

This paper considers the structural-parametric synthesis (SPS) of neural networks (NNs) of deep learning, in particular convolutional neural networks (CNNs), which are used in image processing. It has been shown that modern neural networks may possess a variety of topologies. That is ensured by using unique blocks that determine their essential features, namely, the compression and excitation unit, the attention module convolution unit, the channel attention module, the spatial attention module, the residual unit, the ResNeXt block. This, first of all, is due to the need to increase their efficiency in the processing of images. Due to the large architectural space of parameters, including the type of unique block, the location in the structure of the convolutional neural network, its connections with other blocks, layers, computing costs grow nonlinearly. To minimize computational costs while maintaining the specified accuracy this work set tasks of both the generation of possible topology and structural-parametric synthesis of convolutional neural networks. To resolve them, the use of a genetic algorithm (GA) has been proposed. Parameter configuration was implemented using a genetic algorithm and modern gradient methods (GM). For example, stochastic gradient descent with momentum, accelerated Nesterov gradient, adaptive gradient algorithm, distribution of the root of the mean square of the gradient, assessment of adaptive momentum, adaptive Nesterov momentum. It is assumed to use such networks in the intelligent medical diagnostic system (IMDS), for determining the activity of tuberculosis. To improve the accuracy of solving the classification problem in the processing of images, the ensemble structure of hybrid convolutional neural networks (HCNNs) has been proposed in the current work. The parallel structure of the ensemble with the merged layer was used. Algorithms of optimal choice and integration of features in the construction of the ensemble have been developed

Keywords: *hybrid convolutional neural network, genetic algorithm, ensemble, structural-parametric synthesis*

DESIGN OF HYBRID NEURAL NETWORKS OF THE ENSEMBLE STRUCTURE

V. Sineglazov

Doctor of Technical Sciences,
Professor, Head of Department
Department of Aviation
Computer-Integrated Complexes
National Aviation University
Lubomyr Husar ave., 1, Kyiv,
Ukraine, 03058
E-mail: svm@nau.edu.ua

A. Kot

Postgraduate Student
Department of Computer-Aided
Management and Data
Processing Systems
National Technical University
of Ukraine «Igor Sikorsky Kyiv
Polytechnic Institute»
Peremohy ave., 37, Kyiv, Ukraine, 03056
E-mail: anatoly.kot@gmail.com

Received date 15.12.2020

Accepted date 02.02.2021

Published date 26.02.2021

Copyright © 2021, V. Sineglazov, A. Kot

This is an open access article under the CC BY license
(<http://creativecommons.org/licenses/by/4.0>)

1. Introduction

Currently, healthcare information support systems are actively developing. One of the promising directions of the modern stage of health informatization is the development of intelligent medical diagnostic systems that provide support for decision-making by a doctor. This is primarily due to the lack of sufficient experience from doctors, the rapid development of medicine, and the lack of time resources for improving the skills and experience of staff. As a result, patients undergo duplicative and useless expensive, and unnecessary treatments.

The intelligent element of IMDS is the neural network used both for the image processing from ultrasound studies (USS), computed tomography (CT), magnetic resonance imaging (MRI) studies and to support decision-making regarding the final diagnosis.

Several tasks arise when solving applied tasks in order to increase accuracy and reduce complexity. The first task is to find the optimal network topology. The second is structural (determining the number of hidden layers and

neurons in them, interneuron connections of individual NN) and parametric (setting of weight coefficients) optimization. One of the leading trends in modern computer science is the development of hybrid NNs. Hybrid neural networks (HNNs) consist of different structures united in the interest of achieving the goals based on deep learning. This makes it possible to solve complex problems, first of all, the processing of medical images, which cannot be solved on the basis of individual methods and technologies. The most effective means of image processing are convolutional neural networks (CNNs). A convolutional neural network is built on the basis of a convolution operation, which makes it possible to train CNN on certain parts of the image, iteratively increasing the local learning area of a separate convolutional nucleus.

There are a number of problems when using CNNs. The first is that it is necessary to allocate a large number of signs that determine the object of research. That requires an increase in the number of layers, that is, the complication of the neural network. The second issue is that with an increase in the number of layers, the learning process by an

error backpropagation method is complicated. On the one hand, to implement this algorithm, one needs to calculate a local gradient for each layer. On the other hand, the local gradient falls from layer to layer, thereby the adjustment accuracy is compromised. We have a significant increase in computational costs and a drop in accuracy. Therefore, it is a relevant task to investigate methods for constructing convolutional neural networks and ensembles of neural networks to increase the accuracy of their operation, while reducing the cost and time of their adjustment.

2. Literature review and problem statement

Convolutional neural networks [1, 2] have proved extremely successful for a wide range of computer vision tasks and other applications. However, it should be noted that their configuration is very computationally and temporally consuming if you do it manually. Setting up a CNN involves the optimal choice of the structure of the CNN and the subsequent adjustment of parameters. To date, active scientific activity is underway in this direction. In work [3], the optimal choice of the structure of CNN and parameters is determined using a hybrid genetic algorithm. In order to reduce calculations, a preliminary selection of significant parameters of CNN for a given training sample is made, which takes additional time. In work [4], the optimal choice of structure involves using a genetic algorithm but the options for this choice are limited and cannot be considered optimal. [5] proposes the use of a modified evolutionary algorithm to select the optimal structure but this may only apply to classical CNN and does not provide the high accuracy of the solution to a classification problem. In [6], instead of striving to choose a single optimal architecture, it is proposed to use a specially built matrix, in which a large number of architectures are built. The structure consists of a three-dimensional grid that connects feedback maps at different levels, scales, and channels, with a sparse homogeneous local communication template. This approach is computationally-intensive and costly and actually imposes restrictions on the choice of optimal architecture. Work [7] proposes a meta-modeling algorithm based on the reinforced learning, for the automatic creation of high-performance architectures of CNN for this educational task. The disadvantage of the cited work is the use of the classical structure of CNN, which significantly limits the class of tasks, due to computational difficulties (a drop in the gradient in depth learning). [8] proposes a method of accelerating the choice of architecture by studying the auxiliary HyperNet, which generates the weight of the main model, depending on the architecture of this model. This approach limits the ability to include a new modern topology in the structure of the CNN, which limits its functionality. Work [9] proposes a new paradigm for the design of convolutional architecture and describes a scalable method for optimizing convolutional architectures, which uses a search method of training with reinforcement to optimize the configuration of the architecture. This approach is computationally-intensive and complex and does not provide high accuracy. Paper [10] suggests an approach that combines a new hierarchical scheme of genetic representation that mimics a modular design pattern. This approach is used by experts, but the search space for optimal topology is limited. Deep learning has gained popularity in medical imaging studies, including magnetic resonance imaging of the brain [11],

breasts. ultrasonic detection of cancer [12]. Recently, U-Net is a popular approach to deep learning in biomedical imaging research proposed in work [13]. U-Net makes it possible to use data magnification, including the use of non-tough deformations, to fully use the available annotated sample images to train the model. These aspects suggest that U-Net could potentially provide satisfactory results with a limited amount of biomedical datasets currently available.

The researchers made significant contributions by offering different deep learning structures to identify and segment damage. Work [14] offered very deeply residual networks of more than 50 layers for a two-step segmentation of the lesion frame, followed by classification. It has been argued that deeper networks emit richer and more characteristic signs for recognition. The cited work showed promising results but the two-step structure and very deep networks were expensive in terms of computational costs.

