=q_1^{-1}-q_2^{-1}. \qquad (43)$$

subject to

$$\int_{q_1}^{q_2}q(v)dv=Q(q_2)-Q(q_1)=1-\alpha, \qquad (44)$$

where

$$q(v)=\frac{1}{2^{n/2}\Gamma(n/2)}v^{n/2-1}\exp\left(-\frac{v}{2}\right),\quad v>0,\ \ n>0, \qquad (45)$$

$$Q(u)=\int_{-\infty}^{u}q(v)dv. \qquad (46)$$

B. Analytical Solution of the Problem

It follows from (8) and (43) that

$$\frac{dL(q_1,q_2)}{dq_1}=\zeta(q_2)\frac{d(q_2)}{d(q_1)}-\zeta(q_1)=q_2^{-2}\frac{q(q_1)}{q(q_2)}-q_1^{-2}, \qquad (47)$$

which vanishes if

$$q_2^{-2}\frac{q(q_1)}{q(q_2)}=q_1^{-2}. \qquad (48)$$

For this example, the version 1 is an adequate version of possible decision making.

C. Optimal Analytical Results

It follows from (48) that the optimal solution is given by

$$q_1^2q(q_1)=q_2^2q(q_2). \qquad (49)$$

Numerical results giving values of q_1 and q_2 to four significant places of decimals are available (see Tate and Klett [3]).

D. New Simple Computation Method: the Problem Statement in Terms of the Decision Variable p (Probability) and Quantile Functions q_1 and q_2

Minimize

$$L^2(q_1,q_2)=\left(\int_{q_1}^{q_2}\zeta(\tau)d\tau\right)^2=\left(\int_{q_1}^{q_2}\tau^{-2}d\tau\right)^2=\left(q_1^{-1}-q_2^{-1}\right)^2, \qquad (50)$$

where the p–quantile function q_1 of V is given (via Excel software: CHISQ.INV (probability p, deg freedom n)) by

$$q_1=Q^{-1}(p), \qquad (51)$$

and the $(1-\alpha+p)$ -quantile function q_2 of V is given (via Excel software: CHISQ.INV (probability $1-\alpha+p$, deg freedom n)) by

$$q_2=Q^{-1}(1-\alpha+p), \qquad (52)$$

subject to

$$0\le p<\alpha. \qquad (53)$$

E. Numerical Solutions

The optimal numerical solution minimizing $L(q_1,q_2)$ can be obtained using the computer software "Solver". If, for example, n=3, α=0.1, then the optimal numerical solution is given by

$$p=0.099478,\quad q_1=0.58208,\quad q_2=17.63810464,$$

$$q(q_1)=0.227512,\quad q(q_2)=0.000248 \qquad (54)$$

with

$$q_1^2q(q_1)=q_2^2q(q_2)=0.077. \qquad (55)$$

and the 100(1−α)% shortest-length confidence interval

$$L(q_1,q_2)=1.661282. \qquad (56)$$

The 100(1−α)% equal tails confidence interval is given by

$$L(q_1,q_2\mid p=\alpha/2)=2.714186 \qquad (57)$$

with

$$p=0.05,\quad q_1=0.351846,\quad q_2=7.814728,$$

$$q(q_1)=0.198465,\quad q(q_2)=0.022409. \qquad (58)$$

F. Inference

The proposed method correctly recognized the adequate version 1 of possible decision making and gave accurate numerical results.

G. Relative Efficiency

The relative efficiency of $L(q_1,q_2\mid p=\alpha/2)$ as compared with $L(q_1,q_2)$ is given by

$$\text{rel.eff.}_L\left\{\,L(q_1,q_2\mid p=a/2),L(q_1,q_2)\,\right\}=$$

$$=\frac{L(q_1,q_2)}{L(q_1,q_2\mid p=a/2)}=\frac{1.661282}{2.714186}=0.612074. \qquad (59)$$

VI. CONCLUSION

The novel unified computation technique proposed in this paper represents the conceptually simple, efficient and useful method for constructing exact statistical (or expected) shortest-length or equal tails confidence intervals in terms of pivotal quantities and quantile functions. The exact confidence intervals with the shortest length or equal tails can be found easily and quickly. Applying the proposed novel unified computation technique, we are not in need to use the

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

following: 1) analytical recognition and computational confirmation of adequate versions of possible solutions, 2) tables for numerical solutions, 3) more advanced methods and 4) special computer programs. For example, the special computer program for (49) is given below:

Minimize

$$z = [q_1^2 q(q_1) - q_2^2 q(q_2)]^2 \qquad (60)$$

where the p–quantile function q_1 of V is given (via Excel software: CHISQ.INV (probability p, deg freedom n)) by

$$q_1 = Q^{-1}(p), \qquad (61)$$

and the $(1-\alpha+p)$ -quantile function q_2 of V is given (via Excel software: CHISQ.INV (probability 1-α+p, deg freedom n)) by

$$q_2 = Q^{-1}(1-\alpha+p), \qquad 62)$$

subject to

$$0 \le p < \alpha. \qquad (63)$$

The optimal numerical solution minimizing z can be obtained using the computer software "Solver". If, for example n=3, α=0.1, then the optimal numerical solution is given by

$$p = 0.099478, \quad q_1 = 0.58208, \quad q_2 = 17.63813,$$

$$q(q_1) = 0.227512, \quad q(q_2) = 0.000248,$$

$$q_1^2 q(q_1) = q_2^2 q(q_2) = 0.077085. \qquad (64)$$

with the 100(1−α)% shortest-length confidence interval

$$L(q_1, q_2) = q_1^{-1} - q_2^{-1} = 1.661282. \qquad (65)$$

The main advantage of the proposed technique is that it includes only one decision variable (probability p) for making decisions under constraints. In other words, the two decision variables q_1 and q_2 are reduced to one decision variable (probability p). This technique greatly simplifies the problem of constructing shortest-length or equal tails confidence intervals for unknown parameters of various distributions and is a novelty in the theory of statistical decisions regarding confidence intervals. It allows one to exclude unknown (nuisance) parameters from the problem using the technique of invariant statistical embedding and averaging in terms of pivotal quantities (ISE & APQ) [4–14].

The unified computation method described in the paper is illustrated in detail for some selected cases. Applications of this method to construct shortest-length or equal tails confidence intervals for unknown parameters of log-location-scale or other probability distributions can follow directly.

REFERENCES

[1] R. C. Juola, “More on shortest confidence intervals,” vol. 47, pp. 117–119, 1993.

[2] K. N. Nechval, N. A. Nechval, E. K. Vasermanis, V. Y. Makeev, “Constructing shortest-length confidence intervals,” Transport and Telecommunication, vol. 3, pp. 95–103, 2002.

[3] R. F. Tate and G. W. Klett, “Optimum confidence intervals for the variance of a normal distribution,” J. Am. Stat. Assoc., vol. 54, pp. 674 – 682, 1959.

[4] N. A. Nechval, G. Berzins, M. Purgailis, and K. N. Nechval, “Improved estimation of state of stochastic systems via invariant embedding technique,” WSEAS Transactions on Mathematics, vol. 7, pp. 141–159, 2008.

[5] N. A. Nechval, K. N. Nechval, V. Danovich, and T. Liepins, “Optimization of new-sample and within-sample prediction intervals for order statistics,” in Proceedings of the 2011 World Congress in Computer Science, Computer Engineering, and Applied Computing, WORLDCOMP'11, 18-21 July, 2011, Las Vegas, Nevada, USA, pp. 91–97.

[6] N. A. Nechval, G. Berzins, V. Danovics, “Optimization of statistical decisions for age replacement problems via a new pivotal quantity averaging approach,” American J. of Theoretical and Applied Statistics, vol. 5, pp. 21–28, 2016.

[7] N. A. Nechval, G. Berzins, S. Balina, I. Steinbuka, and K. N. Nechval, “Constructing unbiased prediction limits on future outcomes under parametric uncertainty of underlying models via pivotal quantity averaging approach,” Autom. Control Comput. Sci., vol. 51, pp. 331–346, 2017.

[8] N. A. Nechval, G. Berzins, K. N. Nechval, “Intelligent planning reliability-based inspections of fatigued structures for the crack initiation period in the Weibull case under parametric uncertainty,” Autom. Control Comput. Sci., vol. 52, pp. 184–197, 2018.

[9] N. A. Nechval, K. N. Nechval, and G. Berzins, “A new technique for intelligent constructing exact γ-content tolerance limits with expected (1−α)-confidence on future outcomes in the Weibull case using complete or type II censored data,” Autom. Control Comput. Sci., vol. 52, pp. 476–488, 2018.

[10] N. A. Nechval, G. Berzins, K. N. Nechval, and J. Krasts, “A new technique of intelligent constructing unbiased prediction limits on future order statistics coming from an inverse Gaussian distribution under parametric uncertainty,” Autom. Control Comput. Sci., vol. 53, pp. 223–235, 2019.

[11] N. A. Nechval, G. Berzins, and K. N. Nechval, “A novel intelligent technique for product acceptance process optimization on the basis of misclassification probability in the case of log-location-scale distributions,” F. Wotawa et al. (Eds.) Advances and Trends in Artificial Intelligence. From Theory to Practice, Lecture Notes in Computer Science, vol. 11606, pp. 801–818, Springer Nature Switzerland AG, 2019.

[12] N. A. Nechval, G. Berzins, and K. N. Nechval, “A novel intelligent technique of invariant statistical embedding and averaging via pivotal quantities for optimization or improvement of statistical decision rules under parametric uncertainty,” WSEAS Transactions on Mathematics, vol. 19, pp. 17–38, 2020.

[13] N. A. Nechval, G. Berzins, K. N. Nechval, V. Danovics, “Intelligent constructing optimal airline seat protection levels for multiple nested fare classes of single-leg flights,” J. of Physics: Conference Series, vol. 1603, 2020, 7 pages, 012023.

[14] N. A. Nechval, G. Berzins, and K. N. Nechval, “A new technique of invariant statistical embedding and averaging via pivotal quantities for intelligent constructing efficient statistical decisions under parametric uncertainty,” Automatic Control and Computer Sciences, vol. 54, pp. 191–206, 2020.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Blockchain Systems Review and Analysis for Information Security of Big Data

Igor Zakharov
GeoAnalytics
C-CORE
Ottawa, ON, Canada
igor.zakharov@c-core.ca

Jonathan Anderson
Electrical and Computer Engineering
Memorial University
St. John's, NL, Canada
jonathan.anderson@mun.ca

Garrett Parsons
GeoAnalytics
C-CORE
Ottawa, ON, Canada
garrett.parsons@c-core.ca

Michael Henschel
GeoAnalytics
C-CORE
Ottawa, ON, Canada
michael.henschel@c-core.ca

Abstract. **Data quantities are rapidly increasing in many industry sectors due to the development of new sensors, mobile and cloud technologies, advancements in IoT and AI, and growth of social and entertainment media. Many applications (e.g. in finance, healthcare, government data) have strict information security requirements for simultaneous access, record updating, and validation in an immutable manner, which can be achieved with distributed ledger technology (DLT). In this paper we review and analyze fifty-eight currently available blockchain (BC) systems and their components in context of the DLT for big data storage, provenance tracking, replication, and sharing. We also elaborate the key BC system components and architecture for major information security concerns.**

Keywords: **information security, DLT, blockchain, big data**

I. INTRODUCTION

Data volumes are rapidly increasing across different industries and governments due to the development of new sensors, mobile and cloud technologies, advancements in IoT and AI, and growth of social and entertainment media. The amount of data created, captured, copied, and consumed within an organization can reach volumes on the order of 100s of terabytes to multi-petabytes [1] with millions and even billions of records approaching to big data scales (beyond computing power of modern data centers).

Information security measures to protect data from internal and external cyber threats include access control (authentication and authorization), data integrity maintenance, encryption and digital signatures, and monitoring [2, 3]. Beyond these traditional measures, centralized digital ledger recordkeeping enables verifiable transaction logs to be created and maintained by a central authority to simplify auditing mechanism [4].

In contrast to the centralized ledger, the distributed ledger is a decentralized database which is synchronized and accessible across different users (nodes) on a network. In multiple applications, such as finance, legal services, medicine, earth observation etc., the distributed ledger technology (DLT) can be used to meet requirements in secured transactions, simultaneous data access, validation, record updating and storing in an immutable manner [5, 6]. The immutability helps DLTs access a different point in the consistency-availability-protection (CAP) trade-off space [7, 8] than the traditional solutions [9]. In addition to the key mechanisms, such as, database, consensus algorithm, peer-to-peer (P2P) network and transactions logging, the main DLT features also include [10]:

- immutability (new data can only be appended, but not changed or deleted), and
- cryptography.

Distributed ledgers enable to form and maintain consensus about the existence, status and evolution of a set of shared facts [11]. The major differences between distributed ledgers and traditional distributed databases (distributed across sites in a network) is the use of an adversarial threat model and a very different view of authority.

Blockchains (BCs) can be considered to be a subset of distributed ledgers that share the same adversarial threat model (assuming that not all nodes are trustworthy) over a P2P network and have additional characteristics, such as [10], linked blocks which chronologically stored transactional data.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

In general, the implementation of a custom, private permissioned blockchain from scratch is a difficult task [12] and therefore a comprehensive survey of currently available BCs is essential to achieve advanced performance. The goal of our paper is to review and analyze the currently available BC systems in context of the DLT component configuration for big data management. The novelty of the paper also lies in the elaboration of possibilities for BC implementation for information security of big data.

II. BLOCKCHAIN

A. *BC Surveys*

Numerous publications provided comprehensive and systematic literature reviews of BC technology analyzing its concept, architecture, components and implementation focusing on its:

- evolution and architecture [13, 14],
- performance evaluation [15],
- identity management [16],
- post-quantum cryptography [17],
- use by cryptocurrencies [18],
- industrial IoT applications [19],
- healthcare application [20],
- data management [21],
- engineering and manufacturing applications [22],
- supply chain management applications [23],
- connection with cloud computing [24],
- connection with communication networks [25],
- and other numerous applications [26–30].

A survey on BCs for several smart applications (city, healthcare, transportation and grid), which can generate big data [32], discussed approaches, opportunities, challenges and future directions for secure big data acquisition, data storage, data analytics, and data privacy preservation. BC architecture for massive data storage was recently analyzed and implemented [31] to improve scalability and performance. Various aspects of BC-enabled cyber-physical systems, including security, privacy, immutability, fault tolerance, interoperability, data provenance, atomicity, automation, data/service sharing, and trust were reviewed in [33]. BCs can be categorized as permissionless and permissioned (requiring authorization). Based on the usage and ownership the BCs can be divided [29] into public, private and consortium (to record cross-organizational transactions).

B. *Block Structure*

The block structure includes a block header and block body [29]. The block header specifies the metadata, including various fields:

- hash (ID) of previous block (to connect its previous block called a parent block),
- hash of current block,
- timestamp (creation time of the block),
- Nonce (relates to consensus mechanism for validation),
- Merkle root (to store the transactions for efficient data verification).

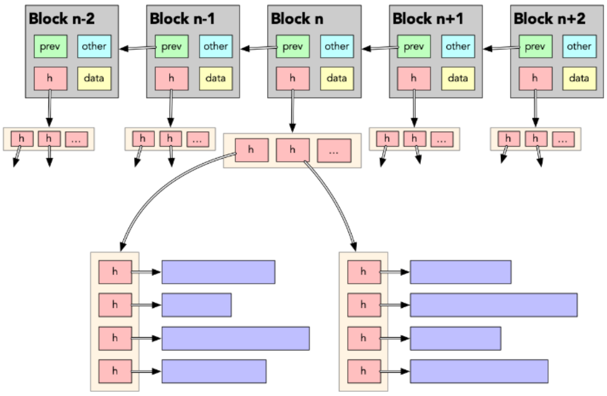


Fig. 1. A typical BC structure

The block's header may also contain other information, for example, block version (software/protocol version), nBits (target threshold of a valid block hash) [34] and confirmation [25]. The block body (also called as block data) stores transactions (work process resulting in a state change) which are assembled using cryptographic functions. All performed transactions (e.g. transfers of money) are hashed and hash values are structured into a Merkle DAG. The linked blocks form a BC (Fig. 1).

C. *BC Architecture*

The three main components which enable BC technology are a decentralized P2P network, distributed consensus and cryptographically secure algorithms [30]. Work [11] highlights importance of ledger and validity rules (when transactions are considered valid and how the ledger gets updated). The basic BC architecture distributed over P2P network is shown in Fig. 2.

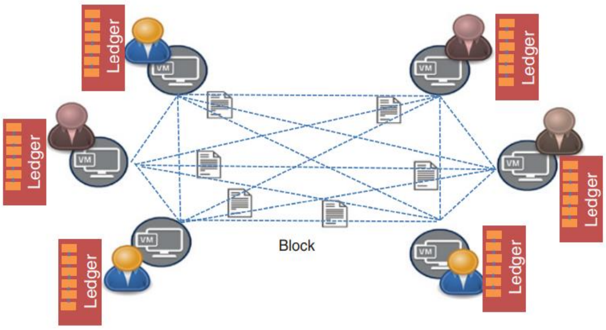


Fig. 2. Blockchain architecure. Adapted from [30]

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

A smart contract is a computer program deployed using cryptographically signed transactions in the distributed ledger (BC network), in which parties agree to be bound by the program's output.

D. Consensus Mechanisms

Consensus algorithms enable agreement among decentralized nodes before a block is included into the blockchain. Since, BC networks are designed to function with no trusted central node [29], there are several major consensus mechanisms: Proof of Work (PoW), Proof of Stake (PoS), Practical Byzantine Fault Tolerance (PBFT), and Delegated Proof of Stake (DPoS). However, BFT still requires either a central authority to identify a BFT node (i.e., permissioning) or PoW to maintain overall consensus. Xie et al. [30] review more consensus mechanisms: Proof of Activity (PoA), Proof of Elapsed Time (PoET), Proof of Luck (PoL), and Proof of Space (PoSpace). Other mechanisms were comprehensively analyzed in different works (e.g. [35]). These include: Multi-signature (majority rule), delayed Proof of Work (DPoW), Proof of Importance, Delegated Proof of Stake (DPoS), Delegated Byzantine Fault Tolerance (DBFT), Partioned Consensus, Smilo BFT+, Stellar Consensus Protocol, Proof of Stake Velocity (PoSV), Proof of Burn (PoB), Proof of History (PoH), Proof of Importance (PoI), Proof of Believability (PoBelievability), Federated Byzantine Agreement (FBA), Combined DPoS+BFT, Proof of Authority (PoAuthority), and Raft.