In [15], convolutional networks were proposed where a parallel integration approach was implemented to segment the damage to ensure the unification of results that improved detection. The end-to-end fully automatic method of segmentation of damage using a 19-layer deep convolutional neural network is proposed in [16]. The loss function was introduced using Jacquard's distance as a measurement. To fine-tune the hyperparameters, a 5x cross-checking was used to train the ISBI dataset to determine the best performer. Paper [17] proposed completely convolutional methods for multiclass segmentation in the ISBI dataset for 2017. Works [18, 19] suggested a two-step segmentation method that employed Faster-RCNN in the first stage, and then a modified version of U-Net and a deep extreme method, respectively, as a second stage to achieve segmentation results. In [20], two deep learning classification models were used to recommend the most appropriate method of segmentation of the ISIC-2017 data set. In [21], a convolutional network (FrCN) was proposed to study the full resolution features of each pixel of the lesion images for segmentation.

Based on the study of literary sources [1–21], we can conclude that at the moment there is no procedure for the structural and parametric synthesis of the construction of hybrid convolutional neural networks of the ensemble structure, which is why it is not possible to solve the problems of classification for the creation of intelligent medical diagnostic systems whose operation requires the identification of as many signs of illness as possible and the processing of a large amount of data in order to make a correct diagnosis. For example, the task of determining the degree of tuberculosis in patients. Therefore, there is a need to determine the properties of unique blocks and their use to create a new topology, namely, hybrid convolutional neural networks.

3. The aim and objectives of the study

The purpose of this work is the structural-parametric synthesis of hybrid convolutional neural networks of the ensemble structure for their use in intelligent medical diagnostic systems.

To accomplish the aim, the following tasks have been set:

- to investigate the unique blocks (modules) of modern convolutional networks, their functionality, and properties for their use in hybrid neural networks;

- to develop a two-step procedure for determining the structure and parameters of a hybrid neural network with the formation of its binary representation;

- to develop an algorithm for the formation of an ensemble of hybrid convolutional neural networks;
- to check the proposed algorithmic support using the example of processing computed tomography of lung examination in order to detect tuberculosis in the presence of tuberculosis.

4. The study materials and methods

The basis of a CNN construction is the use of a convolution operation in order to be able to train the CNN on certain parts of the image. The size of these parts of the image is determined by the dimensionality of the corresponding convolution filter. It should be noted that neurons that correspond to the same convolutional filter have common weights, which provides a decrease in the computational costs of CNN gain compared, for example, with a multilayer perceptron. NN layers that are constructed in this way are called convolution layers.

In order to reduce computational costs, the CNN includes layers of aggregation (pooling), performing the functions of reducing the dimensionality of the sign map. Based on certain signs, an abnormal area is formed, for the classification of which full-linked layers are used (a classifier, which is located at the output of the network). Convolutional networks are built according to the rule, namely: first convolutional layers are placed, their number is determined as a result of solving the problem of structural-parametric synthesis, and then the aggregation layer is placed [22, 23]. The number of such iterations in the network depends on the complexity of the task.

The main parameters of the convolutional neural network [22] are:

- the size of the convolution kernel (filter);
- the number of convolution filters (depends on the number of convolutional layers);
- the amount of displacement when moving the convolution filter along the matrix of the image (a step of the convolution filter);
- parameters (vertically and horizontally) taking into consideration the edge effects (the initial position of the convolution filter on the matrix of the image or the map of signs before moving in order to build an feature map);
- the initial filling of the convoluted filters.

The number of CNN inputs is determined by the number of pixels that make up the image. According to work [22], if we take into consideration the number of inputs, layers, features maps, the number of parameters, the optimal value of which must be determined in the structural and parametric synthesis of CNN, can be very large.

Solving such an optimization problem directly is not possible. Therefore, approaches were considered both to reduce the number of parameters that are optimized and the problems of choosing or developing new optimization methods. Multicriteria methods were studied as optimization methods: genetic, swarm, and modern gradients.

The general rules for reducing the number of parameters do not greatly affect this process – it all depends on the training sample. Therefore, it is proposed to determine the most significant parameters of CNN loss in terms of efficiency as a result of an experiment on a convolutional neural network. The experiment is carried out as follows: step by step, one changes one of the parameters of the CNN,

with fixed other parameters, and determines how much the output has changed.

When choosing optimization algorithms, the following criteria were used: accuracy, computational and time costs. None of the well-known multicriteria optimization algorithms provided the proper results. Therefore, it was decided to develop a hybrid algorithm, in which a genetic algorithm was used as a base, but with the imposition of certain restrictions on the value of individual structural parameters. This is done in order to reduce computing costs. Detailed research in this area was presented in paper [3].

To train a convolutional neural network, this work employs the normalized initialization, which is called Glorot initialization [22].

The task that considers the effectiveness of the proposed approach, the task of structural and parametric synthesis of hybrid CNNs was chosen, to determine the degree of activity of tuberculosis in patients according to the results of CT studies. Processing ultrasound, KT, MRI examinations are considered the most complex because medical images are poorly structured. A modern tomograph was used as the hardware. The study of certain components of the hybrid CNN was carried out using the TensorFlow programming environment.

The sample was borrowed from the state-run institution “National Institute of Phthiology and Pulmonology named after F. G. Yanovsky, National Academy of Medical Sciences of Ukraine” according to the results of the study of patients with suspected tuberculosis (Kyiv, Ukraine). The sample consists of slices of computer tomography studies that corresponded to patients with an accurate diagnosis (either there is a disease or not). The sample was divided into two parts: 80 % – training set, 20 % – test set. The training sample was used to train; the test set determined the accuracy of performance of the trained hybrid CNN or the ensemble of CNN.

5. Results of studying the structural and parametric synthesis of HCNN

5.1. Investigation of unique blocks (modules) of modern convolutional networks, their functionality, and properties

One of the possible options for hybridization is the construction of HCNN based on the use of the known topology of CNN. Therefore, before moving directly to the hybridization process, it is necessary to consider the known topology of HCNN.

Depending on the type of architectural modifications, CNN can be divided into seven different classes:

- CNN based on spatial use;
- CNN based on depth usage;
- branched CNN;
- CNN with a set of connections based on width;
- CNN based on the use of the features map;
- CNN based on boosting channels;
- CNN based on the use of the attention mechanism.

The classification of CNN architectures is visually shown in Fig. 1 [24].

The construction of HCNNs, which consist of different CNNs, is considered impractical due to the increase in computational costs, despite certain advantages over the use of individual modern CNNs by increasing accuracy. More expedient is to use their individual parts.