Some consensus mechanisms have multiple in-service implementations. For example, PoS has several implementations [25], such as, Chain of Activity, Casper, Algorand, and Tendermint.

III. INFORMATION SECURITY WITH BC

BC is not automatically more secure than other DLT systems. Although DLT makes it difficult to manipulate and attack by a single node, the major BC vulnerabilities are related to [30]:

- fork (intentional or unintentional condition in which nodes in the network have diverging views),
- stale or orphaned blocks,
- different attacks (e.g. 51%, DNS, DDoS, selfish mining, consensus delay, double-spending).

Some of these vulnerabilities are specific to cryptocurrency applications and associated with public permissionless BCs. In permissioned BC, multiple validation nodes are trusted to maintain the consensus. To avoid certain nodes accepting inconsistent messages, consensus mechanisms (e.g. PBFT, PoAuthority, DPoS+BFT) are used; decisions are encoded using digital signatures with algorithms such as Elliptic Curve Cryptography and SHA3-256 [30].

IV. BC SYSTEMS FOR BIG DATA

A. Implementations of BC and Other DLT Systems

The key aspects described in the above sections (block structure and BC architecture) have to be implemented and customized based on the type of application and considering the data to be stored, secured and managed with the BC. The block components, algorithms and protocols have to follow the latest standards and meet the defined requirements. The BC implementation has to take into account vulnerabilities and the performance of the hash function considering the data size and file system.

TABLE I. IMPLEMENTATIONS OF BC AND OTHER DLT SYSTEMS

Platform/Framework	**Platform/Framework**
1. Aion	30. Komodo
2. ArcBlock	31. Lisk
3. Ardor	32. MedRec
4. BigchainDB	33. MediChain
5. Binded	34. MultiChain
6. Bitcoin	35. ModelChain
7. BlocHIE	36. Neblio
8. Burst	37. NEM
9. Cardano	38. NEO
10. Chain core	39. Nxt
11. Corda	40. OmniPHR
12. Credits	41. OpenChain
13. Democracy Earth	42. Oracle
14. Elements	43. Qtum
15. Enigma	44. Quorum
16. EOS	45. Paperchain
17. Ethereum	46. Parity
18. Exonum	47. Po.et
19. Follow My Vote	48. Propy
20. Graphene	49. Stellar
21. Guardtime	50. Straitis
22. Hydrochain	51. Tezos
23. Hyperledger Fabric	52. Ubitquity
24. Hyperledger Indy	53. Ujomusic
25. Hyperledger Burrow	54. Verisart
26. Hyperledger Sawtooth	55. Wanchain
27. ICON	56. Warranteer
28. IOTA	57. Waves
29. KSI Blockchain	58. Zilliqa

Multiple resources (e.g. [35, 36]) analyzed available BC and other DLT frameworks and platforms. Their application areas are varying between finance, government, IoT, legal and other industries. A comprehensive list of BC and DLT systems identified on July 15, 2021 from internet and research publications is provided in Table I. This list includes both frameworks (e.g. Ethereum, Hyperledger, Qtum) and the implementations of BCs for specific applications. These implementations were analyzed in context of their applicability and benefits to big data applications (details are presented at the conference). The analysis also considered different factors affecting the implementation, for example, programming language, support documentation, and availability (commercial or open source).

B. BC Systems for Big Data

The BC system can be implemented for increasing level of information security of large and growing data volumes in data centers and cloud environments. The largest cloud service providers (e.g. AWS, Azure, IBM, Oracle) have already developed environments with blockchain as a service (BaaS). In a cloud environment, the usage of BC enables multiple nodes to participate in maintaining transparent and immutable provenance information for tracking data transactions and detect malicious activities [30]. The API and web user interface can be used to monitor BC activity and visualize provenance records. The BC system applications for sharing and enabling integrity of large data files such as medical [37] and Earth Observation [38] images provide examples of the required definition and design activities.

An important problem in BC is scalability, which coincides with growing data volumes. An increasing number of transactions (and block validations) from an increasing number of users, leads to communication overheads that limit the network scalability [35]. BC systems for massive data storage can potentially enable scalability with acceptable performance [25]. To address issue of redundant massive data storage, a secure data storage and recovery scheme in the BC-based network was proposed [39] by improving the real-time monitoring, and supporting the dynamic storage, and update of distributed data.

The main development activities to achieve the best performance and security capabilities of BC system are currently focused on definition and design of its technological elements, including block structure and information, hash function, encryption algorithm, consensus mechanism and integration in file system. The architecture of the BC system and its components for big data were analyzed (presented at the conference with more details).

V. CONCLUSION

This paper reviewed and analyzed current concepts, applications and implementations of BC systems. In total, 58 BC platforms and frameworks were identified. The main elements of a BC system to enable information security of big data were analyzed and defined.

REFERENCES

[1] NASA OCIO, “The Big Data Wave,” IT Talk, vol. 8, no. 3, Sep. 2018. Accessed: Jul. 15, 2021. [Online]. Available: https://www.nasa.gov/sites/default/files/atoms/files/365378_it_talk_design_-_july_2018_0.pdf

[2] M. Stamp, Information security: principles and practice, 2nd ed. Hoboken, NJ: Wiley, 2011.

[3] M. Copeland and M. Jacobs, Cyber Security on Azure: An IT Professional’s Guide to Microsoft Azure Security. Berkeley, CA: Apress, 2021. doi: 10.1007/978-1-4842-6531-4.

[4] B. Camci, S. Bayar, and M. G. Ulkar, “A simple auditing mechanism for financial reports in e-Ledger project,” in 2015 9th International Conference on Application of Information and Communication Technologies (AICT), Rostov on Don, Russia, Oct. 2015, pp. 244–248. doi: 10.1109/ICAICT. 2015.7338555.

[5] World Bank, Distributed Ledger Technology and Secured Transactions. World Bank, Washington, DC, 2020. doi: 10.1596/34007.

[6] M. M. Akhtar, D. R. Rizvi, M. A. Ahad, S. S. Kanhere, M. Amjad, and G. Coviello, “Efficient Data Communication Using Distributed Ledger Technology and IOTA-Enabled Internet of Things for a Future Machine-to-Machine Economy,” Sensors, vol. 21, no. 13, p. 4354, Jun. 2021, doi: 10.3390/s21134354.

[7] E. A. Brewer, “Towards robust distributed systems (abstract),” in Proceedings of the nineteenth annual ACM symposium on Principles of distributed computing - PODC ’00, Portland, Oregon, United States, 2000, p. 7. doi: 10.1145/343477.343502.

[8] S. Gilbert and N. Lynch, “Perspectives on the CAP Theorem,” Computer, vol. 45, no. 2, pp. 30–36, Feb. 2012, doi: 10.1109/MC.2011.389.

[9] M. Litoiu et al., “How do I choose the right NoSQL solution? A comprehensive theoretical and experimental survey,” Big Data Inf. Anal., vol. 1, no. 2/3, pp. 185–216, Sep. 2016, doi: 10.3934/bdia.2016004.

[10] M. Lange, S. C. Leiter, R. Alt, “Defining and Delimitating Distributed Ledger Technology: Results of a Structured Literature Analysis,” in Business Process Management: Blockchain and Central and Eastern Europe Forum, vol. 361, C. Di Ciccio, R. Gabryelczyk, L. García-Bañuelos, T. Hernaus, R. Hull, M. Indihar Štemberger, A. Kő, and M. Staples, Eds. Cham: Springer International Publishing, 2019, pp. 43–54. doi: 10.1007/978-3-030-30429-4_4.

[11] G. Hileman and M. Rauchs, “2017 Global Blockchain Benchmarking Study,” SSRN Electron. J., 2017, doi: 10.2139/ssrn.3040224.

[12] F. Knirsch, A. Unterweger, and D. Engel, “Implementing a blockchain from scratch: why, how, and what we learned,” EURASIP J. Inf. Secur., vol. 2019, no. 1, p. 2, Dec. 2019, doi: 10.1186/s13635-019-0085-3.

[13] M. N. M. Bhutta et al., “A Survey on Blockchain Technology: Evolution, Architecture and Security,” IEEE Access, vol. 9, pp. 61048–61073, 2021, doi: 10.1109/ACCESS.2021.3072849.

[14] W. Yang, E. Aghasian, S. Garg, D. Herbert, L. Disiuta, and B. Kang, “A Survey on Blockchain-Based Internet Service Architecture: Requirements, Challenges, Trends, and Future,” IEEE Access, vol. 7, pp. 75845–75872, 2019, doi: 10.1109/ACCESS.2019.2917562.

[15] C. Fan, S. Ghaemi, H. Khazaei, and P. Musilek, “Performance Evaluation of Blockchain Systems: A Systematic Survey,” IEEE Access, vol. 8, pp. 126927–126950, 2020, doi: 10.1109/ACCESS.2020.3006078.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

[16] T. Rathee, P. Singh, “A systematic literature mapping on secure identity management using blockchain technology,” J. King Saud Univ. Comput. Inf. Sci., p. S1319157821000690, Mar. 2021, doi: 10.1016/j.jksuci.2021.03.005.

[17] T. M. Fernandez-Carames and P. Fraga-Lamas, “Towards Post-Quantum Blockchain: A Review on Blockchain Cryptography Resistant to Quantum Computing Attacks,” IEEE Access, vol. 8, pp. 21091–21116, 2020, doi: 10.1109/ACCESS.2020.2968985.

[18] M. H. ur Rehman, K. Salah, E. Damiani, and D. Svetinovic, “Trust in Blockchain Cryptocurrency Ecosystem,” IEEE Trans. Eng. Manag., vol. 67, no. 4, pp. 1196–1212, Nov. 2020, doi: 10.1109/TEM.2019.2948861.

[19] T. Alladi, V. Chamola, R. M. Parizi, and K.-K. R. Choo, “Blockchain Applications for Industry 4.0 and Industrial IoT: A Review,” IEEE Access, vol. 7, pp. 176935–176951, 2019, doi: 10.1109/ACCESS.2019.2956748.

[20] S. Namasudra and G. C. Deka, Eds., Applications of Blockchain in Healthcare, vol. 83. Singapore: Springer Singapore, 2021. doi: 10.1007/978-981-15-9547-9.

[21] H.-Y. Paik, X. Xu, H. M. N. D. Bandara, S. U. Lee, S. K. Lo, “Analysis of Data Management in Blockchain-Based Systems: From Architecture to Governance,” IEEE Access, vol. 7, pp. 186091–186107, 2019, doi: 10.1109/ACCESS.2019.2961404.

[22] J. E. Kasten, “Engineering and Manufacturing on the Blockchain: A Systematic Review,” IEEE Eng. Manag. Rev., vol. 48, no. 1, pp. 31–47, Mar. 2020, doi: 10.1109/EMR.2020.2964224.

[23] S. E. Chang and Y. Chen, “When Blockchain Meets Supply Chain: A Systematic Literature Review on Current Development and Potential Applications,” IEEE Access, vol. 8, pp. 62478–62494, 2020, doi: 10.1109/ACCESS.2020.2983601.

[24] K. Gai, J. Guo, L. Zhu, and S. Yu, “Blockchain Meets Cloud Computing: A Survey,” IEEE Commun. Surv. Tutor., vol. 22, no. 3, pp. 2009–2030, 2020, doi: 10.1109/COMST.2020.2989392.

[25] M. H. Rehmani, Blockchain Systems and Communication Networks: From Concepts to Implementation. Cham: Springer International Publishing, 2021. doi: 10.1007/978-3-030-71788-9.

[26] D. Cagigas, J. Clifton, D. Diaz-Fuentes, and M. Fernandez-Gutierrez, “Blockchain for Public Services: A Systematic Literature Review,” IEEE Access, vol. 9, pp. 13904–13921, 2021, doi: 10.1109/ACCESS.2021.3052019.

[27] C. Shen and F. Pena-Mora, “Blockchain for Cities–A Systematic Literature Review,” IEEE Access, vol. 6, pp. 76787–76819, 2018, doi: 10.1109/ACCESS.2018.2880744.

[28] T. Ali Syed, A. Alzahrani, S. Jan, M. S. Siddiqui, A. Nadeem, and T. Alghamdi, “A Comparative Analysis of Blockchain Architecture and its Applications: Problems and Recommendations,” IEEE Access, vol. 7, pp. 176838–176869, 2019, doi: 10.1109/ACCESS.2019.2957660.

[29] J. Xie et al., “A Survey of Blockchain Technology Applied to Smart Cities: Research Issues and Challenges,” IEEE Commun. Surv. Tutor., vol. 21, no. 3, pp. 2794–2830, 2019, doi: 10.1109/COMST.2019.2899617.

[30] S. S. Shetty, C. A. Kamhoua, and L. L. Njilla, Blockchain for Distributed Systems Security. 2019. Accessed: Jul. 14, 2021. [Online]. Available: http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9781119519591

[31] X. Chen, K. Zhang, X. Liang, W. Qiu, Z. Zhang, and D. Tu, “HyperBSA: A High-Performance Consortium Blockchain Storage Architecture for Massive Data,” IEEE Access, vol. 8, pp. 178402–178413, 2020, doi: 10.1109/ACCESS.2020.3027610.

[32] N. Deepa et al., “A Survey on Blockchain for Big Data: Approaches, Opportunities, and Future Directions,” ArXiv200900858 Cs, Feb. 2021, Accessed: Jul. 18, 2021. [Online]. Available: http://arxiv.org/abs/2009.00858

[33] W. Zhao, C. Jiang, H. Gao, S. Yang, and X. Luo, “Blockchain-Enabled Cyber–Physical Systems: A Review,” IEEE Internet Things J., vol. 8, no. 6, pp. 4023–4034, Mar. 2021, doi: 10.1109/JIOT.2020.3014864.

[34] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, “An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends,” in 2017 IEEE International Congress on Big Data (BigData Congress), Honolulu, HI, USA, Jun. 2017, pp. 557–564. doi: 10.1109/BigDataCongress.2017.85.

[35] Ismail and Materwala, “Article A Review of Blockchain Architecture and Consensus Protocols: Use Cases, Challenges, and Solutions,” Symmetry, vol. 11, no. 10, p. 1198, Sep. 2019, doi: 10.3390/sym11101198.

[36] TechnoDuet, “A Comprehensive List of Blockchain Platforms,” 2021. technoduet.com (accessed Jul. 15, 2021).

[37] M. Sultana, A. Hossain, F. Laila, K. A. Taher, M. N. Islam, “Towards developing a secure medical image sharing system based on zero trust principles and blockchain technology,” BMC Med. Inform. Decis. Mak., vol. 20, no. 1, pp. 256, Dec. 2020, doi: 10.1186/s12911-020-01275-y.

[38] A. Burzykowska, M. Iapaolo, A. Priit, and A. Sisask, “EO Data Provenance with KSI Blockchain,” ESA, Issue Brief. ESA Blockchain / Distributed Ledgers and EO Community of Practice (CoP), 2020.

[39] W. Liang, Y. Fan, K.-C. Li, D. Zhang, and J.-L. Gaudiot, “Secure Data Storage and Recovery in Industrial Blockchain Network Environments,” IEEE Trans. Ind. Inform., vol. 16, no. 10, pp. 6543–6552, Oct. 2020, doi: 10.1109/TII.2020.2966069.

A Digital Platform for Processing Fluorescence Spectroscopy Data Using Simulation Modelling and Machine Learning Algorithms

Mikalai Yatskou
Dept. of Systems Analysis and Computer Modelling
Belarusian State University
Minsk, Belarus
yatskou@bsu.by

Vladimir Apanasovich
Dept. of Systems Analysis and Computer Modelling
Belarusian State University
Minsk, Belarus
apanasovichv@gmail.com

Abstract. **A digital computational platform is proposed for processing fluorescence spectroscopy data, which implements complex analysis of experimental information based on the simulation modelling and machine learning algorithms. Data analysis includes partitioning biophysical data into clusters according to the degree of likeness in some measure of similarity, finding the median cluster members (medoids), applying the data reduction method and visualizing the experimental data in a two-dimensional space. Analysis of the medoids is carried out by the analytical or simulation models of optical processes occurring in molecular systems. The visualization of data clusters in the original and transformed feature spaces is done with the aim of user interaction. As a demonstrative example, the platform FluorSimStudio is implemented for processing time-resolved fluorescence measurements (https://dsa-cm.shinyapps.io/FluorSimStudio). The digital platform is an open system and allows addition of complex analysis models, taking into account the development of new modelling and analysis algorithms.**

Keywords: **fluorescence spectroscopy, simulation modelling, machine learning, digital platform**

I. INTRODUCTION

Experimental fluorescence spectroscopy methods are applied to study the optical properties of molecular compounds and are commonly used in the studies of artificial photonic materials, protein complexes, biopolymers, DNA sequencing, biological membranes, cell and tissues, medical diagnostics [1]. The considerable development of methods is driven due to the improvements of effective molecular fluorophores, including genetically expressed proteins (for example, GFP), semiconductor nanoparticles and quantum dots, optical systems for laser excitation and registration of radiation, allowing high-precision measurements, computer technologies for data storage and processing [2]. Novel experimental high-throughput techniques, integrating pulsed, phase and modulation methods for recording fluorescence decay times, form the basis of modern fluorescence microscopy and allow obtaining big data, characterized by high spectral, time and spatial resolution [3]. The main fluorescence spectroscopy and microscopy techniques for studying complex molecular systems in "cuvettes" and living cells are fluorescence-lifetime imaging microscopy (FLIM), fluorescence recovery after photobleaching (FRAP) and its derivatives – fluorescence loss in photobleaching (FLIP) and fluorescence localization after photobleaching (FLAP), fluorescence fluctuation spectroscopy (FFS, combining fluorescence correlation spectroscopy (FCS), fluorescence cross-correlation spectroscopy (FCCS), photon counting histogram (PCH) and fluorescence intensity distribution analysis (FIDA)), fluorescence sensing (FS) [4].