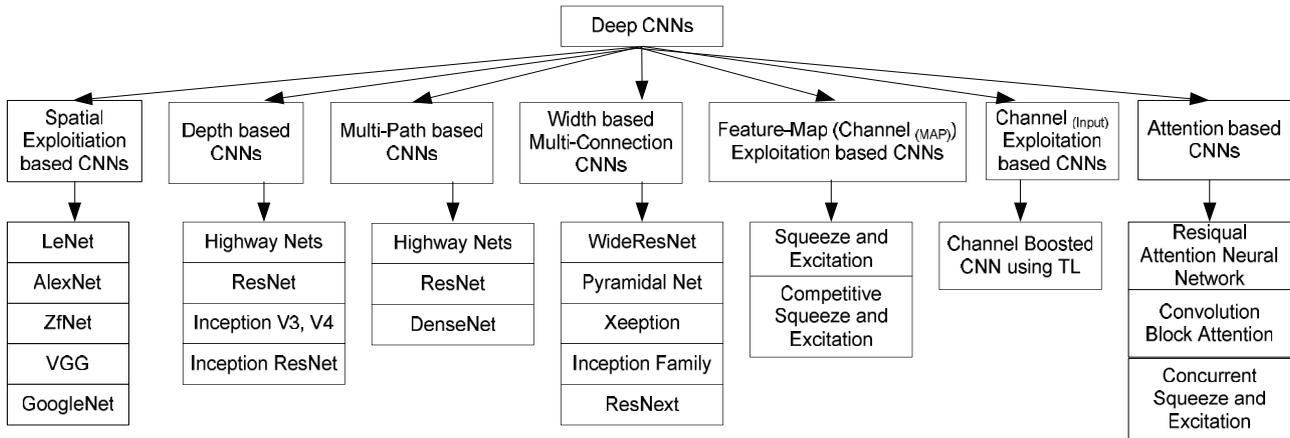


Fig. 1. Classification of CNN deep architectures

A characteristic feature of modern CNNs is the presence of unique blocks that determine their essential features. For example, compression and excitation unit, attention module convolution unit, channel attention module, spatial attention module, residual unit, initial module, ResNeXt unit [25–31]. The topology of the above-mentioned blocks (modules) is shown in Fig. 1. Unique blocks (layers) include Batch Normalization [32] and Reduction Block [25].

As a result of our study, the following unique blocks were selected:

- batch normalization;
- simplification unit;
- compression and excitation unit;
- highway network block;
- residual unit;
- inception unit;
- attention unit.

Batch normalization.

Batch normalization is considered as another layer that is inserted into the model architecture, as well as full-linked or convolutional layers [32]. In practice, batch normalization layers are inserted after a convoluted or fully interconnected layer but before the source data is transferred to the activation function. Batch normalization will be used to normalize the input layer by re-centering and scaling. Each layer of the neural network has inputs with a corresponding distribution, which, during the learning process, is affected by randomness in the initialization of parameters and in the input data. The influence of these sources of randomness on the distribution of inputs into the inner layers during training is defined as an internal covariant shear.

Simplification unit.

The issue related to deep convolutional neural networks is that the number of features maps often increases with the depth of the network. This issue can result in a sharp increase in the number of parameters and calculations required when using larger filter sizes (convolution filters) such as 5×5 and 7×7.

To resolve this issue, a convolutional layer 1×1 is used, which combines channels, often called the pooling (aggregation) of features maps or a projection layer. This

technique is used to reduce the dimensionality, reducing the number of features maps while retaining the noticeable features. This is also used directly to create an individual projection of the features map, to combine features across channels, and to increase the number of features maps after traditional pooling layers.

Squeeze and Excitation Block.

The structure of the Squeeze and Excitation Block (SEB) is shown in Fig. 2.

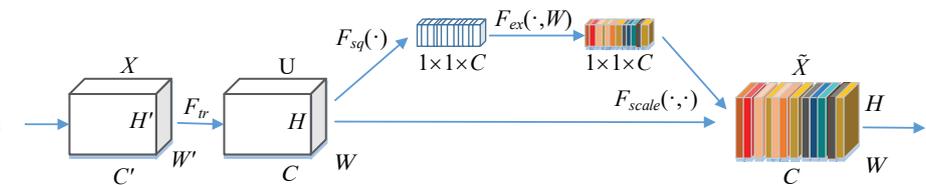


Fig. 2. Structure of the Squeeze and Excitation Block

The following operations are performed in the Squeeze and Excitation Block:

- converting features or a simple convolution operation to the inputs X to get attributes U ;
- compression operation to get one U output value for each channel;
- excitation operation, which is applied to the outputs of compressed data in order to obtain weight coefficients for each channel;
- scale the features map U with these activations to get the resulting output data of the SEB unit. The role this operation plays at different depths differs over the network.

At earlier levels, SEB excites informative features independently of the class, amplifying lower-level general perceptions. At later levels, SEB units are becoming more specialized and responding to various inputs that strictly depend on the specificity of the class.

Highway network block.

Highway networks allow for the unhindered flow of information through multi-layer layers of information through interlayer connection. The structure of the highway network block (HNB) is shown in Fig. 3.

In work [30], the training of the highway neural network unit is reformulated to achieve the following:

- give preference to optimization in the early stages of training, when the selection is mainly directed to unconverted features;
- focus on transforming training features at later stages of training, when transformed features mainly pass through selection;
- rely on a much smaller number of selections to train several layers of transformation features, since the transition of selection is more effective for optimizing and generalizing the model. To achieve the above-mentioned characteristics of training, a new highway unit in the form shown in Fig. 4 is proposed.

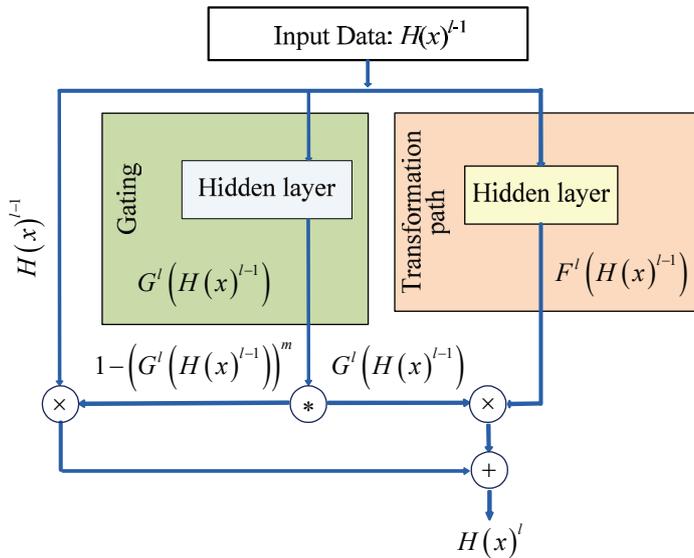


Fig. 3. The structure of a highway network block:
 $H(x)^{l-1}$ – information on the layer ($l-1$) of the highway network;
 $G^l(H(x)^{l-1})$ – output of the selection module;
 $F^l(H(x)^{l-1})$ – conversion trained on a hidden layer l to input $H(x)^{l-1}$

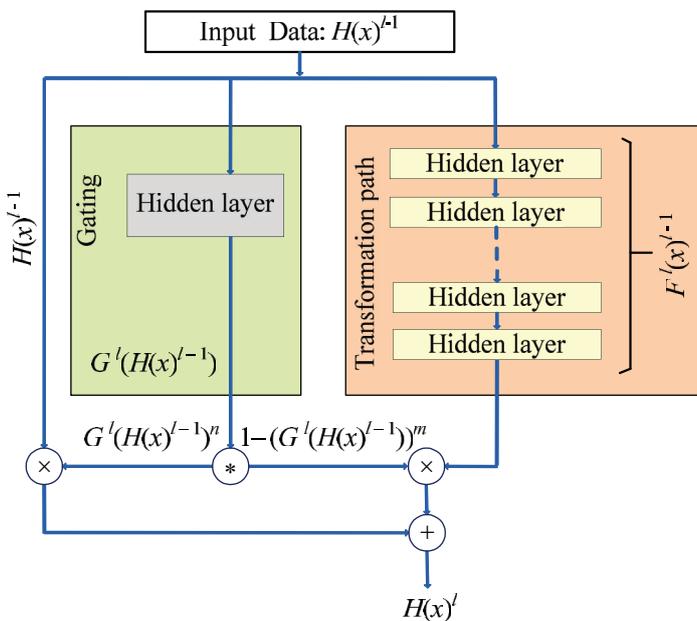


Fig. 4. Proposed highway network block:
 $H(x)^{l-1}$ – information on the layer ($l-1$) of the highway network;
 $G^l(H(x)^{l-1})$ – output of the selection module;
 $F^l(H(x)^{l-1})$ – conversion trained on a hidden layer l to input $H(x)^{l-1}$

Highway network blocks were used in HCNN in order to improve their efficiency for processing video information.