The existing data analysis approaches to processing fluorescence spectroscopy data can be divided into classical and modern, based on machine learning, algorithms. Classical methods consider separate or joint analysis of datasets using deconvolution, least squares, maximum likelihood, Bayesian, target and global analysis to estimate the parameters of mathematical models of optical processes and systems [5]. New approaches are based on: i) projection transformations and following parameter estimation (for example – transformation of fluorescence intensities into the phasor space (phasor analysis), ii) using machine learning techniques, mainly artificial neural networks and ensemble algorithms, to estimate the model parameters, iii) segmentation of cell or tissue images and subsequent classification by a machine learning algorithm [5, 6]. The main disadvantages of existing data processing methods are limited or poor efficiency, that is due to the use of nonphysical analytical models (multi-exponential or polynomial decompositions), poor accuracy in parameter estimating when analyzing noisy data (phasor analysis, neural networks), slow computations (global and Bayesian analysis), the need for the large training datasets (neural networks), special requirements for computing resources (the usage of

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

video cards or multiprocessor nodes to accelerate neural network computing), and finally the lack of specialized software for automated data processing. Therefore, the primary task is to develop an integrated data analysis approach and computational platform that eliminates the main drawbacks of existing methods, which would include physical models of the processes and systems under study, effective methods and software for processing a series of fluorescence spectroscopy data.

A computational approach for processing large sets of time-resolved fluorescence data using simulation modelling and data mining algorithms was developed [7, 8]. By this methodology it is possible to increase the accuracy of the estimated parameters of biophysical and optical processes occurring in the studied molecular systems. Specialized and general-purpose software tools and products, both commercial and freely available, have been developed for statistical processing, analysis and simulation of fluorescence spectroscopy data. However, there are no unified integrated software tools for processing large datasets using simulation modelling and machine learning methods. The development of a digital software platform for simulation and machine learning analysis of fluorescence data in various biophysical systems under experimental studies is an critically important and urgent task.

In this paper, we propose the conception of a digital software platform for the simulation modelling and machine learning analysis of optical processes in molecular systems studied by the fluorescence spectroscopy methods. As a demonstrative example, developed integrated methodology is implemented into the computational platform FluorSimStudio for processing fluorescence kinetic curves obtained through FLIM experiments.

II. METHODOLOGY

A. *Review of the Computational Tools for a Digital Platform.*

A digital computational platform in this case is an intellectual software resource or a programming environment designed to model and analyze large experimental fluorescence spectroscopy data studied in biophysical research. The platform includes a programming environment, integrated coding languages, software tools for automation, code debugging and creating an application interface, models of research objects, methods for analyzing and visualizing data, assessing the quality of analysis and the reliability of models. The choice of the optimal software platform primarily implies the choice of a programming environment and interface development tools for interacting with the user.

Various computing platforms and programming technologies are used to implement the software. In most publications on benchmarking open access packages, there is no clear leader in machine learning and data mining. Currently, a large number of software tools are actively used, including WEKA, Tanagra, Rapid Miner, KNIME, Orange, Java, Python and R projects, as well as platforms implemented using high-performance programming languages C++ and Scala. The advantages of these software resource are computational performance, a wide range of libraries for statistical analysis, cross-platform integrity, the ability to develop user interfaces, parallel computing, work directly with existing databases and data warehouses. The main disadvantages include the lack of versatility, significant requirements for computing resources, and the limitation of the integration of the above fascinating properties in a single format. The most promising projects for organizing the digital environment are Scala-, Python- and R-platforms. A platform based on the Scala language (for example, Apache Hadoop) is designed to analyze big data in production projects and is used to solve industrial programming problems. Python applications are aimed at solving general engineering and data analysis problems with an emphasis on neural network approaches and programming. R-projects are developed primarily with the aim of optimizing and validating applied statistical analysis, which includes approaches using classical and data mining methods. Let take a closer look at the R environment.

The main advantages of the statistical programming environment R are the presence of optimized structures for representing data objects, which greatly simplifies data processing, optimization of programming tools and implementation of computation algorithms (in the sense of minimizing the introduction of errors into the program code), the ability to use a huge set of processing algorithms, statistical and data mining, various computing resources of the scientific community [9]. The main drawback is the low computational performance in the basic version of the environment layout, which is especially critical when working with large datasets and developing simulation models. This limitation can be partially or completely eliminated by connecting program codes of high-performance programming languages Scala, Java, C++ (packages rscala, rjava, Rcpp, inline), parallel computing procedures (managed by packages parallel, Rmpi, snow, snowfall), additional packages for efficient processing big data (readr, LaF, data.table, ff, bigmemory) and the use of third-party software resources (Microsoft R Open and Intel Math Kernel Library libraries, H2O big data analysis platforms, Apache Hadoop and Spark systems, with using h20, Rhadoop and SparkR packages).

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

An important issue is the development of the interface of a software application. The most popular R-code integrating user interface development packages are gWidgets, rpanel, svDialogs, RGtk2, qtbase, tcltk. A new direction in the development of R-applications for the analysis of biophysical systems [10] is associated with the creation of "reactive" web interfaces using the Shiny package and the subsequent placement of the software implementation on the shinyapps.io resource provided by the open source software developers RStudio. The advantage of this approach is the ability to remotely work with a web application for a wide scientific audience of users online via the global Internet. To implement the software application, the R computing environment and the Shiny package were chosen to create a web interface for the developed application.

The computing platform is organized according to the example of open projects of network resources CRAN (https://cran.r-project.org), R-Forge (https://r-forge.r-project.org), Bioconductor (https://www.bioconductor.org), Github (https://github.com). It is a programming and simulation environment that contains updated and supplemented libraries of analytical and simulation models of optical processes in molecular systems, built-in tools for machine learning methods and assessment of the quality of analysis and modelling, provides the scientific community with opportunities to develop new algorithms and simulation models.

B. *Conception of the Digital Platform.*

The digital platform can integrate the research scheme for a certain biophysical process or molecular compound using a complex approach based on simulation modelling and machine learning methods [7]. A schematic diagram of the methodology for spectral or/and time-resolved fluorescence spectroscopy data analysis of the platform is shown in Fig. 1. Consider the main stages of data analysis.

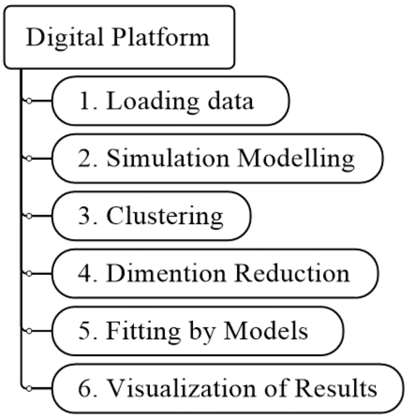


Fig. 1. Main stages of the fluorescence data analysis of a digital platform using simulation modelling and machine learning

The platform is designed to analyze experimental or simulated data. Data loading and graphical presentation is carried out in block 1. Visual assessment of two-dimensional and three-dimensional fluorescence datasets allows predetermining the choice of a mathematical model for describing the physical processes, making a supposition regarding the number of data clusters, and limiting the choice of measures for calculating the similarity of samples based on the noise level of the data.

Modelling and visualization of fluorescence data are carried out in block 2. Integrated models of optical processes are considered. Simulation modelling is carried out using Monte Carlo algorithms [11]. The input characteristics of the simulation are the type and parameters of the model, the number of samples and the number of simulations. 2D or 3D visualization are intended for expert analysis of modeled data, study of the behavior of models when changing their parameters, manual selection of the most optimal modelling parameters, such as the number of simulations and data points, as well as initial approximations of parameters for subsequent precise determination during fitting using mathematical models. New and improved models of optical-physical processes in molecular systems can be developed and integrated into the software environment.

In block 3, cluster analysis of fluorescence data is performed in the space of experimentally detected features. Clusters of data are identified according to some degree of similarity (Euclidean, Minkowski, Manhattan, maximum or Canberra distance). The number of clusters is determined intuitively, automatically from the hierarchy dendrogram of data constructed on the basis of the cluster binding measure (Ward, nearest neighbor, far neighbor, or middle bond), or on the basis of a statistical criterion [12, 13]. The median representatives of the clusters are calculated – medoids, samples or data objects having the smallest average distances to the rest of the objects of the corresponding clusters.

A data reduction is carried out in block 4. The consideration of a large group of uninformative experimentally detected features leads to difficulties in data analysis, namely, to their noise, an increase in the amount of data, and distortion of reliable information about clusters of similar samples. To improve the quality of data analysis, in particular, the visual assessment of data partitioning into clusters, it is expedient to carry out the stage of data analysis, which includes the transition to a low-dimensional space of new informative features, in which the fluorescence data form clusters. To perform this transformation, it is required to use data dimensionality reduction algorithms, among which the method of principal component analysis is the most widely known [14]. Conversion of fluorescence data using principal component analysis is performed. The proportion of relative variation attributed to principal components is set, limiting the number of components. Principal components are selected that correspond to a given variation in the data (for example, 0.95 out of 1). A diagram of the proportions of variation of the first ten principal components is constructed, according to which the contribution to the total variance in the data is estimated. Clusters and their medoids are displayed in the scatter diagram of the first two principal components. Medoids are calculated in the space of initial features or in the space of the main components that explain a given value of variability. For example, if the data clusters are not separated, then it can be assumed that there is only one kind of fluorescent compounds. Otherwise, the presence of several forms of compounds (fluorophores) is allowed. For the convenience of visual control of cluster separability, histograms of frequencies are plotted on the axis of the first three principal components. Good separability of clusters is characterized by the presence of a multimodal form of histogram distributions.

In block 5, cluster medoids are analyzed to accurately determine the parameters of fluorescent compounds using an optimization algorithm and mathematical models. To approximate the fluorescence data, represented by the found medoids, analytical and simulation models for describing photophysical processes are used. Optimization methods are applied for the optimal selection of the parameters of mathematical models during the approximation of experimental data. In this work, the Nelder–Mead method [15] is chosen, which does not take into account the derivative of the objective function, which greatly simplifies the use of simulation models in the parameter estimation procedure. The best approximation is chosen according to a criterion (or a set of criteria) that determines the degree of deviation of the theoretical model from the experimental data. As a rule, such a criterion is presented analytically in the form of a function of experimental and theoretical data, the form of which is determined by the field of application, the direct modelling method and the conditions of the experiment. In our experiments, we consider the normalized chi-square criterion, diagrams of weighted residuals and their autocorrelation function [12].

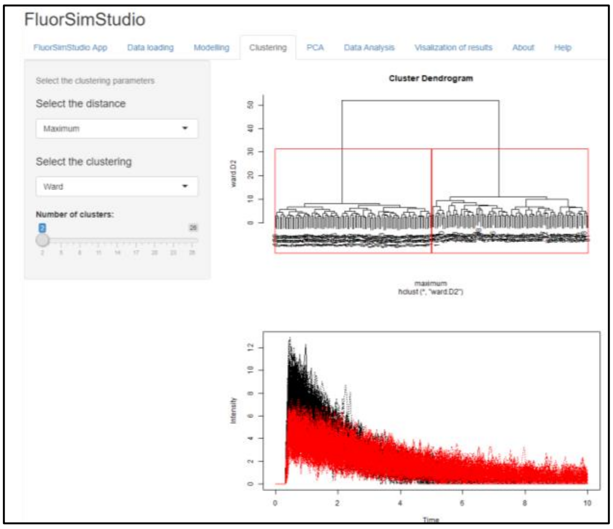
The visualization of the results and the analysis of graphical images of the estimated data clusters are carried out with the aim of interpreting, explaining, improving the understanding of the research object and its behavior (block 6). Reduced data are plotted in the three principal component space, the original feature space, and the principal component coordinates that explain the given fraction of the variation in the data. The presentation of a diagram of three main components, interactive for user interaction, allows to visually assess the proximity of the found clusters and their shapes, the location of individual data points, the influence of experimental effects. Diagrams of a set of informative components enable to determine data clusters for a possible assessment of the parameters of models in the space of the main components. The latter helps to improve the accuracy of parameter estimation by reducing noise in the data due to the elimination of uninformative components describing the experimental noise. The procedure for estimating the parameters of models in the space of principal components can be additionally implemented in the platform. An interactive domain data cluster diagram let to qualitatively explore groups of processed data.

III. RESULTS

For the practical implementation of the digital platform conception, integrating simulation modelling and machine learning algorithms, the computational platform FluorSimStudio is developed for processing fluorescence kinetic curves at FLIM experiments. It is launched on an R server hosted on a network resource, such as shinyapps.io. To implement simulation models, it is proposed to use the C++ programming language. The choice and development of algorithms for data analysis is carried out by direct programming or by connecting ready-made machine learning packages provided by the scientific community of developers through open projects CRAN, Bioconductor, Github. The user's work is carried out through a web application. In the structure of the computational approach, the platform integrates the implementation of simulation models, analysis algorithms, provides computational tools for applying the developed simulation models and methods to the analysis of datasets, instruments for assessing its quality, visualizing and interpreting data.

The programming implementation of the platform FluorSimStudio is organized using the Shiny R package and contains a set of functions that integrate the methodology for an integrated approach to data analysis. The web application is hosted at https://dsa-cm.shinyapps.io/FluorSimStudio. An example of the interface window is shown in Fig. 2. The main interface window consists of nine panels corresponding to six stages of analysis: loading, modelling and clustering data, reducing data dimensionality by the principal component analysis (PCA), fitting medoids (data analysis), visualizing and interpreting the results, information about the authors of the development, and instructions for using the computational resource.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.



Select the clustering parameters

Select the distance: Maximum

Select the clustering: Ward

Number of clusters: 2

Fig. 2. FluorSimStudio web application interface window. Example of clustering fluorescence decay curves

The performance of the computational platform FluorSimStudio was tested by examples of the analysis of datasets representing systems of free fluorophores and in the presence of the Förster electronic excitation energy transfer process [1]. The obtained results are in good agreement with those previously published for analytical models of single- and stretch-exponential fluorescence decay laws [7]. Comprehensive analysis using simulation models and machine learning lets successfully to restore the parameters of optical processes from the experimental data.

IV. CONCLUSIONS

The conception of a digital platform for processing fluorescence spectroscopy data is developed, which is an implementation of an integrated approach for the complex machine learning analysis and modelling of optical processes in biophysical systems. Integrated data analysis pipeline comprises partitioning data into clusters, finding the cluster medoids, applying the data reduction method and visualizing the experimental data in a two-dimensional space, analyzing the medoids with analytical or simulation models. By this data analysis approach, it is possible to enhance the efficiency of the biophysical research. The digital platform is a programming environment designed to model and analyze large experimental fluorescence spectroscopy data. It includes a development framework, coding languages, tools for automation, code debugging and creating an application interface, models and methods for processing and visualizing data, assessing the quality of analysis. The R computing environment and the Shiny package are selected to create a web interface and online version for the developed software application. The C++ programming language is used for accelerating simulation modelling algorithms. The proposed methodology of the digital platform is realized in the computational platform FluorSimStudio, intended for processing fluorescence decay curves in molecular systems. FluorSimStudio provides high productivity of processing large fluorescence datasets, is hosted on the server and can be used in the educational process and for the study of experimental systems. Computational efficiency of the digital platform can be increased by connecting software tools for high performance big data computing (for example, H2O, Apache Hadoop, Spark resources).

REFERENCES

[1] J. R. Lakowicz, Principles of Fluorescence Spectroscopy, 3rd ed. Springer, New York, 2006.

[2] Z. Gryczynski, I. Gryczynski, Practical Fluorescence Spectroscopy. CRC Press, Boca Raton, 2019.

[3] G. Cox, Ed., Fundamentals of Fluorescence Imaging. Jenny Stanford Publishing Pte. Ltd, Singapore, 2019.

[4] D. M. Jameson, Introduction to Fluorescence. CRC Press, Boca Raton, 2014.

[5] R. Datta, T. M. Heaster, J. T. Sharick, A. A. Gillette, M. C. Skala, “Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications”, J. Biomed. Opt., vol. 25(7):071203:1-43, May 2020.

[6] R. Datta, A. Gillette, M. Stefely, M. C. Skala, “Recent innovations in fluorescence lifetime imaging microscopy for biology and medicine”, J. Biomed. Opt., vol. 26(7):070603:1-11, July 2021.

[7] M. M. Yatskou, V. V. Skakun, V. V. Apanasovich, “Method for processing fluorescence decay kinetic curves using data mining algorithms”, J. Appl. Spectr., vol. 87(2), pp. 333–344, May 2020.

[8] M. M. Yatskou, V. V. Skakun, L. Nederveen-Schippers, A. Kortholt, V. V. Apanasovich, “Complex Analysis of Fluorescence Intensity Fluctuations of Molecular Compounds”, J. Appl. Spectr., vol.87(4), pp. 685–692, September 2020.

[9] R. Gentleman et al., "Bioconductor: open software development for computational biology and bioinformatics", Genome Biology, vol. 5(10):R80, September 2004.

[10] V. Yuan, D. Hui, Y. Yin, M. S. Peñaherrera, A. G. Beristain, W. P. Robinson, "Cell-specific characterization of the placental methylome", BMC Genomics, vol. 22(1):6, January 2021.

[11] M. M. Yatskou, Computer Simulation of Energy Relaxation and Transport in Organized Porphyrin Systems. Wageningen University, The Netherlands, 2001.

[12] M. M. Yatskou, Data Mining [in Russian]. BSU, Minsk, 2014.

[13] H. Shimodaira, "Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling", Annal. Statist., vol. 32(6), pp. 2616–2641, December 2004.

[14] T. Jolliffie, Principal Component Analysis. Springer, New York, 2002.

[15] J. A. Nelder, R. Mead, "A simplex method for function minimization", Comput. J., vol. 8(1), pp. 308–313, January 1965.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Eigen Transformations of Symmetric Matrices in Information Processing Problems

Valery Demko
United Institute of Informatics Problems
of NAS of Belarus
Minsk,Belarus
selforg@newman.bas-net.by

Viktoria Zaitseva
United Institute of Informatics Problems
of NAS of Belarus
Minsk,Belarus
zaitsevaria@gmail.com

Abstract. **The mathematical justification of the algorithm for synthesis of proper transformation and the finding the eigenvalue of a symmetric matrix of dimension $N \times N$, $N = 4$ based on orthogonal rotation operators is given. Analytical relations for calculating the eigenvalues of symmetric matrix is optained. It is shown that the proper transformation has factorized structure in the form of a product of rotation operators. Each operator is a direct sum of elementary rotation matrices.**

Keywords: **symmetric matrix, eigenvalues, rotation operators**

I. INTRODUCTION

The analysis and processing of large amounts of data involves compression using fast algorithms, since it is important to have a high speed of information flow to the appropriate data analysis and processing systems. In statistics, the principal component method is used to compress information without significant loss of its informativeness [1]. It consists in a linear orthogonal transformation of the input vector X dimension n in the output vector of Y dimension p, where $p < n$. In this case, the components of the vector Y are uncorrelated and the total variance after the transformation remains unchanged.

The covariance matrix of the input data X is defined as

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix},$$

where r_{ij} - covariance between the i -th and j -th components of the input images.

The eigenvalues λ_k of the matrix R characterize the variance of the principal components. In this case, the sum of the variances in the space of the initial signs is equal to the sum of the variances in the space of the output signs:

$$\sum_{i=1}^{n} r_{ii}(x) = \sum_{i=1}^{n} \lambda_i .$$

II. METHODOLOGY

To solve the problem of eigenvalues of symmetric matrices, different approaches are used, in particular, in [2], a QL algorithm is proposed, which requires a preliminary reduction of the original matrix to a tridiagonal form. However, the QL algorithm uses similarity transformations of a rather complex structure, which do not allow to effectively organize the computational process when implementing its own basis. The method for rotations [2] allows to find all the eigenvalues and eigenvectors of a symmetric matrix without using the characteristic equation. It is known [4] that for a symmetric matrix A an orthogonal matrix Ψ exists when

$$\Psi^T A \Psi = \Lambda ,$$

where $\Psi^T = \Psi^{-1}$ (orthogonality condition), Λ is the diagonal matrix. In the method of rotations, the matrix Ψ is constructed as the limit of the sequence of products of matrices of simple rotations. Rotation operators [3], in contrast to rotation matrices, have invariant properties (Lemmas 1–4), that let solve this problem in an optimal way.

In this paper, we propose an approach for solving the symmetric eigenvalue problem based on the use of orthogonal rotation operators. The rotation operator is a matrix of dimension $n \times n$, $n = 2^i$ ($i = 2, 4, \ldots, N$), which is the direct sum of the elementary Givens and Jacobi rotation matrices. In this case, the symmetric matrix is preliminarily reduced to the persymmetric one (where the symmetry is less than and relative to the side diagonal), and then the synthesis of the proper transformation of the resulting matrix is carried out [3].

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Suppose that a symmetric matrix $A = \left[a_{ij} \right]_{4\times4}$ is given, as well as the rotation operators:

$$\bar{T}_2^{(4)} = \begin{bmatrix} c_1 & s_1 & 0 & \\ s_1 & -c_{11} & & 0 \\ 0 & & -c_2 & s_2 \\ & 0 & s_2 & c_2 \end{bmatrix},$$

$$T_0^{(4)} = \begin{bmatrix} c_3 & 0 & 0 & s_3 \\ 0 & c_4 & s_4 & 0 \\ 0 & s_4 & -c_4 & 0 \\ s_3 & 0 & 0 & -c_3 \end{bmatrix},$$

$$G_1^{(4)} = \begin{bmatrix} c_5 & 0 & s_5 & 0 \\ 0 & c_5 & 0 & s_5 \\ s_5 & 0 & -c_5 & 0 \\ 0 & s_5 & 0 & -c_5 \end{bmatrix},$$

$$T_1^{(4)} = \begin{bmatrix} c_6 & 0 & 0 & s_6 \\ 0 & c_6 & s_6 & 0 \\ 0 & s_6 & -c_6 & 0 \\ s_6 & 0 & 0 & -c_6 \end{bmatrix},$$

$$c_k = \cos\alpha_k, \quad s_k = \sin\alpha_k$$

The following basic properties of the rotation operators are proved:

Lemma 1.

Let $\left[b_{ij} \right]_{4\times4} = \bar{T}_2^{(4)} A \bar{T}_2^{(4)} = B_1$, $b_{11} = b_{22}$, $b_{33} = b_{44}$,

$\left[d_{ij} \right]_{4\times4} = T_0^{(4)} B_1 T_0^{(4)} = B_2$, $d_{11} = d_{44}$, $d_{22} = d_{33}$.

Then the diagnal elments of the matrix B_2 will be equal, i.e. $d_{11} = d_{22} = d_{33} = d_{44} = \frac{1}{2} L_1$, where

$$L_1 = \frac{1}{2} \sum_{i=1}^{4} a_{ii}$$

Lemma 2. Let $\left[l_{ij} \right]_{4\times4} = G_1^{(4)} B_2 G_1^{(4)}$. Then for all α_5 the relation is valid:

$$l_{11} - l_{44} = l_{22} - l_{33}$$

Lemma 3. Suppose that $X = \left[x_{ij} \right]_{4\times4}$ is a symmetric matrix, and $\left[y_{ij} \right]_{4\times4} = G_1^{(4)} X G_1^{(4)}$ and $\left[\bar{y}_{ij} \right]_{4\times4} = T_1^{(4)} X T_1^{(4)}$. Then equality $x_{14} = x_{23}$ follows that:

$$y_{14} = y_{23}, \ \bar{y}_{14} = \bar{y}_{23}.$$

Lemma 4. Let $X = \left[x_{ij} \right]_{4\times4}$, $\left[y_{ij} \right]_{4\times4} = \bar{T}_2^{(4)} X \bar{T}_2^{(4)}$. If the conditions are met $x_{13} = x_{24} = 0$, $x_{14} = x_{23}$, that $y_{14} = y_{23}$.

The invariant properties of the operators allow us to transform the symmetric matrix $A = \left[a_{ij} \right]_{4\times4}$ into a persymmetric form and then use the diagonalization algorithm proposed in [3]. It should be noted, since the rotation operators are a direct sum of elementary rotation matrices that do not commute among themselves, and also, taking into account the invariant properties of the operators, the system of nonlinear equations for determining their parameters is a system with separated variables.

Thus, the eigenvalue of a symmetric matrix has a factorized structure in the form of a product of rotation operators, which implies its use in information processing problems.

REFERENCES

[1] S. A. Ayvazyan et al. Applied Statistics: Classification and Dimensionality Reduction, Ref. Pub..; edited by. S. A. Ayvazyan, Moskva, Finance and statistics, 1989.

[2] B. Parlett, Symmetric eigenvalue problem. Numerical methods, B. Parlett, Moskva, Mir, 1983, 382 p.

[3] V. M. Demko, Orthogonal representation of the proper transformation of a persymmetric matrix based on rotation operators, Computer science, 2018, T. 15, № 1, pp. 34 – 50.

[4] A. G. Kurosh, The course of higher algebra, Moskva, Nauka, 1975, 431 p.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Enhancement of Land Cover Classification by Training Samples Clustering

Artem Andreiev
Dept. of Geoinformation Technologies in Remote Sensing of the Earth Scientific Centre for Aerospace Research of the Earth
Institute of Geological Sciences
National Academy of Sciences of Ukraine
Kyiv, Ukraine
a.a.andreev@casre.kiev.ua
ORCID 0000-0002-6485-449X

Anna Kozlova
Dept. of Geoinformation Technologies in Remote Sensing of the Earth Scientific Centre for Aerospace Research of the Earth
Institute of Geological Sciences
National Academy of Sciences of Ukraine
Kyiv, Ukraine
ak@casre.kiev.ua
ORCID 0000-0001-5336-237X

Abstract. **In this study, a hybrid approach is proposed to enhance land cover classification accuracy by clustering training samples into homogenous subclasses. The proposed approach implies the integration of both supervised and unsupervised classification methods into a holistic framework. A criterion of training sample separability is developed as separability index of training samples. The approach was applied to enhance the land cover classification of the highly heterogeneous natural landscapes by the case of the Shatsky National Natural Park.**

Keywords: **land cover classification, clustering, hybrid approach, training samples separability**

I. INTRODUCTION

Land cover classification is a key research field in remote sensing, which is still challenging in heterogeneous landscapes [1]. The problem mainly arises from the mixing of land cover classes. Nowadays, the solution of the problem is seen in the application of hybrid classification models based on combining both supervised and unsupervised learning [2].

As known, the classification process implies that expert selects the training samples of each land cover class. Hence, it is necessary to obtain a description of each class. However, due to the human factor, selected training samples tend to be inaccurate as well as presented classes are subjective, which, in turn, decreases classification accuracy.

The study aims to enhance land cover classification accuracy by clustering training samples into homogenous subclasses. For this, an approach to land cover classification is developed as a hybrid of supervised and unsupervised methods. The potential of the proposed approach is explored by mapping the highly heterogeneous natural landscapes of the Shatsky National Natural Park.

II. METHODOLOGY

A. Hybrid approach to classification

The hybrid approach to classification is developed to reduce the impact of problems caused by the high heterogeneity of land cover classes [3]. The core of that approach is the integration of both supervised and unsupervised classification methods into a holistic framework. This conception aims to lessen the subjectiveness of expert-selected classes and the mixing of training samples. This point is reached by clustering of classes training samples with the unsupervised methods. Another point is to provide a reasonable interpretation of classes, which is inherent in supervised methods.

Input data of the proposed approach to classification consists of an image and training samples of each class. The training samples should satisfy such requirements as completeness, sufficiency, and purity [4]. The algorithm of the proposed approach is described in Fig. 1.

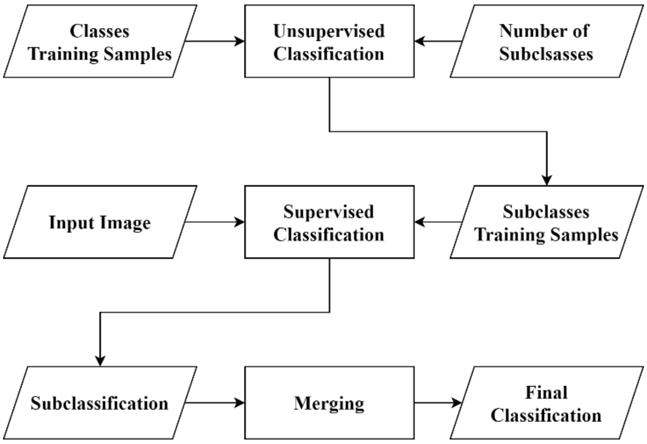


Fig. 1. The scheme of hybrid approach to classification

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The first step of this algorithm implies training samples clustering via unsupervised classification. Firstly, initial classes are subdivided into subclasses. Secondly, classes training samples are clustered into subclasses training samples in the form of subclusters. The method of unsupervised classification could be specified individually for each class since the clustering is performed for each class separately.

The second step engages the supervised classification. The subclusters obtained in the previous step are used as subclasses training samples. Whereas input image is divided into subclasses, but not to classes, the result of this procedure is named subclassification instead of classification. The method of supervised classification should be assigned taking into account the features of an input image and training samples.

The final step is to merge the subclasses of subclassification into initial classes. Since this procedure is pixelwise, each pixel of subclassification requires identifying the initial class of its subclass. This step is necessary to transform subclassification to classification.

B. Number of subclasses

According to the algorithm of the hybrid approach to classification, the parameter “Number of subclasses” is not defined a priori. This parameter sets the number of subclasses for each initial class. The most appropriate value of this parameter is the one that maximizes the separability of training samples, thereby minimizing their mixing.

Fig. 2 describes the algorithm of selection of the most appropriate number of subclasses. Presented algorithm iterates over combination sets, which contains a number of subclasses for each class. Such a combination set is illustrated in Fig. 2 as “Number of subclasses”.

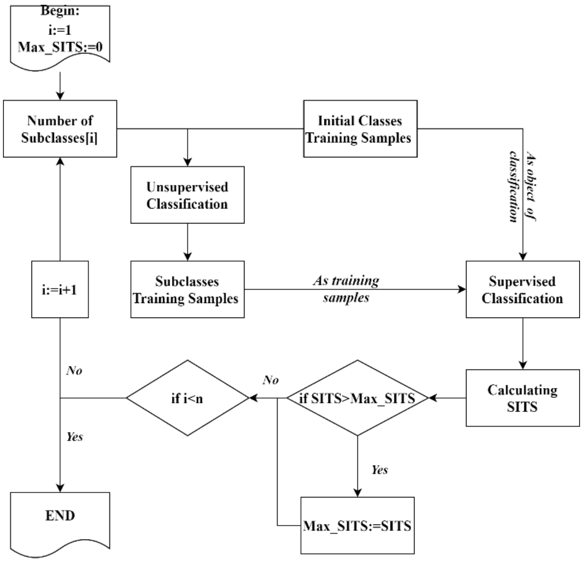


Fig. 2. The algorithm of selection of the most appropriate number of subclasses

In order to limit the iterative process, the maximum number of subclasses should be assigned for each class. An expert makes this decision taking into account the available computational power or some features of training samples, namely their size, density, etc. Thereafter, the number of iterations is calculated by following combinatorial formula:

$$n = \prod_{i=1}^{K} sub_i \quad , \tag{1}$$

where K is the number of initial classes, sub_i is the maximum assigned number of subclasses for i class.

The first step of this algorithm also involves the training samples clustering, as the algorithm of the hybrid approach to classification. However, it is worth noting that the number of subclasses varies with each iteration.

The second step is to perform supervised classification for initial classes training samples and their subclasses training samples as an object of classification and training samples, respectively. The method of supervised classification cannot differ from the one chosen for the hybrid approach to classification.

Estimation of training sample separability is engaged at the third step. In order to carry out this task, the separability index of training samples (SITS) is proposed. SITS quantifies the separability of training samples by measuring the ratio of the number of correctly classified training samples to the total number of training samples. Calculation of SITS is similar to the standard calculation of classification overall accuracy [5]:

$$SITS = \frac{\sum_{i=1}^{K} TS_{\mathrm{corr}\,i}}{\sum_{j=1}^{K} TS_{\mathrm{total}\,j}} \quad , \tag{2}$$

where K is the number of initial classes, TS_{corri} is the number of correctly classified training samples of i class, TS_{totali} is the total number of training samples of i class.

In order to calculate SITS, the number of correctly classified training samples is provided by classification obtained in the second step. Since training samples and their total number are specified a priori as input data, the calculation of SITS could be performed automatically, unlike the calculation of classification overall accuracy.

After all n iterations, the highest value of SITS will be determined. This value refers to the most appropriate combination set among others. Thus, returning to the hybrid approach to classification, this set assigns the number of subclasses for each class.

III. EXAMPLE

A. *Study Area*

The proposed hybrid approach was tested at the area of the Shatsk National Natural Park. It is situated in northwest Ukraine, within Volyn' oblast, between 51º 28'25"N and 23º 49'29"E (Fig. 3). Lying in the vast wetland region of West Polissia, the Park encompasses diverse forests, peat bogs, transitional mires, meadows, and lakes. As a study area, the Park was chosen due to the high heterogeneity of its natural landscapes. During land cover classifications, it often results in the subjectiveness of expert-selected classes and the mixing of training samples.

Since 2007, the Park belongs to the Ukrainian network of the test sites for satellite-based products validation [6]. Above 100 georeferenced sample plots were set here to provide comprehensive ground truth information about the representative landscapes of the West Polissia region.

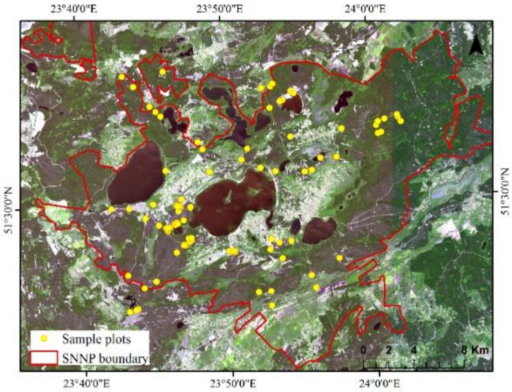


Fig. 3. Location of the study area and sample plots within the Shatsk National Natural Park. They are shown on the fragment of the true-colored composite of the Sentinel-2 Multispectral Instrument (MSI) image acquired on 1 June 2018

B. *Input Image*

A cloud-free Sentinel-2A multispectral instrument (MSI) image acquired on 01 June 2018 was downloaded from the U.S. Geological Survey (USGS) archive (https://earthexplorer.usgs.gov). The image was obtained at the top of the atmosphere reflectance (TOA, Level 1C) and then atmospherically corrected to the bottom of the atmosphere reflectance (BOA, Level 2A) using the Sen2Cor tool (https://step.esa.int/main/snap-supported-plugins/).

During the processing, Sen2Cor discarded the three bands (B1, B9, and B10) that consider the effects of aerosols and water vapour on reflectance. Then, the Sentinel-2 bands acquired at 20 m data were resampled using the nearest neighbour method to obtain a layer stack of 10 spectral bands at 10 m. Finally, the obtained image was resized to the extent of the study area and account for 2284x1554 pixels.

C. *Training samples*

Six broad land cover classes were the focus, as follows: artificial surfaces, forest, natural grassland, agricultural areas, water bodies, and inland wetlands.

As it was mentioned, completeness, sufficiency, and purity are the key requirements for training samples. An extensive analysis of representative features of each class all over the study area provided the satisfaction of the requirements. Updated information from the georeferenced sample plots has also contributed to initial training sample completeness and purity.

The given classes varied considerably both in spatial extent and heterogeneity. The relatively small class included diverse features (e.g. agricultural areas) while the bigger one could be quite homogenous (e.g. natural grassland). Therefore, the number of training pixels of each class also varied disproportionally. The overall number of all training pixels accounted for 2684. Table I shows labels, description, and training pixel amount for the land cover classes assigned for the experiment.