Residual block.

Deep networks pull low-, medium-, and high-level features through multilayered ways, and increasing the number of layers or blocks consisting of several layers can enrich the levels of features. However, as the depth of the learning network increases, it becomes unstable, and the accuracy achieved begins to decrease (degrade). This is due to damping the gradient in the reverse run in an error backpropagation method and, as a result, the deterioration of neural network performance. The structural diagram of the residual block is shown in Fig. 5.

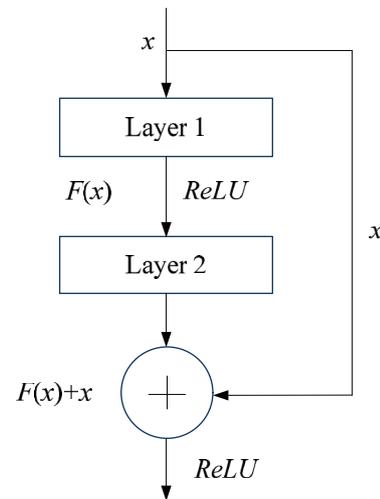


Fig. 5. Residual Block Structural Diagram

This block uses the so-called shortcut connections, that is, an identical mapping is explicitly added. As a result, the reverse run in an error backpropagation method produces: $dF(x)/dx+1$. Thus, the gradient will not fade, because there will always be a reverse run. This design requires that the output data of the two convoluted layers have the same shape as the incoming layers so that they can be put together. To change the number of channels, somebody must introduce a simplification unit (an additional convolution layer 1×1) to convert the input information x into the desired form to perform the addition operation. Convoluted 1×1 simply displays the input pixel with all its channels to the original pixel, no matter what around is. This convolution is used to reduce the number of depth channels. Taking into consideration the need to use batch normalization and an additional convolutional layer 1×1 , the structure of the residual block takes the form shown in Fig. 6. Thus, the next layer does not lose its degree of freedom to shift and scale the input data but is engaged only in assessing their structural properties; as a result, the acceleration of the convergence of the learning process is obtained.

The use of the residual block as part of the HCNN makes it possible to reduce computational costs and improve the processing results by increasing the number of layers of NN.

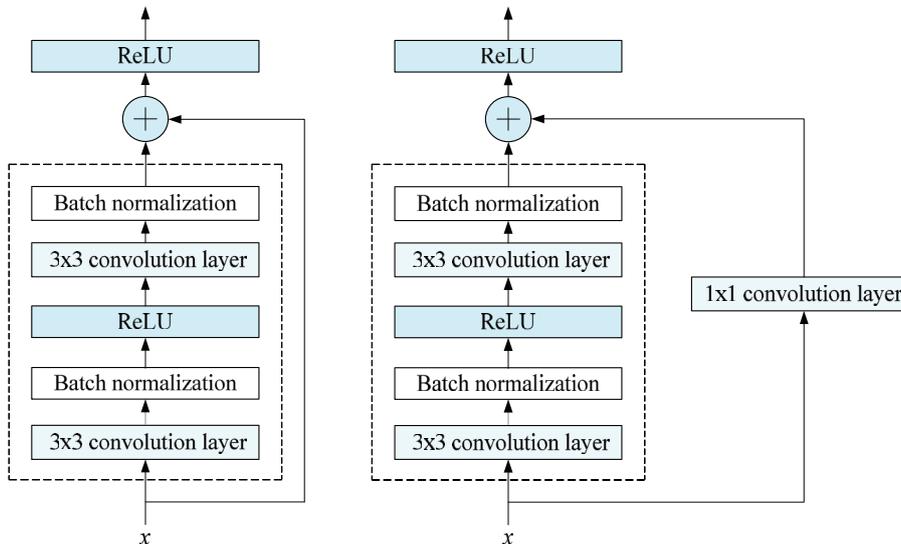


Fig. 6. Residual block in the presence and absence of a convolution layer 1x1

Inception unit.

There are two ways to improve network quality – increasing the depth and increasing the width, but this relates to an increase in the likelihood of retraining and inefficient use of computing resources – the sparse structure of convolution is inefficient in the computational sense. To eliminate these shortcomings, the correlation structure of activations of the previous layers is used. The structure of the Inception block is shown in Fig. 7.

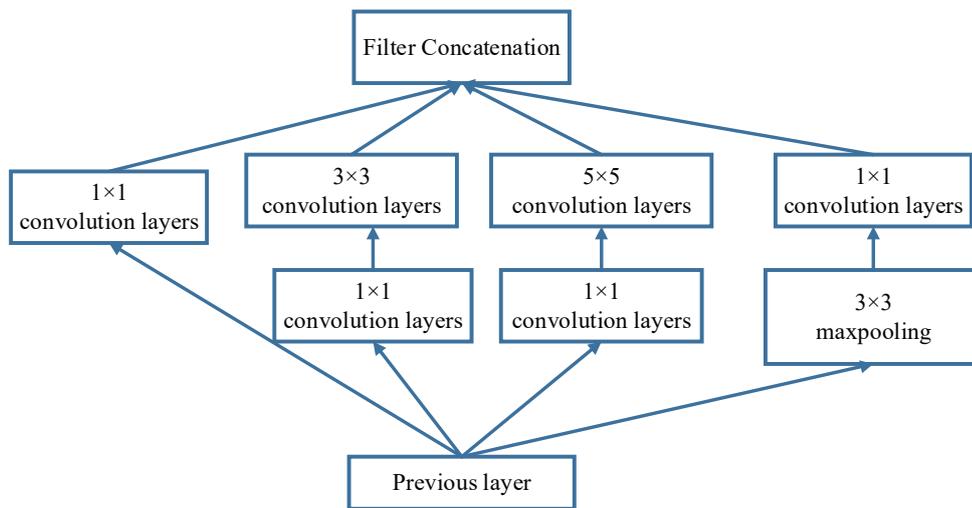


Fig. 7. Inception unit structure

Each block has convolution layers with a filter of different sizes to recognize features of different scales. In addition, in this model, convolution 1x1 is used to reduce the dimensionality of the tensors that will be fed to the input of the next layer. In order not to lose the information obtained in the previous layer, a subsampling layer is used. After it, a convolutional layer with a convolution filter of 1x1 is also used, in this case, in order to align the dimensionality of tensors at the output after each parallel layer. Then there is the concatenation of the features maps obtained on each parallel layer.

Attention unit.

Attention mechanisms are an approach in machine learning, which implies separating part of the input data (regions of images, text fragments) for more detailed processing.

Often, to solve the task of image classification, you do not need to process all the pixels of the image: for example, in a classification problem, the background often plays a minor role. However, convolutional networks, which are the most popular method of solving such a task, use the same amount of computing resources on all parts of the image. The attention block (module) is implemented in two versions: the channel attention module and the spatial attention module.

Channel attention module and the spatial attention module.

Channel attention module. A channel attention map is created using the relationship of features between channels. Since each channel of the features map is considered as an object detector, the channel's attention focuses on what matters given the input image. To effectively calculate the channel's attention, compress the spatial size of the input card. For the generalization of spatial information, averaging aggregation is still common (Fig. 8).