TABLE I. THE CLASSIFICATION SCHEME USED IN THE EXPERIMENT

Land Cover Class	**Description**	**Training pixels**
Artificial surfaces	Urban public and industrial built-up areas, transport units, and construction sites	370
Forest	Broadleaved, coniferous, and mixed forests, roadside tree lines, areas with tree cover more than 30%	611
Natural grassland	Natural herbaceous vegetation, permanent grasslands of natural origin, pastures	544
Agricultural areas	Arable land, permanent crops, fallow lands, heterogeneous agricultural areas, open soils	313
Water bodies	Lakes, rivers and streams of natural origin, including man-made reservoirs and canals.	403
Inland wetlands	Non-forested areas of peat bogs, transitional mires, eutrophic marshes, and reed beds	438

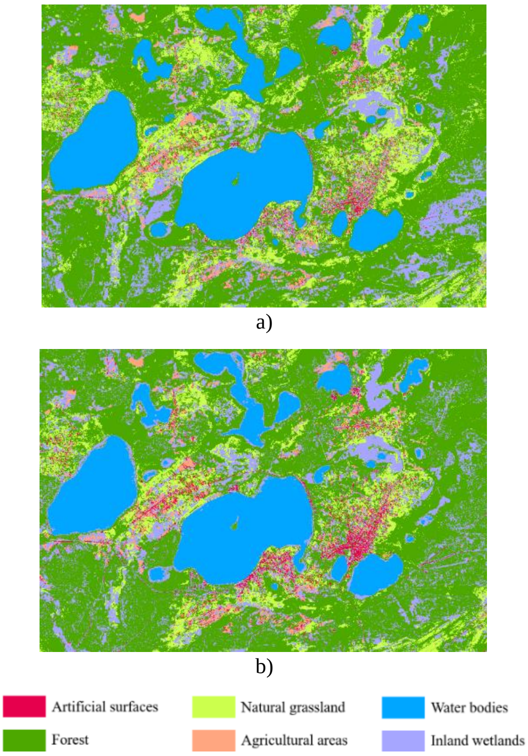
D. *Classifications*

The initial set of training samples were used to obtain a land cover map (Fig. 4a) applying Mahalanobis distance as a method of supervised classification [7]. To estimate the separability of this set, the value of its SITS was calculated by the formula (2):

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

$$\mathrm{SITS}_{\text{initial}} = \frac{TS_{\text{corr}_1} + TS_{\text{corr}_2} + TS_{\text{corr}_3} + TS_{\text{corr}_4} + TS_{\text{corr}_5} + TS_{\text{corr}_6}}{TS_{\text{total}_1} + TS_{\text{total}_2} + TS_{\text{total}_3} + TS_{\text{total}_4} + TS_{\text{total}_5} + TS_{\text{total}_6}} = \\ = \frac{339 + 525 + 538 + 307 + 403 + 356}{370 + 611 + 544 + 313 + 403 + 438} = 0.92$$

Another land cover map was obtained using the hybrid approach, which provides clustering of the initial training samples. The application of this approach involves a definition of subclasses number for each initial class of the training samples. According to the proposed algorithm, an expert should set the maximum number of subclasses for each initial class. Taking into account the size of the initial set of training samples, the maximum number of subclasses for each class was set to 10. Under the formula (1), there are 10^6 different combinations of training samples, derived by their clustering.


a)

b)

Fig. 4. Land cover maps of the study area were obtained using (a) initial training samples and (b) training samples after clustering

Among all iterations of the K-Means clustering [8], the maximum value of SITS is achieved by subdividing initial classes into the combination of subclasses, which is shown in Table II.

TABLE II. THE DETERMINED NUMBER OF SUBCLASSES FOR EACH INITIAL CLASS

Land Cover Class	1	2	3	4	5	6
Number of subclasses	10	3	1	4	4	6

The SITS value of obtained set of training samples is calculated herein (2):

$$\mathrm{SITS}_{\text{final}} = \frac{TS_{\text{corr}_1} + TS_{\text{corr}_2} + TS_{\text{corr}_3} + TS_{\text{corr}_4} + TS_{\text{corr}_5} + TS_{\text{corr}_6}}{TS_{\text{total}_1} + TS_{\text{total}_2} + TS_{\text{total}_3} + TS_{\text{total}_4} + TS_{\text{total}_5} + TS_{\text{total}_6}} = \\ = \frac{363 + 609 + 542 + 308 + 403 + 428}{370 + 611 + 544 + 313 + 403 + 438} = 0.99$$

After that, the final classification (Fig. 4b) is carried out using the determined set of training samples.

E. Accuracy Assessment

The initial and final land cover maps were verified independently from each other using proportionate stratified random samplings. Such sampling technique produces sample set sizes that are directly related to the size of the classes. It is widely used in assessing the classification accuracy of heterogeneous landscapes. To determine the required total sample size, the minimum sample size was set to 0.01% of the total number of the input image pixels. Therefore, validation samples were equal to 355 pixels for each land cover map.

As a primary source of reference data, high spatial resolution satellite images (QuickBird) available in Google Earth TM for 2018 were used for verification.

Confusion matrices were constructed to assess overall accuracy (OA), producer's accuracy (PA), and user's accuracy (UA) of the land cover maps.

Table III shows the confusion matrix of the initial land cover map. Its overall accuracy was 77%. Both producer's and user's accuracy were very low for the classes #1 artificial surfaces (PA – 22%, UA – 50%) and #6 'inland wetlands' (PA – 52%, UA – 29%). User's accuracy was also low for class #3 'natural grassland' (63%), while producer's accuracy was very low for class # 4 'agricultural areas' (40%). Only classes #2 'forest' and # 5 'water bodies' had high both producer's and user's accuracy.

TABLE III. CONFUSION MATRIX OF THE INITIAL LAND COVER MAP

Class#		**Actual Class**							
		1	***2***	***3***	***4***	***5***	***6***	***Σ***	***UA,%***
Predicted Class	***1***	2	0	0	2	0	0	4	50
	2	3	160	3	6	0	7	179	89
	3	2	8	30	2	0	6	48	63
	4	1	0	0	10	0	0	11	91
	5	0	0	0	0	56	2	58	97
	6	1	24	9	5	0	16	55	29
	Σ	9	192	42	25	56	31	**355**	
	PA, %	22	83	71	40	100	52		***OA, % 77***

Table IV shows the confusion matrix of the final land cover map. Its overall accuracy was 81%. Only class #4 'agricultural areas' had low producer's accuracy (31%). However, at the same time its user's

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

accuracy was the highest (100%). For classes #1 ‘artificial surfaces’ and #3 ‘natural grassland’ both producer’s and user’s accuracy were equal (67%) or almost equal (PA – 62%, UA – 66%), but still not high. All other classes had high both producer’s and user’s accuracy.

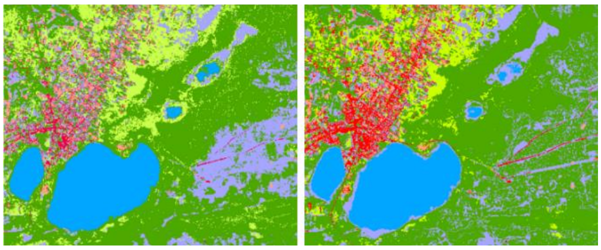
TABLE IV. CONFUSION MATRIX OF THE FINAL LAND COVER MAP

Class#		**Actual Class**							
		1	***2***	***3***	***4***	***5***	***6***	***Σ***	***UA,%***
Predicted Class	***1***	6	0	0	3	0	0	9	67
	2	1	168	6	6	0	2	183	92
	3	0	6	23	6	0	0	35	66
	4	0	0	0	11	0	0	11	100
	5	0	0	0	0	55	1	56	98
	6	2	18	8	9	0	24	61	39
	Σ	9	192	37	35	55	27	**355**	
	PA, %	67	88	62	31	100	89		***OA, %*** **81**

IV. DISCUSSION AND CONCLUSION

The experiment has revealed, that the land cover classification of the study area was enhanced by application of developed approach. This is evidenced by a 4% increase in overall accuracy from 77% to 81%.

The most significant enhancement was appeared in two classes, namely #1 ‘artificial surfaces’ and #6 ‘inland wetlands’ (Fig. 5). PA values of both #1 and #6 classes were increased by 3 and 1.7 times, respectively. This indicates that those predicted classes became more referenced to actual ones. Meanwhile, UA values of both #1 and #6 classes were increased by 1.3 each. This points that those predicted classes became less misclassified.



a) b)

Fig. 5. Fragments of initial (a) and final (b) land cover maps illustrating enhancement of artificial surfaces and inland wetlands classification

Described enhancement is reflected by increase of SITS value after training samples clustering. It is significant, that among with increasement of PA and UA values of mentioned classes, number of their correctly classified training samples (TS_{cor1} and TS_{cor6}) increased from 339 and 356 to 363 and 428, respectively.

Thus, the developed approach enhances land cover classification accuracy of heterogenous landscapes by clustering training samples into homogenous subclasses.

Further research should be aimed at approbation of the developed framework with application of other supervised and unsupervised methods. Also, this approach could be extended by additional criteria of training samples separability.

REFERENCES

[1] J. Paneque-Gálvez, J. Mas, G. Moré, J. Cristóbal, M. Orta-Martínez, A. Luz, M. Guèze, M. Macía, V. Reyes-García, “Enhanced land use/cover classification of heterogeneous tropical landscapes using support vector machines and textural homogeneity.” Intern. J. Appl. Earth Obs. Geoinformation, vol.23, pp. 372-383, Aug. 2013.

[2] J. Xiao, Yuhang Tian, X. Ling Xie, Jiang and Jing Huang. “A Hybrid Classification Framework Based on Clustering.” IEEE Transactions on Industrial Informatics, vol. 16, pp. 2177–2188. Jan. 2020.

[3] A.A. Andreiev, “Hybrid approach to classification of remote sensing data.” CERes Journal, vol. 6, issue 2, pp. 32–37. Dec. 2020.

[4] W.G. Cochran, Sampling Techniques. New York: John Wiley & Sons, 1977.

[5] M.O. Popov, “Methodology of accuracy assessment of classification of objects on space images”, J. Autom. Inf. Sci., vol. 39, pp. 1-10. 2007. (In Russian).

[6] V.I. Lyalko, M.A. Popov, S.A. Stankevich, J.I. Zelyk, S.V. Cherny, “Calibration/Validation Test Sites in Ukraine: current state and directions of further research and development.” Ukrainian Metrological Journal, vol. 2, pp. 15-26. 2014. (In Russian)

[7] L. Bruzzone, B.A. Demir, “A review of modern approaches to classification of remote sensing data”, in Land use and land cover mapping in Europe, I. Manakos, M. Braun, Eds. Springer: Dordrecht, Netherlands, pp. 127–143. 2014.

[8] A.K. Jain, R.C. Dubes, “Algorithms for Clustering Data.” Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Performance Analysis of Deep Learning Models for Heart Segmentation in Chest X-ray Images on a Small Dataset

Ahmedkhan Rdzhabov
Biomedical Image Analysis Department
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
axmegxah@outlook.com

Vassili Kovalev
Biomedical Image Analysis Department
United Institute of Informatics Problems
of NAS of Belarus
Minsk, Belarus
vassili.kovalev@gmail.com

***Abstract.* The widespread practice of screening of the lungs by radiography makes it possible to analyze the chest area for the presence of extrapulmonary pathologies, such as cardiac pathologies. In many cases, it is advisable to assign the process of solving the problem of analyzing and marking up images to automated algorithms. This paper discusses the performance comparison of multiple deep learning models for heart segmentation on chest x-ray images. The information obtained can be used to improve the algorithms for recognizing pathologies in chest X-ray images.**

***Keywords:* Deep Learning, Neural Networks, Computer Vision, Automation, Radiology, Computational Experiment**

I. INTRODUCTION

Due to the wide practice of lung screening by the method of chest radiography, extensive databases of chest X-ray images have been accumulated and there is a possibility, using these databases, to diagnose extrapulmonary pathologies.

Despite the rather long history of image recognition systems [1], they still have not received widespread acceptance. This is largely due not only to insufficient technical development, but also to a lack of systematized knowledge in specific areas [2], which can be overcome by using deep learning methods (neural network methods) along with large amounts of data [3].

The processing of the entire volume of the obtained images and their diagnosis for a wide list of pathologies are complicated for medical institutions by the limited resources. In this regard, it is advisable to use the automation of segmentation and recognition processes, which already at the first stages of technology development makes it possible to redistribute the attention of doctors, focusing their attention on potentially pathological cases and returning attention to cases mistakenly identified as non-pathological.

One of the ways to create an algorithm for the automatic analysis of heart pathologies on X-ray images of the chest is the use of neural networks. To train such models, a large amount of labeled data is required, which is not always available. In this paper, we consider the case of using a small dataset of marked up images (100 pieces) for marking (segmentation) a larger images dataset (more than 2 million images), which has only textual data on pathologies. That can be used for further semi-automatic images segmentation.

Checking the impact of the training model hyperparameters on the result obtained during segmentation can help to reveal some patterns and indicate the disadvantages of the approaches chosen for the algorithm for the automatic analysis of heart pathologies on X-ray images of the chest.

II. DATA

To train the models, an open database of annotated chest X-ray images from General Blockchain Inc, originally associated with COVID studies, containing 100 images of various resolutions, was used (Fig. 1). An example of an image with a superimposed mask, rescaled to an aspect ratio of 1:1, is shown in Figure 1. For testing, an image database obtained from two clinics of the Republic of Belarus was also used. From it, a sample of 70,000 images was made, containing 35,000 images with diagnosed cardiomegaly and 35,000 images of healthy hearts (Fig. 2). All images were converted to 512x512 resolution, because under these conditions, the model did not rest against memory and performance limitations, and the result when working with this resolution turned out better, in comparison with the 256x256 resolution.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

The peculiarity of this dataset is that the images for the final tests and the images for training were obtained on different equipment and their quality varies.

Fig. 1. Sample from the training set

The computational experiment was carried out on a computer equipped with an Nvidia RTX2070 video card (8 GB of video memory). Each run took 1-30 minutes, depending on the number of epochs, the number of neural network layers, and other features of each launch.

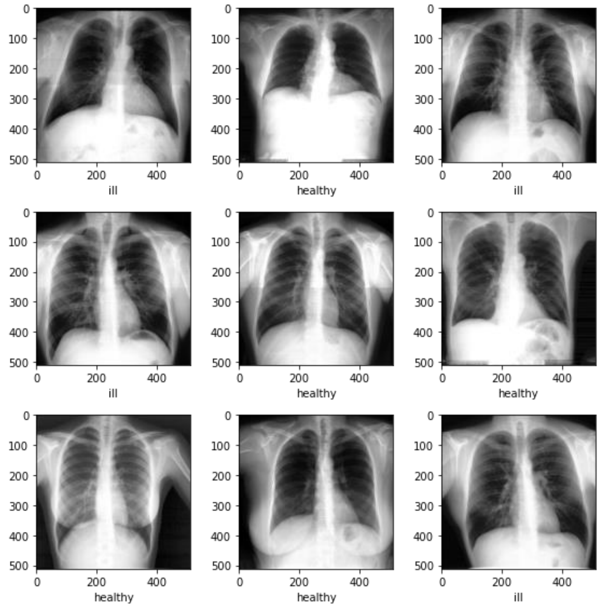


Fig. 2. Samples of used X-ray images

Several preparatory algorithms have been developed using Python, the SQLite language and UNet deep learning model variations to provide cardiac segmentation on chest X-ray images. The operation of this algorithm can be divided into several stages: preparing a dataset by groups of interest in the SQLite database, collecting images by groups of interest in local image databases with simultaneous image preprocessing, training on the collected data in turn by all prepared training models.

III. PERFORMANCE COMPARATION

The launch of training UNet, Simple ASPP, SegResNet, SegResNetVAE, DynUNet, VNet, RegUNet of neural network models showed the greatest promise of using UNet.

The effect of changing size of region of interest (ROI) was also investigated. Below are the masks generated by the trained model for four chest x-ray images (Fig. 3).

a) ROI = 96x96

b) ROI = 384x384

c) ROI = 448x448; with cropping the result

d) Best ROI, U-Net 16-2048 configuration

e) Using residential units and cropping the result

Fig. 3. Examples of the obtained masks for images that were not involved in training

At the stage of analyzing the impact of ROI changes, UNet was used in the configuration: 64, 128, 256, 512, 1024.

The behavior of the segmentation algorithm when using ROI of different sizes aroused interest. So, in the case of ROI = 96x96 (Fig. 3.a), the neural network learned to exclude from the mask only areas with a much lower density, creating a mask for the entire extrapulmonary region, despite learning from heart masks.

The case with an increased ROI up to 384x384 (Fig. 3.b) is characterized by an increase in the sensitivity of the neural network, more parts of the image were removed from the mask.

Adding an image cropping stage at the training stage before evaluating the accuracy at ROI = 448x448 (Fig. 3.c) increased the focus on the heart region, but reduced the integrity of the image data perception by the algorithm, which resulted in a slightly changed nature of the resulting masks.

Further optimization of the neural network parameters led to an increase in the number of UNet layers from 5 to 7, namely to the following configuration: 16, 64, 128, 256, 512, 1024, 2048. This configuration already allows us to achieve visually [4] and numerically better results (Fig. 3.d). Adding residential units (Fig. 3.e) further improves the quality of the resulting masks and reduces the number of large elements. At the last stage, it is possible to use basic algorithms for removing small elements and smoothing shapes to achieve a better result, but since the purpose of this work was only to study the behavior of models, such algorithms were not used.

Typical masks obtained by the latest model, the neural network, which showed the best result, are shown in Fig. 4. Despite the small amount of training data, it was possible to approach a satisfactory result.

a) healthy

b) ill

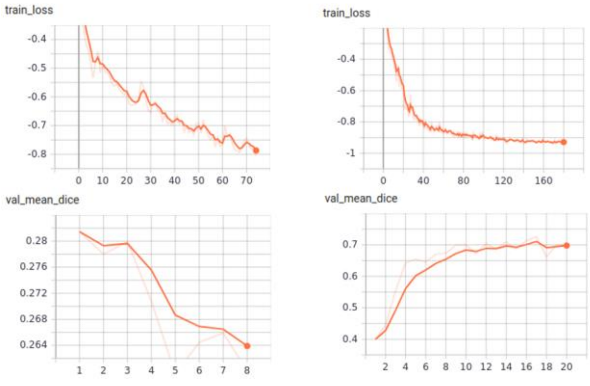
Fig. 4. Best results

The graph for all the described models is shown in Fig. 5. The best final result on the Dice-metric was 0.8484, losses -0.2983. The progress graphs of the learning model can be seen in Fig. 6.

IV. CONCLUSION

The results of these experiments can be used for automatic and semi-automatic segmentation of the heart region on X-ray images of the chest, i.e. to expand the database of the marked-up data, which partially solves the problem of the availability of such data. The trained neural network can be used as a stage in training new models [5]. This data can also be used to build more accurate models of heart segmentation and classification of heart diseases, subject to additional image processing with standard computer vision algorithms and additional training of the model with updated output masks. Also, retraining of pre-trained models can be a promising way to improve performance [6].

Fig. 5. Comparison of the considered models



Low ROI

Balanced ROI

Fig. 6. Evaluation of results in the learning process

In the future, it is planned to work on an algorithm for the detection and classification of extrapulmonary pathologies on X-ray images of the chest, where the obtained information and results will be applied.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

ACKNOWLEDGMENT

In this work at various stages advices, algorithms and experience of my colleague researcher Vitali Liauchuk were used regarding the automation of the process, the preparation of the image base and other issues. I thank him for his help.

REFERENCES

[1] E. H. Shorliffe, Computer-based medical consultations: MYCIN, New York, Elsevier scientific publishing company, 1976, 263 p.

[2] J. S. Pekker, Computer technologies in biomedical research. Signals of biological origin an d medical images: textbook. manual, Tomsk, Ed. TPU, 2002, 240 pp. 7.