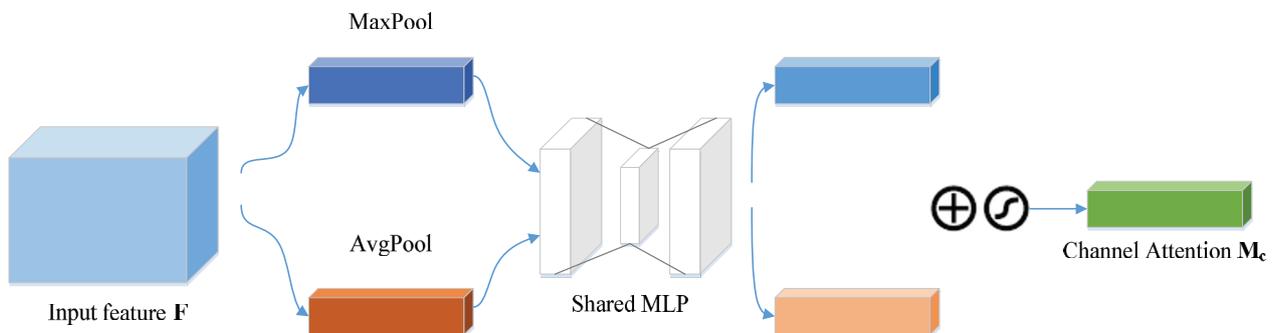


Fig. 8. Structural diagram of the channel attention module

Spatial attention module. A spatial attention map is generated using an interspatial relationship of features. Unlike the channel's attention, spatial attention focuses on where there is an informative part that complements the channel's attention. To calculate the spatial attention, first, apply the operations of the mean and maximum join along the channel axis and combine them to create an effective feature descriptor. It has been shown that the use of join operations along the channel axis effectively affects the selection of informative areas [21]. On the combined feature descriptor, a convolution layer is used to form the spatial attention map $M_s(\mathbf{F}) \in \mathbf{R}^{H \times W}$ that encodes where to emphasize or suppress. A detailed structural diagram is shown in Fig. 9.

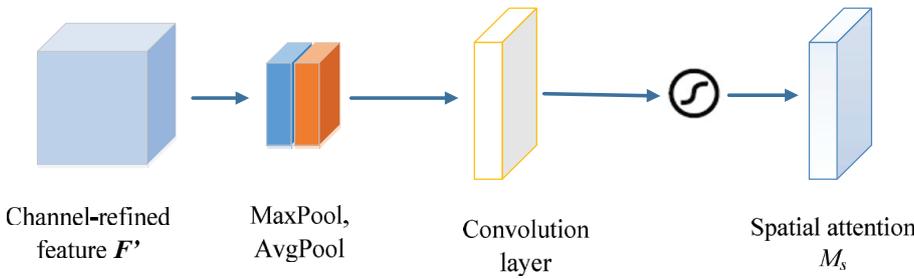


Fig. 9. Structural diagram of the spatial attention module

One of the main issues that hinder the further progress of using CNN is a large architectural space of parameters, including the type of unique block, the location in the structure of the CNN, its links with other blocks and layers. As a result, there is a task of the structural and parametric synthesis of CNN.

5.2. Two-step procedure for determining the structure and parameters of a hybrid neural network

The synthesis of hybrid convolutional neural networks used mainly for imaging is much more complicated than the synthesis of convolutional neural networks. Such synthesis requires determining the type of unique blocks to be used, aligned with adjacent blocks, their locations.

These difficulties require first the solution to the structural and parametric synthesis of classical CNN based on a given training sample. Classic CNN consists of convolutional layers, each neuron of which performs a convolution of some area of the previous layer, aggregation layers (pooling), performing the functions of reducing the dimensionality of the features map, and full-reconnected layers (a classifier located at the output of the network). Convolutional layers and aggregation layers may alternate. Most often, the layers of aggregation are placed after the layers of convolution [29, 30]. As a basic architecture of CNN, based on the application of the approach to determine the most significant, in terms of efficiency, parameters of CNN and train a convolutional neural network, the structure and parameters of the basic CNN are determined. For example, for a training set, the structure of NN was as follows: convolutional layer, convolutional layer, subsampling layer, convolutional layer, convolutional layer, subsampling layer, full-linked layer (classifier).

The formation of a hybrid structure by including unique blocks or an ensemble of hybrid convolutional neural networks in the basic CNN provides new opportunities for increasing the efficiency of solving the set problem. The task of the structural and parametric synthesis of hybrid CNNs is solved on the basis of using a multi-stage procedure for

determining the structure and parameters of hybrid CNN with the formation of its binary representation.

The procedure for the synthesis of hybrid neural networks can be represented as the following sequence of operations:

1. The structural and parametric synthesis of basic CNN.
2. The structural and parametric synthesis of hybrid CNN with the determination of types and sequences of unique blocks that are introduced into the hybrid CNN.
3. The structural and parametric synthesis of the ensemble, consisting of separate hybrid CNNs.

In accordance with the proposed procedure, as a result of the first stage, a basic convolutional NN was obtained, which does not have unique blocks. In the case when this does not

meet the quality criterion for solving image processing, you should use a hybrid CNN. Let us call the hybrid convolutional network the one that includes various unique blocks, the choice and placement of which will be determined on the basis of the use of a genetic algorithm. It is possible to use more complex blocks, namely:

- Attention-Inception Block;
- Inception-Dense Block;
- Inception-Residual Block;
- Inception-Recurrent Residual Block;
- Deep Pyramid-Highway Block;
- Deep Pyramid-Self Attention Block;
- Squeeze and Excitation-Attention Block;
- Squeeze and Excitation-Dense Block;
- Squeeze and Excitation-Inception Block;
- Squeeze and Excitation-Residual Block;
- Residual-Squeeze and Excitation-Attention Block;
- Residual-Attention Block.

Many modern network structures [33, 34] can be divided into several groups. In each group, the geometric dimensions (width, height, and depth) of the grouping cube remain unchanged. Neighboring groups are connected by a spatial pooling operation. The structure of the basic NN in a general case represents the alternation of two convolutional layers, followed by a pooling layer, so the hybrid CNN is built on the basis of basic CNN by replacing the convolutional layers with a grouping. All convolution operations in the same group have the same number of filters or channels.

Binary term representation is provided for the network structure in a limited case. First of all, note that many modern network structures can be divided into several blocks. In each block, the geometric dimensions (width, height, and depth) of the layer cube remain unchanged. Adjacent blocks are connected by a spatial pooling operation, which can change spatial resolution. All convolution operations on one block have the same number of filters or channels (Fig. 10).

Each hybrid CNN consists of S groups; the s -th grouping, $s=1, 2, \dots, S$, contains K_s blocks denoted v_{s, k_s} , $k_s=1, 2, \dots, K_s$. Nodes in each block are sorted, and connections from a node with a lower number to a node with a higher number are allowed. Each node has a unique block. The full-linked network part is not encoded. In each group, use $\frac{1}{2}K_s (K_s-1)$ bits to encode cross-site links. The first bit represents the relationship between $(v_{s,1}, v_{s,2})$, then the next two bits represent a connection, between $(v_{s,1}, v_{s,3})$ and $(v_{s,2}, v_{s,3})$, etc. This process continues until the last bits K_s-1 are used to represent a connection between $v_{s,1}, v_{s,2}, \dots, v_{s, K_s-1}$ and v_{s, K_s} .