[3] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, The MIT Press, 2016, 800 p.

[4] T. F. Tikhomirova, Normal X-ray anatomy and basic syndromes of pathology of the heart and large blood vessels, Minsk,BSMU, 2009, 48 p. (in Russian).

[5] Xupeng Chen, Deep Mask For X-ray Based Heart Disease Classification, Cornell University, Aug., 2018. [Online]. Available: https://arxiv.org/abs/1808.08277.

[6] Hari Sowrirajan, Jingbo Yang, MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models, Medical Imaging with Deep Learning conference, Lubeck, Jul., 2021, pp. 727–743.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

An Automatic Pruning Method for Skeleton Images

Jun Ma
Dept. of Infocommunication Technologies
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
majun1313@hotmail.com

Xunhuan Ren
Dept. of Infocommunication Technologies
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
renxunhuan@bsuir.by

Valery Konopelko
Dept. of Infocommunication Technologies
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
volos@bsuir.by

Victor Tsviatkou
Dept. of Infocommunication Technologies
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
vtsvet@bsuir.by

***Abstract.* Skeletons can be regarded as a compact shape representation in that each pattern can be completely reconstructed from its skeleton. One of the limitations of the application of the skeleton for object analysis and recognition is the existence of redundant skeleton branches. Skeleton pruning is an effective way to remove the redundant skeleton branches in the skeleton images. However, most of the existing pruning methods require manual tuning of the parameter to control the power of pruning, which is not convenient to use. In this paper, we have proposed a fully automatic pruning method that adjusts the power of the pruning according to the image and achieves good pruning results in the experiment.**

***Keywords:* skeleton, automatic skeleton pruning, skeletonization**

I. INTRODUCTION

Skeleton, also termed as medial axis, is an efficient and compact descriptor for pattern recognition. Skeletonization is a set of methods that extract the skeleton from the 2D shapes or patterns while preserve topology characters of the original image. After Blum [1] proposed the primitive skeletonization by simulating the fire-grass, a large amount skeletonization emerged. These methods can be majorly divided into three major groups [2], which are continuous geometric approach, continuous curve propagation approach, and digital approach.

Continuous geometric approach generate the skeleton by focus on the properties of Blum's medial symmetry axis. The most popular method under this category is the Voronoi skeletonization. These methods require many vertices to generate a proper polygonal approximation of a shape. Therefore, Voronoi skeletonization yield many unwanted skeletal branches [3] that contribute little information for overall shape.

The continuous propagation curve approach [4] is modeled using partial differential equations, in which certain singularities occur, which are referred to as shocks. The flaw of this method is that the topology of the resulting skeleton may altered.

The Digital approach of skeletonization is based on simulating Blum's grassfire propagation as an iterative erosion on a digital grid under certain predefined topological and geometric rules [2]. By adopting the parallel strategy, these methods are very fast. In addition, these methods can suppress tiny boundary noise and avoid generating unwanted branches to some extent.

One of the limitations of the use of the skeleton in the application is caused by the fact that there are many spurious skeleton branches that resulted from the small deformations along the boundary [5]. In order to overcome this problem and obtain clean skeletons, a lot of pruning algorithms have proposed by the researchers. Xiang Bai and his team have proposed a pruning method by using the techniques of contour partition with discrete curve evolution [6]. Their method is guaranteed to preserve skeleton topology, does not shift skeleton, and does not shrink the remaining branches. Wei Shen [7] proposed another pruning methods by considering significance measure of bending potential ratio (BPR), in which the decision regarding whether a skeletal branch should be pruned or not is based on the context of the boundary segment that corresponds to the branch. Since their BPR

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

evaluates both local and global shape information, thus it is insensitive to local boundary deformation. HongZhi Liu and his colleagues have proposed a skeleton pruning algorithm based on information fusion [8]. They treat skeleton pruning as a multi-objective decision-making problem. They combined different measurements of branch significance including region reconstruction, contour reconstruction and visual contribution. Guo Siyu has proposed a skeleton pruning method based on saliency sorting [9]. Their method decomposes a skeleton into a number of skeletal components (SCs), and terminal SCs are removed one by one according to a saliency measure. Then SCs may be merged into a new one after each removal. The removal process continues until the desirable number of terminals are achieved.

These pruning methods are all worked well, but all of them have some parameters need to manual tuning. For examples, Bai’s method need to provide the stop parameter k, Shen’s method need provided the value of the filter threshold t, for Liu’s method, there are more parameters rather than one need to specify. Because of these parameters are not very intuitive, a suitable parameter is hard to find. It is necessary to develop a method for automatically chose a proper parameter for pruning without any intervention of the people.

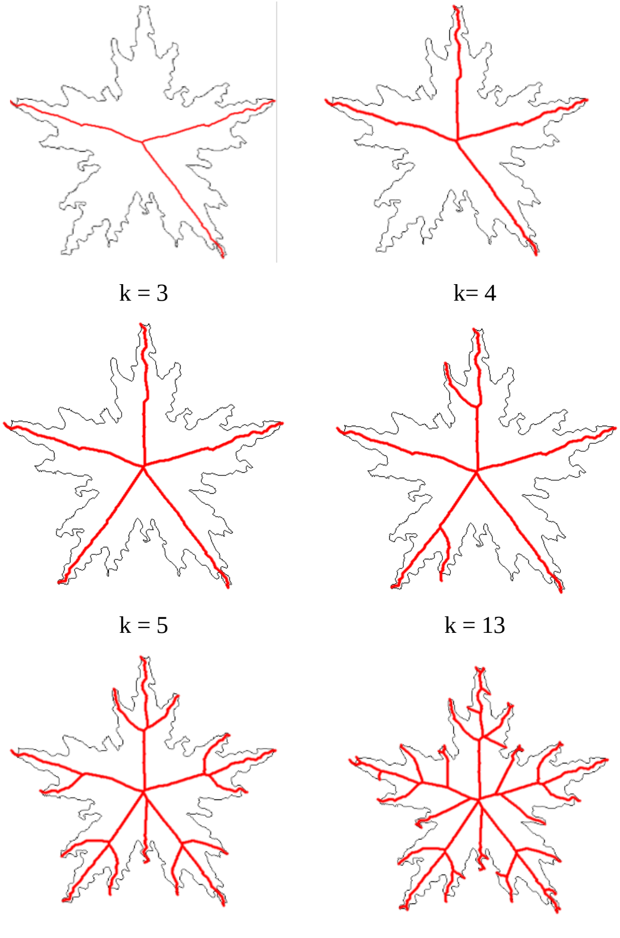
In this paper, we have proposed an automatic pruning method based on the Bai’s pruning skeleton algorithm. The experiment on the images from the benchmark of MPEG-7 has proved that the proposed method can properly pruning the input skeleton for different shape image and produce a relative clean skeleton.

II. PROPOSED METHOD

A. *Motivation*

For Bai's pruning algorithm, which is based on the discrete curve evolution partition method [10–12], a parameter k (which should above or equal three) is required for calibrating the strength of pruning. The value of the parameter k can dramatically alter the level of completeness of the resulting skeleton. As shown in Fig. 1. Therefore, to generate a proper skeleton, It is necessary to manually select a proper k for each shape according to the real human visual perception. Whereas, the intervention of people will dramatically reduce the efficiency of the skeletonization. A complete automatic pruning method is very helpful to extract a relatively good skeleton in a high-efficiency way.

Since Bai's method can generate hierarchical skeletons, it has the potential to conduct pruning automatically. It can be noticed that in Fig. 1., when k is set as the lowest value 3, there only three skeleton branches are reserved, which is not good for representing the original star. With the increase of the k parameter, more and more skeleton branches are retained.



k = 3

k= 4

k = 5

k = 13

k = 30

k = 70

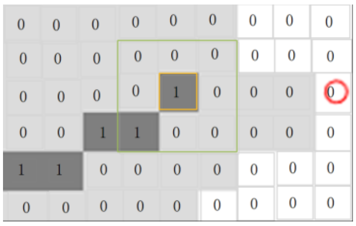
Fig. 1. The result skeleton under different parameter k

When the k is rise to 5, the extracted skeleton already can represent the original structure well. But if we continue to enlarge the value of the parameter k, many unwanted branches are also introduced. This inspires us to propose an automatic pruning method based on Bai's method.

B. *Detail of the Proposed Method*

Before introducing the proposed method, the original input binary image *I* should first respective processed by the skeletonization algorithm to extract the skeleton and discrete curve evolution algorithm to extract the salient points. For the skeletonization algorithm, we chose to use the algorithm proposed in [13], since it has a good performance in terms of computational speed and can suppress slight boundary noise, which can effectively reduce workload in the pruning procedure. In the discrete curve evolution algorithm, we followed Bai’s method. The skeleton and salient points are denoted as *S* and *P* respectively.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.



0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0
0	0	1	1	0	0	0	0	0
1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Fig. 2. The black pixels are skeletal pixels, the grey pixels are the deleted pixels during skeletonization, the yellow pixel is the endpoint, the green window is the 8-neighborhood of the endpoint, the red circle highlight the salient point

The proposed method begins with pairing all the endpoints *E* in the skeleton *S* with their nearest salient point *P*. Endpoints are referred to those foreground pixels that have only one foreground neighbor in his 8-neighborhood (See Fig. 2). Then we sort these endpoints from large to small according to the value of their paired salient points. Hereafter, we iteratively reconstruct the original image from the partial skeleton which consisted by the partial skeleton branches that directly connected to the selected endpoints. In the first iteration, we only select the top three endpoints. Next, two comparisons are conducted which are the area comparing of current reconstruction image *R* with the original image *I*, and area comparing of current reconstruction with the last reconstruction. If the reconstruction image *R* cannot cover most of area of the original image or there is obvious area increment in reconstruction images between two successive two iterations, then including more endpoints in sequence and continue iteration. We summarize the whole process into the following algorithm.

Automatic pruning method

Input **:** Skeleton *S*, Salient Points *P*, Original image *I*

Output**:** Pruned Skeleton *PS*

Step 0: Initialize the all the pixels in Pruned Skeleton *PS* and in the reconstruct image *R1* are backgrounds pixels, set *count* as 0.

Step 1: Search all the endpoints in *S*, paired them with the nearest *P*, and conduct the sort.

Step 2: Reconstruct the image by using top (*count+3*) endpoints along with their skeletal branches and saved it as *R2.*

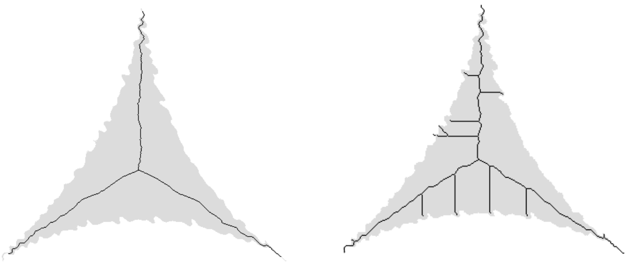
Step 3: If the *area*(*R2-R1*)*/*(*area*(*R1*)*+1*) is above 0.1 **or** *area*(*I-R2*)*/area*(*I*) is less than 0.9 **then** jump to step 4, **otherwise** jump to step 5.

Step 4: *count = count+1, R1 = R2,* jump to step 2.

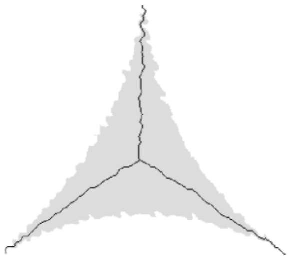
Step 5: Pruned Skeleton *PS* are constructed by the top (*count+3*) endpoints along with their skeletal branches. Return *PS.*

III. EXPERIMENTS AND COMPARISON

In this section, we will compare the result of the proposed method and the result of the Bai's method. We have specified the proper parameter k according to our eye-level view assumption in advance. In the other hand, to enhance the contrast, we also presented the original skeleton that extracted by [13]. All the test images are come from the benchmark of the MPEG-7.

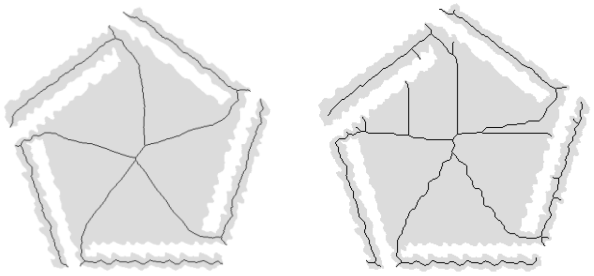


Bai's method (k = 3)   Skeleton without pruning

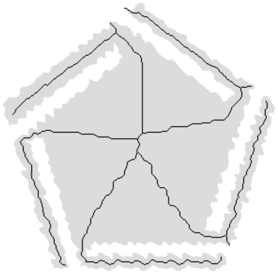


Pruned Skeleton by using proposed automatic pruning

Fig. 3. Skeleton generated from the device4-17 of the MPEG-7



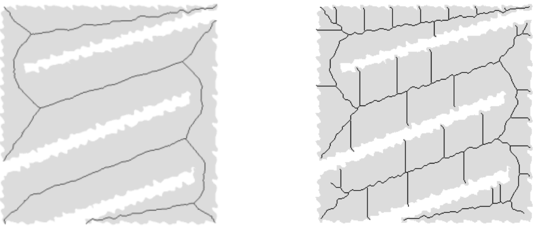
Bai's method (k = 15)   Skeleton without pruning



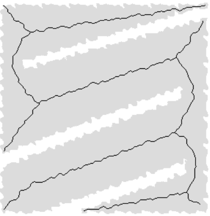
Pruned Skeleton by using proposed automatic pruning

Fig. 4. Skeleton generated from the device6-18 of the MPEG-7

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.



Bai’s method (k = 10) Skeleton without pruning



Pruned Skeleton by using proposed automatic pruning

Fig. 5. Skeleton generated from the device3-18 of the MPEG-7

From Fig. 3 to Fig. 5, it can be noticed that for the different input images, although Bai’s method generates very good skeletons, which maintained the topology component and succeeds in deleting unwanted branches that are caused by the boundary perturbation, the parameter k are different. In Fig. 3, the proper parameter of k is 3, whereas in Fig. 4 and Fig. 5, appropriate value of k is changed into 15 and 10 respectively. In the contrast, the proposed method does not need to change any parameter and can provide a relatively satisfactory result. In each figure, the image of the skeletons without pruning are the input of the proposed method. They are not the input of the Bai’s method. By the way, the skeletonization method used in Bai’s method is the skeleton growing algorithm [14], which is slow but more accurate. Then when we compared the result of the proposed method and its input original skeleton, it is obvious that there are a lot of unwanted branches that have been filtered. Therefore, the proposed method is a good automatic pruning method that without the need to manually tuning parameter for different images.

IV. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an automatic pruning method based on the Bai’s pruning algorithm that by using the discrete curve evolution for partition contour. In the proposed method, manually tuning of the pruning parameter is not necessary, which is inevitable in many pruning algorithms. Therefore, proposed method is more convenient than others pruning algorithm to use. On the other hand, the proposed method generates an enough clean skeleton, in which many unwanted branches are removed. The result of the proposed method is close to the Bai’s method, so it also has the potential to use in shape similarity, which is also one of the merits of the Bai’s method.

In the future, it is interesting to further inspection the performance of the proposed method, such as comprehensive comparing the skeleton extraction with Skeleton Grafting method [15], which is an automatic extract method rather than a pruning method. Another interesting direction is to explore the shape similarity by using the result of the proposed method.

REFERENCES

[1] H. Blum, “A transformation for extracting new descriptors of shape, Cambridge,” MA: MIT press, 1967, pp. 362–380.

[2] P. K. Saha, G. Borgefors, G. Sanniti di Baja, “A survey on skeletonization algorithms and their applications,” Pattern Recognition Letter, 76. 2016, pp. 3–12.

[3] R. L. Ogniewicz, O. Kübler, “Hierarchic Voronoi skeletons,” Pattern Recognition, 28. 1995, pp. 343–359.

[4] F. Leymarie, M. D. Levine, “Simulating the grassfire transform using an active contour model,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 14, 1992, pp. 56–75.

[5] P. K. Saha, G. Borgefors, G. Sanniti di Baja, “Skeletonization: Theory, Methods, and Applications,” Academic Press. London, 2017.

[6] X. Bai, L. J. Latecki, and W. Y. Liu, “Skeleton pruning by contour partitioning with discrete curve evolution,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 3, pp. 449–462, 2007, doi: 10.1109/TPAMI.2007.59.

[7] S. Wei, B. A. I. Xiang, Y. Xingwei, L. L. Jan, “Skeleton pruning as trade-off between skeleton simplicity and reconstruction error,” vol. 56, no. April, pp. 4–9, 2013, doi: 10.1007/s11432-012-4715-3.

[8] H. Liu, Z. H. Wu, X. Zhang, D. F. Hsu, “A skeleton pruning algorithm based on information fusion,” Pattern Recognit. Lett., vol. 34, no. 10, pp. 1138–1145, 2013, doi: 10.1016/j.patrec.2013.03.013..

[9] G. Siyu, H. Pingping, L. Zhigang, W. He, L. Min, “A skeleton pruning method based on saliency sorting,” 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI). pp. 593–599, 2019.

[10] L. J. Latecki and R. Lakämper, “Convexity Rule for Shape Decomposition Based on Discrete Contour Evolution,” Comput. Vis. Image Underst., vol. 73, no. 3, pp. 441–454, 1999, doi: 10.1006/cviu.1998.0738.

[11] L. J. Latecki and R. Lakämper, “Polygon evolution by vertex deletion,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 1682, pp. 398–409, 1999, doi: 10.1007/3-540-48236-9_35.

[12] L. J. Latecki and R. Lakamper, “Shape Similarity Measure Based on Correspondence of Visual Parts,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 10, pp. 1185–1190, 2000, doi: 10.1109/34.879802.

[13] J. Ma, X. H. Ren, T. V Yurevich, “A Novel Fast Iterative Parallel Thinning Algorithm,” Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing. – 2020. – C. 1–5.

[14] W. Choi, K. Lam, W. Siu, “Extraction of the Euclidean skeleton based on a connectivity criterion,” Pattern Recognition, vol. 36, pp. 721–729, 2003.