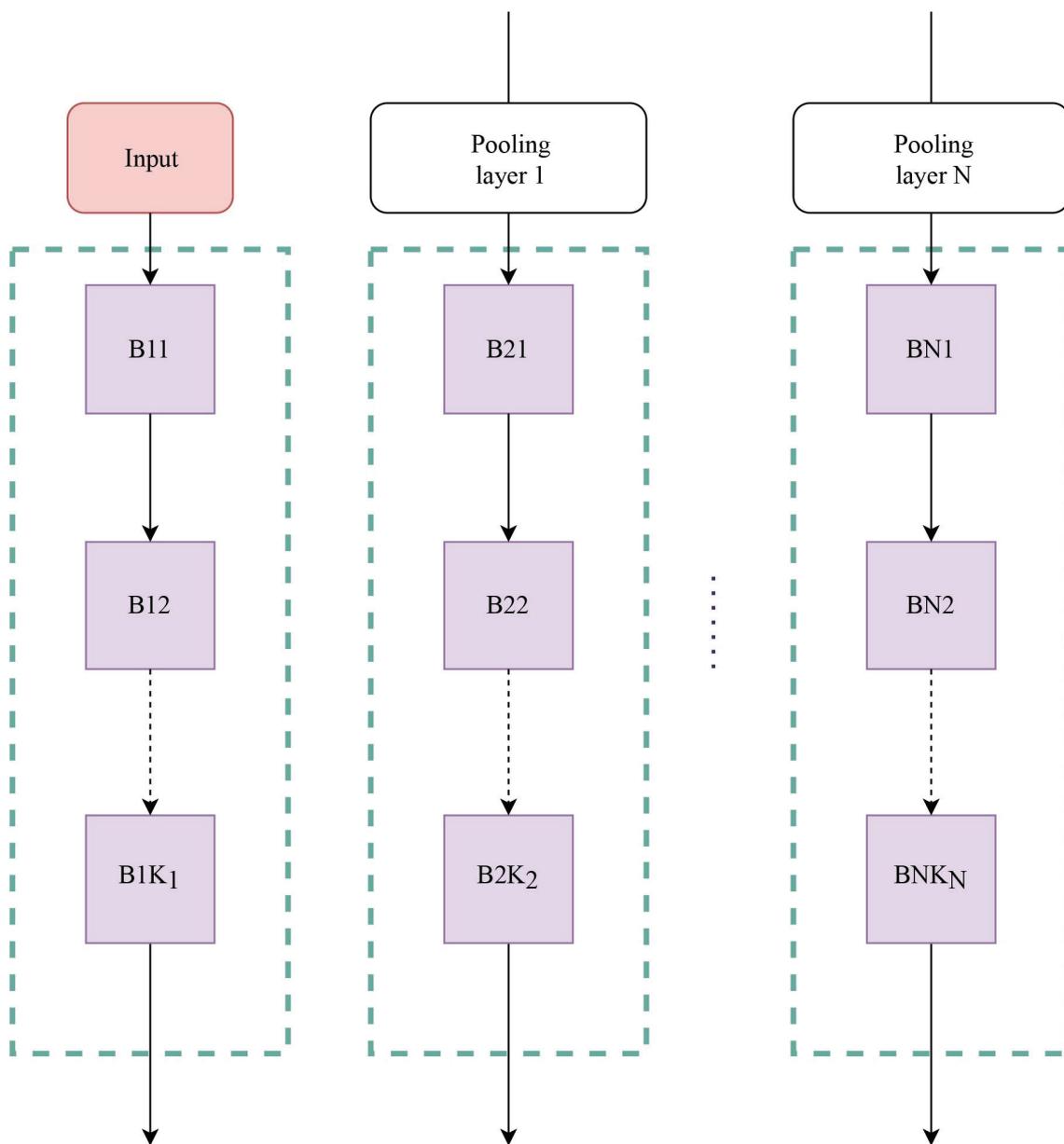


Fig. 10. Multiblock hybrid convolutional neural network

The selection process is carried out at the beginning of each generation. The t -th generation of the n -th individual $M_{t-1,n}$ is assigned a fitness function defined as the $r_{t-1,n}$ recognition speed, obtained in the previous generation or $r_{t-1,n}$ initialization directly affects the probability that $M_{t-1,n}$ is stored in the selection process [4].

The two-step optimization algorithm employed the following settings of the genetic algorithm: population size – 25 individuals, archive size – 25 individuals, number of iterations – 10, probability of crossing – 80 %, probability of mutation – 20 %. The objectives of the genetic algorithm were to minimize the value of the classification error and maximize the accuracy of the neural network classification.

The following optimizer settings were used during the experiments: for the fastest descent algorithm, Adagrad, RMSProp, and Adam, we used a training factor of 0.01; for an accelerated Nesterov gradient, the training coefficient – 0.01, the pulse factor – 0.9 were used; the learning factor

with a value of 1 was used for the Adadelta optimizer. The values of coefficients were chosen experimentally in order to increase the efficiency of the algorithms.

To compare, the learning using each optimization algorithm was carried out ten times, after which the average number of learning ages required to achieve an accuracy of 85 % was calculated. The comparison of the use of optimization algorithms with a two-step algorithm is given in Table 1.

Table 2 gives the percentage of performance improvements for the two-step algorithm compared to optimization algorithms and gradient descent in particular.

Table 3 compares the use of a two-level optimization algorithm using the algorithm of the fastest descent at the last stage and with the use of other two-step algorithms at the second stage of neural network configuration. The table shows a comparison of the number of epochs required to achieve a classification accuracy of 85 % and a percentage of the speed of algorithms.

Thus, the use of a two-step algorithm to optimize the parameters of the neural network has made it possible to increase the efficiency of training compared to using only optimization algorithms. This effect of performance increase is noticeable even with the use of a small number of iterations of the genetic algorithm at the first stage of the two-step algorithm.

Number of epochs to achieve 85 % accuracy

Optimization algorithm	Number of epochs	
	When configured only by the optimizer	When configuring with a two-level algorithm
The fastest descent	3,666	2,876
Accelerated Nesterov gradient	1,732.5	1,425.5
Adagrad	1,257	1,037
RMSProp	309	295.2
Adadelata	203.2	183.4
Adam	52.7	48.7

Table 1

Comparison of two-step algorithm optimizers

Optimization algorithm	Two-level algorithm performance gain, %	
	Compared to the algorithm of the fastest descent	Compared to a single-step optimization algorithm
The fastest descent	–	27.47
Accelerated Nesterov gradient	157.17	21.54
Adagrad	253.52	21.21
RMSProp	1,141.87	4.68
Adadelata	1,898.91	10.79
Adam	7,427.72	8.21

Table 2

Comparison of two-step algorithms with different optimizers

Optimization algorithm at the second stage	Faster operation compared to the two-step algorithm of the fastest descent	
	Epochs	%
Accelerated Nesterov gradient	1,450.5	101.75
Adagrad	1839	177.34
RMSProp	2,580.8	874.25
Adadelata	2,692.6	1,468.16
Adam	2,827.3	5,805.54

Table 3

5. 3. Algorithm of formation of an ensemble of hybrid convolutional neural networks

The ensemble of neural networks is a group of topologies, combined into a single structure, which can differ in architecture, learning algorithm, learning criteria, and types of resulting neurons [35–37]. In another version, the term ensemble refers to the “united model”, the output of which is a functional combination of outputs of individual models [38].