[15] C. Yang, B. Indurkhya, J. See, M. Grzegorzek, "Towards Automatic Skeleton Extraction with Skeleton Grafting," in IEEE Transactions on Visualization and Computer Graphics, 2020. doi: 10.1109/TVCG.2020.3003994.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Software Control System for Equipment of Integrated Circuit Layout Inspection

Alexander Doudkin
United Institute of Informatics
Problems of NAS of Belarus
Minsk, Belarus
doudkin@lsi.bas-net.by

Alexander Voronov
United Institute of Informatics
Problems of NAS of Belarus
Minsk, Belarus
voronov@lsi.bas-net.by

Valentin Ganchenko
United Institute of Informatics
Problems of NAS of Belarus
Minsk, Belarus
ganchenko@lsi.bas-net.by

Abstract. **This paper describes the main functions and architecture of software control system for equipment of critical sizes inspection for layouts of integrated circuit on basis of computer vision. Advantages of the developed architecture are described, as well as its application for image processing of integrated circuit layouts. The system allows identifying effectively defects what it is especially important for Very large-scale integration manufacturing based on submicron technology.**

Keywords: **image processing, VLSI, automatic layout inspection, control and measurement equipment**

I. INTRODUCTION

The equipment of integrated circuits (ICs)layout in section is characterized by a large diversity and essentially differs in degree of complexity: from simple visual inspection tools for mass production to the most complex automatic inspection and measurement systems which are used both in R&D of new technologies and devices, as well as in the large-scale production. Automatic inspection and measurement system for defects in semiconductor wafers is already used in microelectronics [5]. Scientific and technical group of dedicated designers, technicians and production engineers which develops and supplies special systems for realization of critical technologies in microelectronics and precise engineering are foundation for development and further growth of modern technology.

Modern means of developing ICsare aimed at reducing time for mastering and launching new products into production, as well as reducing cost of digital equipment during its mass production. Such an opportunity is provided by technological base, including machine (computer) vision systems, which are an integral part of modern technology for the design and production of ICs.

In connection with updating new submicron design standards and increasing complexity of the ICs themselves, it becomes necessary to solve problems of processing, storing, receiving and transmitting large amounts of data obtained during lithographic process of ICs design. Original approaches for image processing allow to fully complying with conditions of submicron manufacture of Very large-scale integration (VLSI) and to reduce cost of production. The object of the study is process of critical dimensions inspection on the photomasks and VLSI layouts. The processing consists of image analysis, generating reports based on the previous analysis results, controlling the focusing system, coordinate table and other external devices, as well as synthesis of routines for the automatic operation of control equipment for monitoring of layout critical sizes.

The developed Software Control System (SCS) for equipment of ICs layout inspection is based on machine vision and provides the following functions: image preprocessing taking into account design and technological constraints; image processing and analysis with support for third-party video camera equipment; image analysis to control design and technological constraints; storage and access to data with the ability to import and export data in various formats; program synthesis for automatic operation; management of third-party mechanisms; data visualization. The main analogs of the SCSare the Olympus MicroSuite FIVE software systems from Olympus Corporation (Japan) and NIS-Elements Microscope Imaging Software from Nikon Instruments Inc. (USA). Analogs of installations for monitoring critical dimensions and their approximate cost: LEICA LWM – 4.8 million dollars, KLA Tencor IPRO – 5 million dollars.

II. SCS USE CASES

Use cases can be divided into main and minor. The main cases consist of: loading an object (as emiconduct or wafer or its photomask) – preparing the object for further work (preliminary orientation in space, moving to the working area); unloading of the object – removal of the object from the working area to the storage (container, cassette); initialization of the installation and basing of mechanisms – loading data

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

describing the initial state of the equipment for solving a specific problem, and setting the mechanisms to the appropriate state/position; movement control of the coordinate table – formation of generalized commands to control the movement of the coordinate table; alignment and orientation – binding of the reference system and the coordinate system of the object to the coordinate system of the installation; control of dimensions – launching algorithms for control and measurement of dimensions; automatic measurement – launching algorithms for automatic measurement of dimensions; sizing of elements – launching algorithms for determining the size of the image; mechanism control – preparation of equipment control commands (meta-commands) and corresponding parameters; generating of control commands – transformation of meta-commands into the format required by microcontroller for equipment control; creation of a control program for automatic mode – generate a list of control actions with appropriate parameters and their saving as a file or record in a database (DB); creation of a map-structure of an object – preparation of a description of the structure of an object for its further study; save the results of control – the results of PC operation into a structure intended for further storage in the DB and the subsequent saving of the received data block using the DB.Minor use cases for user identification are required to delimit access to SCSfunctionality.

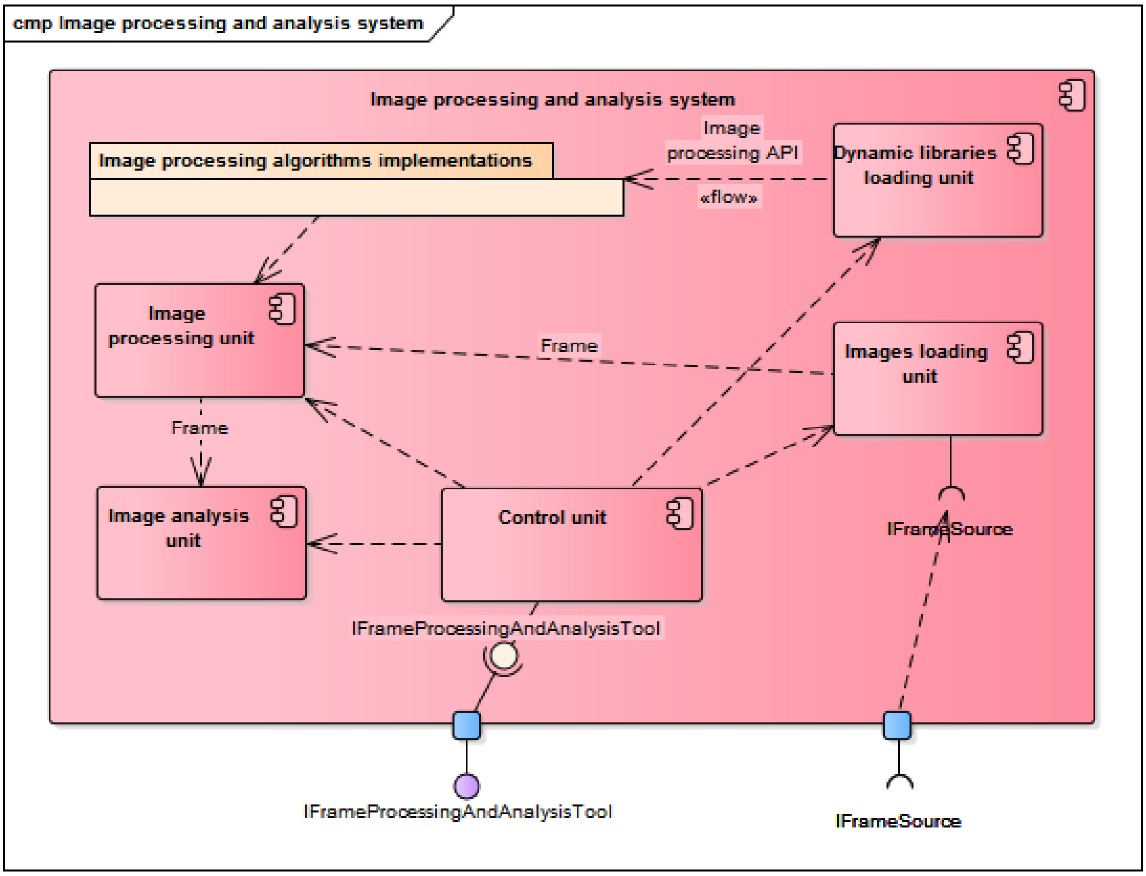


Fig. 1. Architecture of the image processing and analysis system

III. SCS ARCHITECTURE

SCSconsists of several subsystems; one of them is an image processing and analysis system. The architecture of the image processing and analysis system can be described in form of a diagram of its components (Fig. 1). The diagram (Fig. 1) describes composition and interconnections of image processing modules control module, which is designed to select algorithm implementations specified by parameters and coordinate for the interaction of the modules; the module for operation uses data received from source ICommandSource; the image loading module, which is used to retrieve individual images from an IFrameSource image source; the dynamic library loading module that loads dynamically loaded libraries containing implementations of various algorithms for image processing and analysis, and also provides a general interface for accessing the loaded functionality; the modules that implement image processing algorithms –they are dynamically loaded from libraries, as well as an interface processing algorithm that allows

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

you to call various algorithms sequentially; the image processing module that calls the selected implementation of the image processing algorithm (or their chain); the image analysis module that is designed to analyze processed images and transfer the analysis results to IProcessingResultStorage storage module.

The ability to use the dynamically loaded libraries provide a compatibility of the system with image processing algorithms that are implemented in form of such libraries. Thus, the system under development can use the following image processing algorithms [1–11]: ContrastCorrection; GammaCorrection; Invertion; LaplaceFiltration; MeanFiltration; MedianFiltration; Morphological Closing; MorphologicalDilatation; Morphological Erosion; Morphological Opening; ThresholdBinarization. The functioning of the system is carried out according to the generalized algorithm, the steps of which are given below.

Step 1. Loading operation parameters from the command source ICommandSource.

Step 2. Loading dynamic libraries.

Step 3. Preparation of the algorithms' implementations required for processing.

Step 4 If the images in the IFrameSource are still available, then go to step 5, otherwise go to step 10.

Step 5. Loading image from source.

Step 6. Image processing with the selected algorithm or algorithms.

Step 7. Analysis of the processed image.

Step 8. Saving the analysis result to the IProcessingResult Storage result storage.

Step 9. Go to step 4.

Step 10. Completion of work. As noted earlier, the software module for image processing and analysis, almost all elements of the SCS should be a set of dynamically loaded libraries containing supported functions. The following classes were implemented:

- interfaces: IFrameProcessor– interface for implementing image processing algorithms; IFrameAnalyser - interface for image analysis algorithms; IFrameProcessingAndAnalysisTool– interface of the control module;

- data classes: FrameProcessingState - the result of image processing; FrameAnalysisResult – result of image analysis; ProcessingAndAnalysisResultsContainer – combined result of image processing and analysis;

- class factories: FrameProcessorFabric – the factory for image processing algorithms; FrameAnalyserFabric – the factory for image analysis algorithms; FrameProcessingAndAnalysisToolFabric – control module implementations factory;

- processing/analysis chains: FrameProcessorChainProxy – a chain of image processing algorithms; FrameAnalyserChainProxy – image processing analysis chain; FrameProcessingAndAnalysisTool – implementation of the control module; DynamicLibraryLoader – module for loading dynamically loaded libraries; DynamicLibraryLoader::Handle – handle to the loaded dynamic library (the class is a subclass of the DynamicLibraryLoader class);

- classes that implement image processing algorithms: ContrastCorrection-FrameProcessor– contrast correction;GammaCorrectionFrameProcessor– gamma correction; InvertionFrameProcessor– color inversion; LaplaceFiltration-FrameProcessor– Laplace filter; MeanFiltrationFrame Processor – averaging filter; MedianFilterationFrame Processor – median filter; MorphologicalClosingFrameProcessor– morphological closure; Morphological-DilatationFrameProcessor – morphological dilatation; MorphologicalErosionFrameProcessor– morphological erosion; MorphologicalOpeningFrameProcessor – morphological opening; ThresholdBinarizationFrame-Processor – threshold binarization.

The SCS includes implementation of main function by special systems and subsystems: an image processing and analysis system, including a video camera subsystem for preparing data for use; a control system for functional linking of other systems and subsystems; the mechanism control system for generating unit control commands; graphic user interface for the user to control the functioning; subsystem of interaction with the DB for storing the results of inspection; a subsystem for control program generation (description preparation of the configurations of operation used in the automatic mode of operation). In addition, each of the systems must be implemented with a sufficiently high level of abstraction to ensure uniform operation when using different video equipment and control equipment. So when choosing an design pattern for software package, the following criteria were used: modularity; openness; configurability; separation of graphical user interface and functionality. The most convenient design pattern based on the listed before criteria is MVC (Model-View-Controller). MVC pattern with some modifications allows to take into account mentioned above criteria and requirements to the architecture of

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

the SCS: the control system must be able to receive commands from several sources – an interface for receiving control commands is added for converting general view commands (meta-commands) into specific commands for such equipment; the control system must be able to receive a video stream from several sources – an interface for receiving video data is added, which converts various data formats to a single format an interface for receiving control commands is added for converting general view commands (meta-commands) into specific commands for such equipment, the virtual data model that stores a description of the state of the parameter control process is implemented.

The developed architecture of the SCS is shown on Fig. 2. The work of the SCS is carried out as follows:

1. Initialization of user work by issuing control actions to the control system (by the user interface or loading the configuration for automatic operation).

2. Transformation of control commands, if necessary.

3. Receiving data from video camera using the appropriate SDK, convert data for processing.

4. Processing and analysis of data by appropriate subsystems.

5.Transmission of video stream and analysis results in the control system, if necessary, format conversion.

6. Development of control actions by the decision-making subsystem (decision based on the received commands, video stream and analysis results).

7. Transferring of data at the control process to the virtual data model.

8. Signal transmission when changes the state of the model from the virtual data model to the graphical user interface.

9. Request the required data graphical user interface from the virtual data model and retrieving them.

10. Transmission commands of control to equipment by the decision-making subsystem through the appropriate next path (interface, the mechanism control system and the corresponding Software Development Kit (SDK)).

11. Save the results of control in the DB.

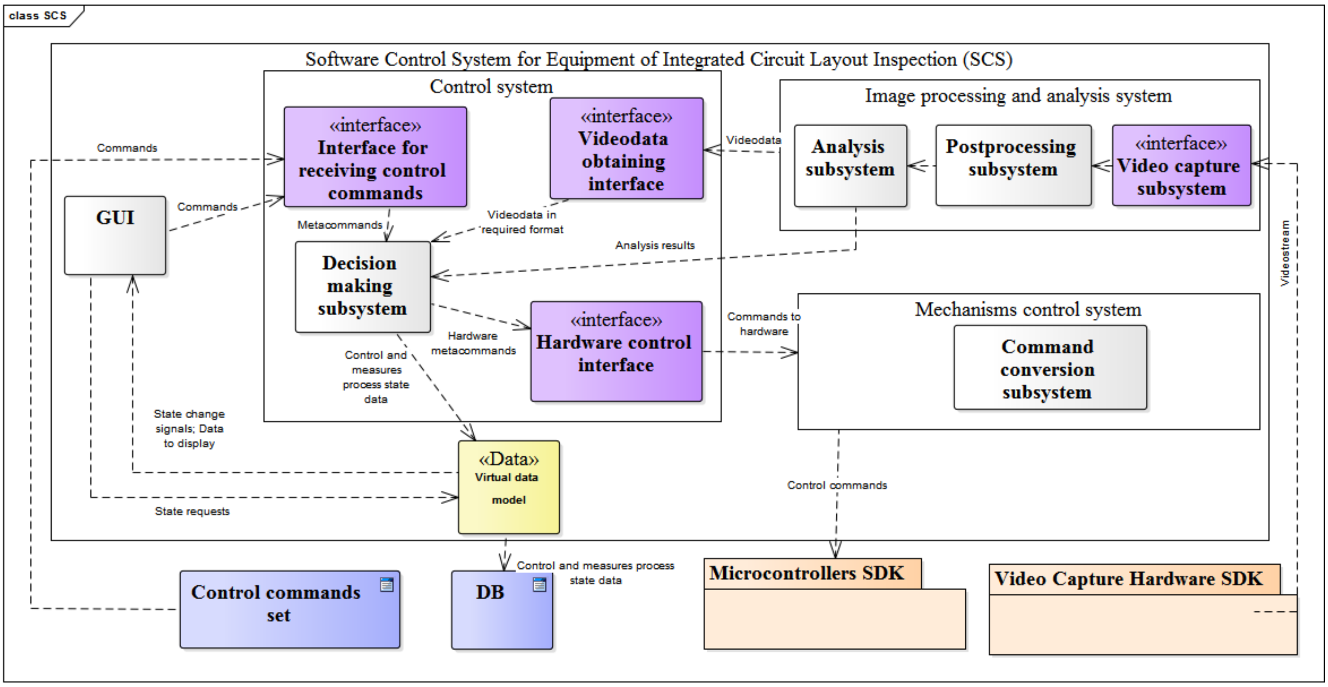


Fig. 2. Architecture of the SCS

IV. CONCLUSION

The general architecture of program complex for control of equipment for monitoring of critical size based on machine vision systems has been developed, which allows working with big input data and easily adapted to specific equipment.

In this paper, the requirements and the structure of the SCSare described. technology allows identifying effectively defects that is especially important for software engineering for equipment of critical sizes inspection of VLSI manufacturing based on submicron technology.

The developed software provides the following functions:

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

- image preprocessing taking into account design and technological restrictions;
- image processing and analysis with support of different video equipment;
- image analysis for inspections of manufacturing operations;
- storage and access to data with the ability to import and export data in various formats;
- synthesis of the program for automatic operation mode;
- control of different mechanisms;
- data visualization.

The architecture of the software system was developed, providing possibility of flexible adjustment of general algorithm of image processing and analysis. The user can independently compose chains of simple algorithms to obtain more complex ones. It is also possible to connect external routines. The operation parameters and automatic processing programs are stored in the database.

The software is used in production of competitive precision equipment for VLSI manufacturing what determines its practical importance: for automatic photometry with precision laser focusing system; for automated microsize inspection system; for mask pattern coordinates measurement system, equipment for mask pattern generation and inspection.

A significant advantage of equipment of the controlled of SCS and developed by JSC Planar for the production of VLSI over foreign counterparts is that it is designed on a single design and technological base, realizing full hardware, software and metrological compatibility of the entire set of installations operating in a single technological cycle for Embodiments in silicon of critical technologies of the microelectronic industry.

REFERENCES

[1] R. Gonzalez, R. Woods, The world of digital processing, Digital image processing, Moscow, Technosphere, 2005, p. 660.

[2] V. V. Starovoitov, Local geometric methods of digital image processing and analysis, Minsk, In-t techn. Cybernetics, p. 282, 1997 (In Russian)

[3] V. A. Soifer, Theoretical foundations of digital image processing : Textbook, Samara, SSAU, 2000, p. 255.

[4] A. A. Dudkin, A. A. Voronov, E. Marushko, Algorithm for stitching frames of the VLSI topology layer by key points, Bulletin of the Brest State Technical University, 2015, No. 5 (94): Physics, mathematics, computer science, pp. 11-14 (In Russian)

[5] A. P. Dostanko, S. M. Avakov, O. A. Ageev, M. P. Batura Technological complexes of integrated processes for the production of electronic products, Minsk, Belorusskaya Navuka, 2016, pp. 251.