As a result of CT work, a set of CT images corresponding to certain slices was obtained. Each slice is processed by a separate component of the segmentation ensemble. The algorithm of formation of the ensemble of hybrid convolutional neural networks includes the solution to two problems: segmentation and classification. The algorithm for constructing

an ensemble of segmentators includes the following: the choice of aggregation method, the choice of segmentator types that are included in the ensemble, and the determination of criteria by which the quality of segmentation is evaluated.

Aggregation of results occurs in one of three methods:

- Ensemble-ADD (used in Fig. 11): combines the results of Mask R-CNN, DeepLabV3, and Deep Pyramid Attention Module to create the final segmentation mask;
- Ensemble-Comparison-Large: selects a large segmented area by comparing the number of pixels in the output data of all segmentators;
- Ensemble-Comparison-Small: on the contrary, selects a smaller segmented area at the output of all segmentators.

The ensemble-ADD method was used to build the ensemble. Ensemble-ADD components used the results of segmentator such as Mask R-CNNADD, Deeplab-V3+ADD, Deep Pyramid-Attention Module ADD to create the final segmentation mask.

Evaluate the effectiveness of algorithms using the Jacquard Similarity Index (JSI), sensitivity, specificity, accuracy, dice similarity coefficient, and Matthew correlation coefficient (MCC) [39, 40]:

$$\text{Sensitivity} = \frac{TP}{TP+FN}, \quad (1)$$

$$\text{Specificity} = \frac{TN}{FP+TN}, \quad (2)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}, \quad (3)$$

$$\text{JSI} = \frac{TP}{(TP+FP+FN)}, \quad (4)$$

$$\text{Dice} = \frac{2+TP}{(2+TP+FP+FN)}, \quad (5)$$

$$\text{MCC} = \frac{TP+TN-FP+FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \quad (6)$$

where TP – real positive results, TN – real negative results, FP – falsely positive results, FN – falsely negative results.

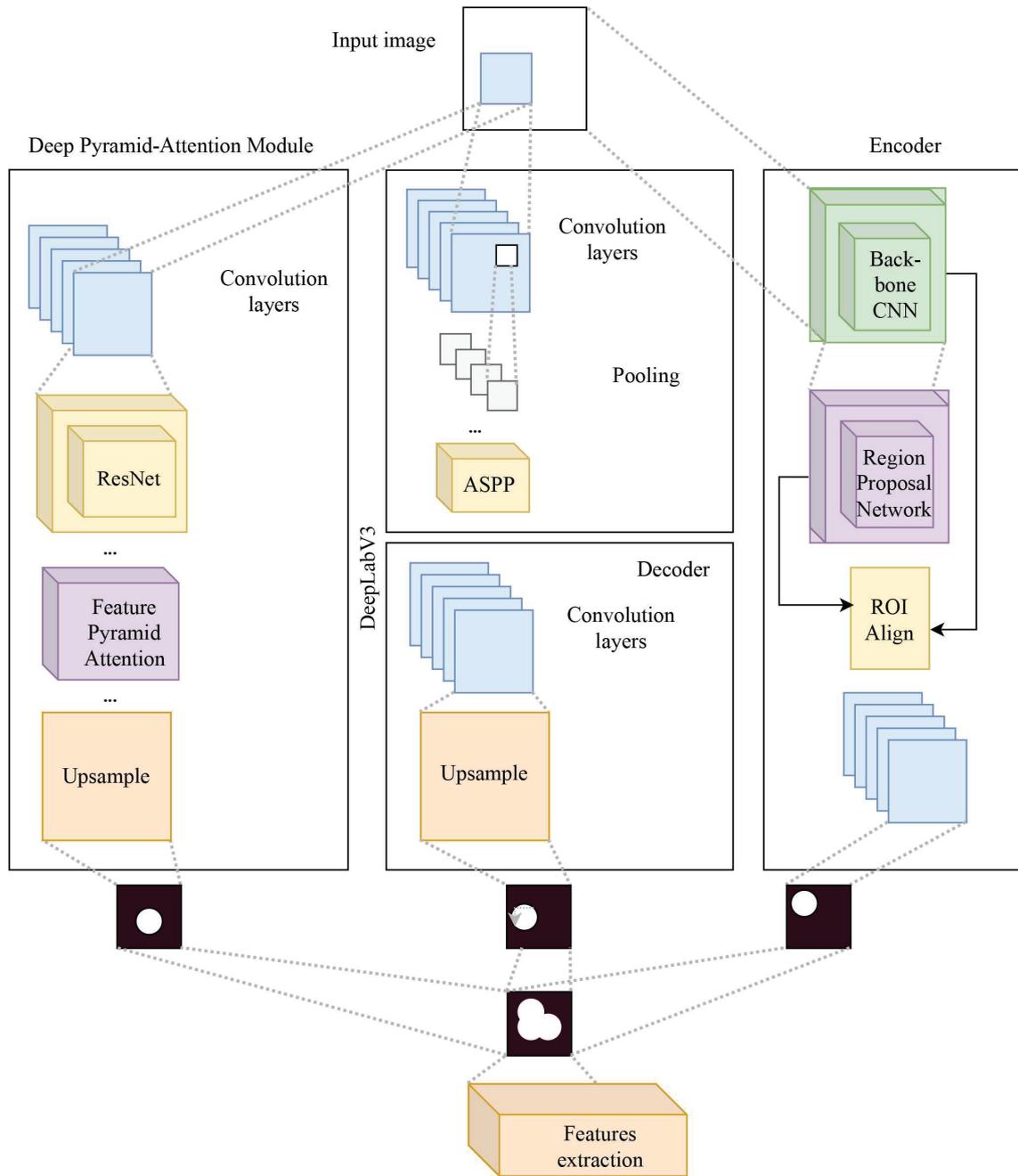


Fig. 11. The architecture of segmentation ensemble

5. 4. Testing the proposed algorithmic support

The proposed procedure of structural and parametric synthesis to create an ensemble segmentation system, which includes the choice of the segmentation method and the composition of the ensemble (Mask R-CNNADD, DeepLabV3+ADD, Deep Pyramid-Attention Module ADD), was used in the processing of CT studies (based on the cuts from a tomograph with the presence of areas suspicious of the disease) in determining the stages of tuberculosis activity in the diseased (Fig. 12).

Sensitivity is defined from equation (3), where TP are the true positive results and FN are the falsely negative ones. High sensitivity (close to 1.0) indicates good performance during segmentation, all lesions have been successfully seg-

mented. On the other hand, specificity (from equation (4)) shows the proportion of true negative (TN) among the intact. High specificity indicates the ability of the method not to segment the lesion cell. The accuracy of segmentation methods determines the percentage of pixels in an image that have been correctly classified, from equation (5). JSI and Dice are a measure of how similar predictions and reliability are by measuring the number of TP detected and FP fines found by the method, both from equation (6) and (7), respectively. MCC has a range from -1 (completely incorrect binary classifier) to 1 (the fully correct binary classifier). It is used to evaluate the effectiveness of segmentation algorithms based on binary classification (lesions compared to non-destructions), from equation (8).