[6] M. Fowler, Architecture of corporate software applications, Moscow, Publishing house "Williams", 2006.

[7] L. Craig, Application of UML 2.0 and design patterns, Ed. 3-E, Moscow, Izd. house "Williams", 2013.

[8] G. Erich, Techniques of object-oriented design, SPb. Ed. house "Peter", 2015, pp. 368.

[9] E. Evans, Domain-oriented design (DDD): structuring of complex software systems, Moscow, "Williams", 2011.

[10] D. Spinellis, G. Gusios, Ideal architecture. Leading experts on the beauty of software architectures, St. Petersburg, Symbol-Plus, 2010.

[11] N. Ford, R. Parsons, P. Kua, Building evolutionary architecture (Support constant change), St. Petersburg: Symbol-Plus, 2019.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Cloud Detection in Aerospace Imagery of Environmental Monitoring

Alexander Labokha
ITS Department
Belarusian State University
of Informatics and
Radioelectronics
Minsk, Belarus
labokha.poit@bsuir.by

Artsiom Shamyna
ITS Department
Belarusian State University
of Informatics and
Radioelectronics
Minsk, Belarus
shamyna@bsuir.by

Alexei Ardyako
ITS Department
Belarusian State University
of Informatics and
Radioelectronics
Minsk, Belarus
ardyakon@gmail.com

***Abstract.* In this paper, the problem of cloud detection in the processing of aerospace images is considered for further use of images in monitoring systems for natural complexes. The existing methods of cloudiness recognition on aerospace images are analyzed, an approach based on an admissible range of values is proposed for solving a particular class of problems. An experimental comparison of the algorithms is carried out, and on the basis of the results obtained, the effectiveness of the proposed method is determined. An approach is identified for the most complete removal of cloudiness from the image.**

***Keywords:* cloud detection, aerospace images, monitoring, GIS**

I. INTRODUCTION

Aerospace images can be used in monitoring systems for natural and other complexes. In this case, it is important to improve the regularity of earth remote sensing data updates. The presence of partly or overcast clouds in images can significantly increase the refresh period, so it is important to use informative pixels even from images that contain clouds. One of the possible solutions to this problem is to identify and replace pixels containing clouds with pixels from the previous image with informative data.

The research was carried out as a part of the development of a methodology for determining the level of fire hazard using aerospace images in fire hazardous periods (from March to October) 2019-2020 for the territory of the Volozhin forestry enterprise. During the selected interval (245 days) in 2019, Sentinel-2 satellites took 61 images with a cloud percentage of no more than 20% and 81 images with no more than 40% cloud cover; in 2020, 36 and 64 images were taken, corresponding to a cloud level of no more than 20% and 40%, respectively. So, the average frequency of taking images with an acceptable cloud level is at least 3 days, which confirms the relevance of the task.

II. CLOUD DETECTION ALGORITHMS

The Braaten-Cohen-Yang method [1] uses the threshold condition to determine the presence of cloudiness in the considered pixel based on the values in spectral channels B3 and B4 of Sentinel-2 satellite images [2]. The efficiency of this fairly simple algorithm on the Hollstein dataset reaches 73% of the classification accuracy. Most of the errors arise from the definition of snow as a cloud and the inability to detect thin clouds.

A Scene Classification Layer (SCL) layer is available for Sentinel-2 satellite images. The Scene Classification Algorithm [3] generates classification maps that include four different classes of clouds (including cirrus) and six different classifications for shadows, cloud shadows, vegetation, soil or deserts, water and snow. The algorithm detects pixels with clouds or cloud shadows and is used for replacing pixels with clouds based on the SCL classification.

The approaches described above are universal, they affect in their changes only pixels that contain clouds, and do not affect other pixels that can potentially contain information useful for research. However, in the case of solving a certain problem, it is possible to sacrifice pixels, that don’t contain cloudiness, and will not be used in the future. Then, for the pixels of interest a range of values can be determined to be enough large in order to guarantee the extremely low probability of the value of the pixels of interest to fit in range. The pixels with values outside the specified range are determined as potentially containing clouds and they can be replaced with pixels from the previous image.

The interest of this study is an information about forest vegetation, the analysis of the spectral channel

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

B2 of various images with cloud percentages from close to 0% to 50% and higher was carried out. Information on forest vegetation is predominantly in the range from 200 to 400, while pixels containing cloudiness generally have values greater than 800 and 1000. Pixel values that are greater than 800 and 1000 also corresponded to agricultural areas and settlements. This confirms the previously presented hypothesis and suggests that replacing pixels with a value greater than 1000 (or 800) in spectral channel B2 will not affect pixels with useful information about forest areas, but will replace most of the pixels "polluted" by clouds.

It should be noted that such an approach will not allow identifying all types of cloudiness, and does not allow taking into account the presence of cloud shadows in the images. Since each method has both strengths and weaknesses, it is likely that sequential application of several methods for replacing cloudy pixels will reveal the largest number of cloudy pixels.

III. EXPERIMENTAL COMPARISON OF ALGORITHMS

When conducting an experimental study, the effectiveness of the described methods was analyzed both separately and as a result of joint application for the same image. For the best assessment of the methods, an image of the Volozhin forestry enterprise from 06/17/2020 was selected, containing various types of clouds, as well as cloud shadows; a fragment of its spectral channel B2 is shown in Fig. 1, a. As the previous donor image contains a low percentage of cloud cover, the image from 06/17/2020 was used; a fragment of the spectral channel B2 in Fig. 1, b. The result of applying the Braaten-Cohen-Young method is shown in Fig. 1, c; method based on the classification of scenes Sentinel-2 – in Fig. 1d; method for analyzing channel B2 values with a range of values from 0 to 1000 – in Fig. 1, e. The result of sequential application of these three methods is shown in Fig. 1, f. For a representative visual comparison, all images are displayed in a gradient from black to white in the range from 0 to 2000.

IV. RESULTS AND CONCLUSIONS

Visual analysis of the obtained images allows us to make the following conclusions:

1) The Braaten-Cohen-Young method is better than other methods to recognize transparent and translucent clouds, as well as cloud shadows. At the same time, this method is much worse than the others in processing continuous clouds, without replacing individual pixels within the cloud.

2) The Sentinel-2 scene classification method is well suited for replacing pixels within overcast clouds. This method poorly handles transparent and semi-transparent edges of clouds, replacing only part of the pixels in the center of such clouds. The algorithm does not detect pixels covered with haze or shadowed by clouds.

3) The method based on the analysis of the values of the spectral channel B2 with an allowable range of values from 0 to 1000 according to the results of the work is close to the method based on the classification of scenes, but it is better for processing semitransparent clouds and the boundaries of continuous clouds. The handling of transparent clouds and cloud shadows is also low.

4) The combined use of the three considered methods allows you to use the advantages of each of the methods and mutually reduce their disadvantages. The resulting image contains a minimum of cloudiness compared to the results of using the methods separately.

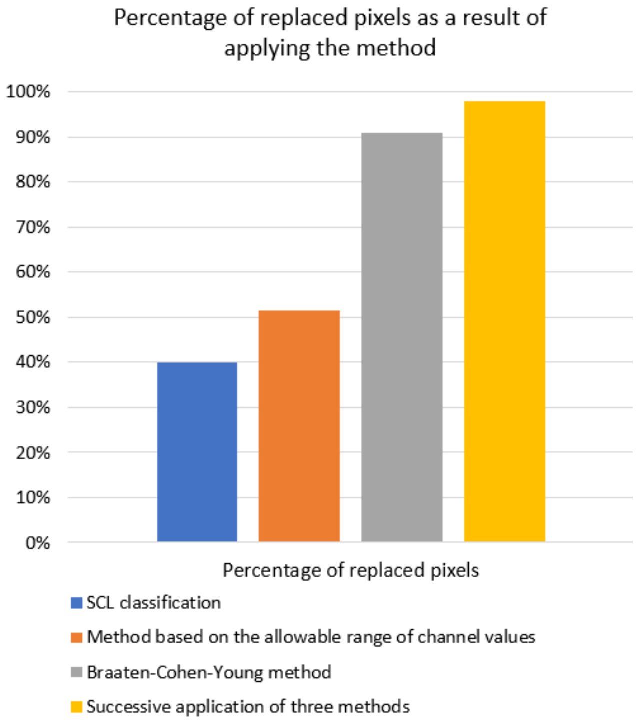
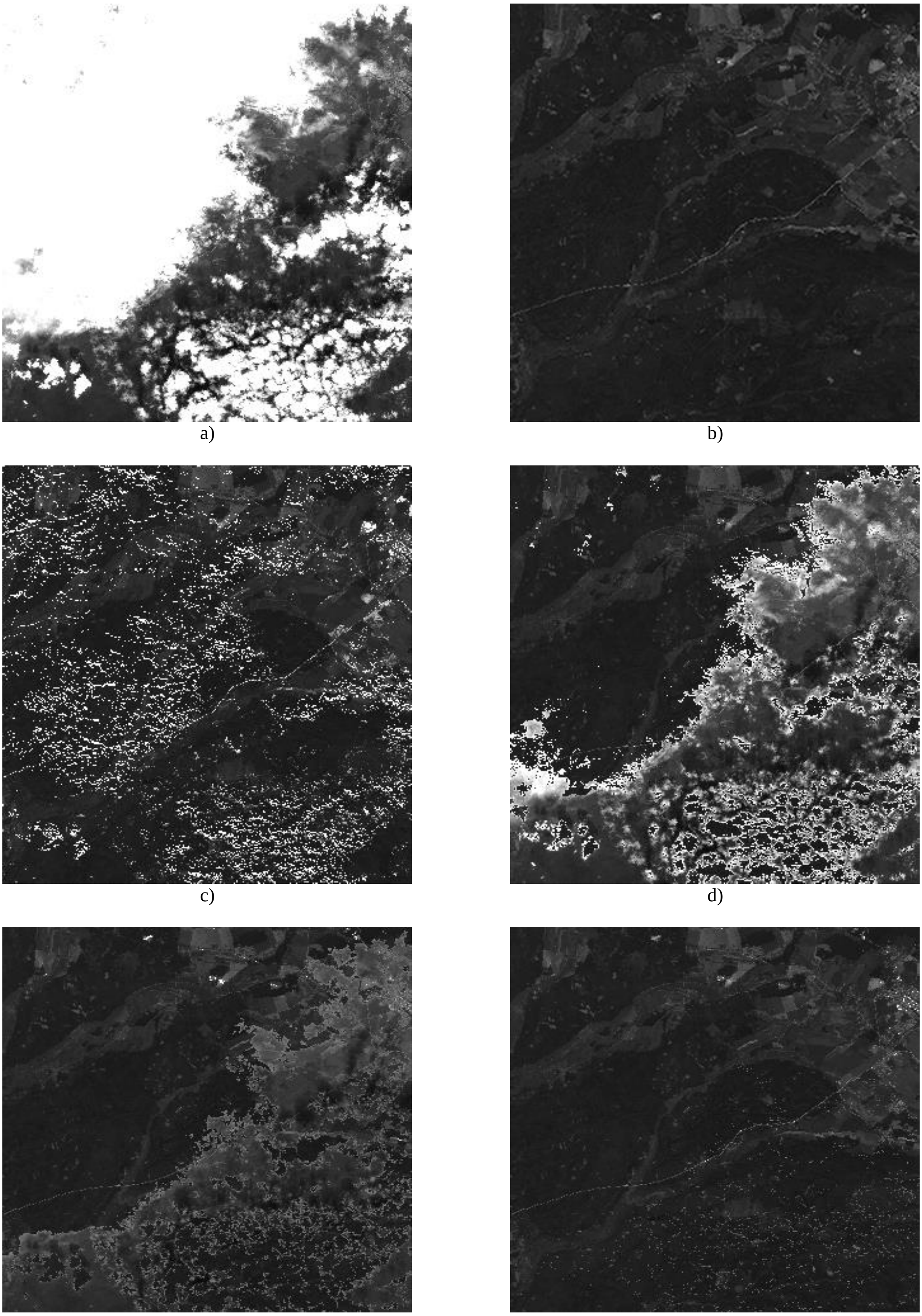


Fig. 2. Histogram of the percentage of replaced pixels

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.



a)

b)

c)

d)

e)

f)

Fig. 1. Spectral channel B2: a) image for June 17, 2020, without processing; b) image for 06/10/2020, without processing; c) image for 06/17/2020, cloud replacement using the Braaten-Cohen-Young method; d) image for June 17, 2020, cloud cover change according to SCL classification; e) image for 06/17/2020, cloud cover replacement according to the allowable range of values from 0 to 1000; f) image for 06/17/2020, replacement of clouds by successive application of three methods

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

A numerical assessment of the effectiveness of each of the methods is the percentage of detected pixels containing clouds of the total number of pixels belonging to forest areas, it is shown in Figure 2. For a selected day with a high percentage of cloudiness, the method based on the classification of Sentinel-2 scenes marked about 40% of pixels as containing clouds. The proposed approach based on the range of values noted by a quarter more pixels (more than 50%), while the spatial distribution of the detected cloudiness correlates with the results of the SCL method, expanding the selected areas (Fig. 3, a, b). The Braaten-Cohen-Yang method revealed almost twice number of pixels (about 90%) containing clouds compared to other methods. However, upon visual analysis (Fig. 3, c), it can be seen that this method did not reveal individual pixels within the areas of continuous clouds, which were replaced by other methods. Therefore, the combined application of all three methods for the test image revealed the largest number of pixels with cloudiness (98% of the total number), making it possible to combine the advantages of different methods.

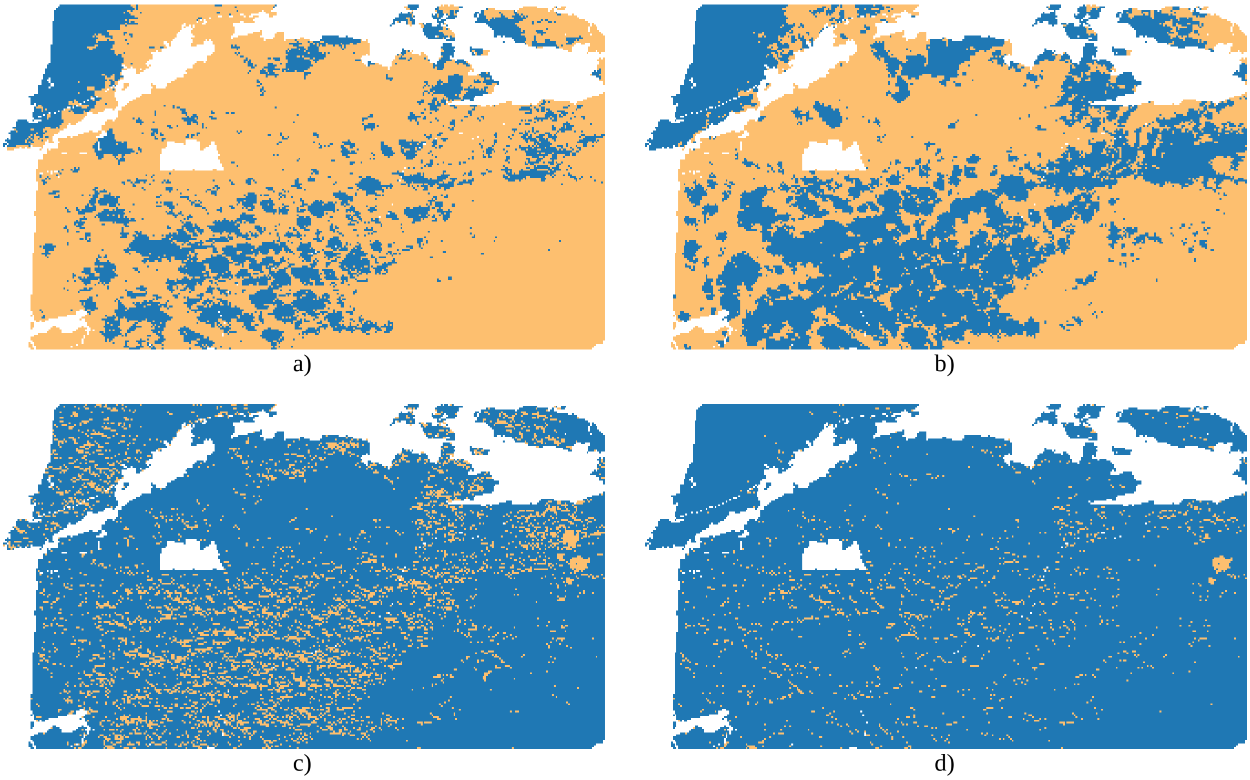

a)
b)
c)
d)

Fig. 3. Map of forest areas (blue color – pixels marked as containing clouds; orange color - pixels marked as not containing clouds; white color - pixels outside forest areas) as a result of cloudiness replacement: a) according to SCL classification; b) by the allowable range of values from 0 to 1000; c) by the Braaten-Cohen-Young method; d) sequential application of three methods

The research clearly demonstrates that the sequential application of various methods for identifying pixels containing clouds is most effective. At the same time, in the case of solving a highly specialized problem, filtering areas based on the allowable range of values is also a reason-able approach and allows you to identify a larger number of pixels containing clouds.

REFERENCES

[1] Braaten, Justin, W. B. Cohen and Zhiqiang Yang. “Automated cloud and cloud shadow identification in Landsat MSS imagery for temperate ecosystems.” Remote Sensing of Environment, 2015, pp. 128–138.

[2] Braaten-Cohen-Yang cloud detector | Sentinel-Hub custom scripts [Online].Available:https://custom-scripts.sentinel-hub.com/sentinel-2/cby_cloud_detection.

[3] Level-2A Algorithm - Sentinel-2 MSI Technical Guide - Sentinel Online [Online]. Aviabvle: https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

AUTHOR INDEX

Pattern Recognition and Information Processing

Proceedings of the 15th international conference

PRIP'2021

Распознавание образов и обработка информации

Доклады 15-й Международной конференции

На английском языке

Ответственный за выпуск С. С. Мойсейчик

Подписано в печать 20.10.2021. Формат 70×100 1/16.
Гарнитура Таймс. Бумага офсетная. Усл. печ. л. 39,0.
Уч.-изд. л. 42,0. Тираж 30 экз. Заказ 7.

Государственное научное учреждение «Объединенный институт проблем
информатики Национальной академии наук Беларуси».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 1/274 от 04.04.2014. ЛП № 02330/444 от 18.12.2013.
Ул. Сурганова, 6, 220012, Минск.