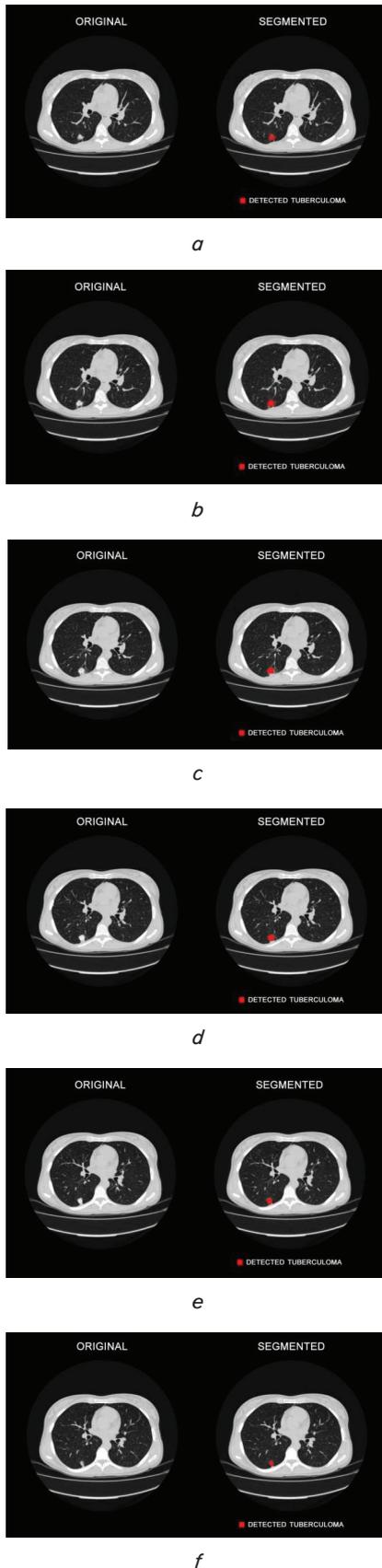


Fig. 12. Slices from a tomograph with the presence of areas suspicious of the disease when determining the stages of tuberculosis activity in patients: *a* – slice 1; *b* – slice 2; *c* – slice 3; *d* – slice 4; *e* – slice 5; *f* – slice 6

Based on ensemble segmentation, features are distinguished, which are input information for solving the classification problem. The problem of classification should be solved with the help of an ensemble of classifiers. Work [3] presents a detailed study of this issue: the need to build ensembles is substantiated, the optimal structure of ensembles is determined, parallel, the number and type of criteria for the selection of NN (classification models) in the ensemble – accuracy and diversity, the author developed an algorithm for determining the contribution of each component to the overall result of the ensemble to solve the problem of classification in order to build a rating of components.

Input data can be broken down into specific groups to handle different NN or be submitted to all networks at the same time.

The main difficulty of combining networks in an ensemble is the training of all components to solve the problem. In order to increase the effectiveness of training, NN learn separately, and then unite into a single structure. However, if the algorithms for setting up the selected topology belong to different classes of training, synchronous training of all ANN included in the ensemble is required, and, therefore, it is necessary to develop a single algorithm for setting up all the CNN in an ensemble. In the example under consideration (processing of CT studies (based on the slices from a tomograph with the presence of areas suspicious of the disease), NNs are trained separately.

Neural networks were trained on the NVIDIA Tesla K80 computing processor with 12 GB of dedicated video memory. For the implementation of neural networks, the Python programming language was used using the Keras library (with TensorFlow backend) as a high-level neural network library. The data were obtained as a result of processing the results of CT scans of tuberculosis patients with the detection of tuberculosis, their volume, density.

Such data, namely the accuracy of the solutions of individual convoluted classifiers, is given in Tables 4–8.

The architecture of the classification ensemble is shown in Fig. 13.

Table 4

Accuracy of solutions of individual convolutional classifiers without proposed segmentation architecture

Classifier	Training accuracy, %	Test accuracy, %
Residual-Squeeze Module	92.11	90.39
Squeeze and Excitation-Attention Module	97.98	91.93
Inception-Attention Block	97.11	91.74
Squeeze and Excitation-Residual Module	90.66	90.02

Table 5

The accuracy of the solution of the complete convolutional majoritarian ensemble

Training accuracy, %	Test accuracy, %
97.65	93.36

Table 6

Accuracy of solving individual classifiers with the proposed segmentation architecture

Classifier	Training accuracy, %	Test accuracy, %
Multilayer perceptron	94.23	91.82
Decision Tree	95.91	92.56
NefClassM	92.67	90.41

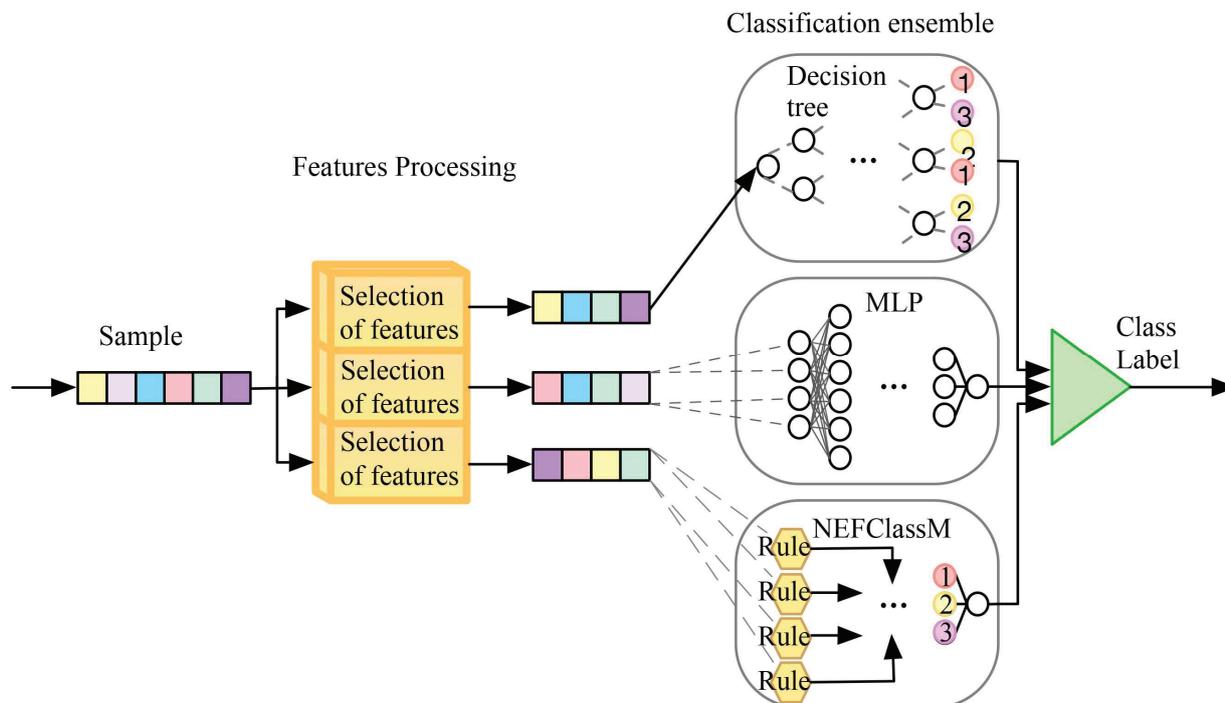


Fig. 13. The architecture of classification ensemble: Rule – rule; MLP – multilayer perceptron

Table 7

Accuracy of the full ensemble of classifiers with segmentation ensemble

Training accuracy, %	Test accuracy, %
96.6	97.14

Table 8

Accuracy of other different ensemble architectures

Ensemble	Training accuracy, %	Test accuracy, %
Random Forest	97.32	87.15
AdaBoost	54.96	54.25
Bagging Decision Tree	94.41	83.85
CNN bagging	94.96	91.39

The use of this structure will improve the quality of the solution to the classification problem, especially under difficult conditions in the presence of a large number of features of different nature.

6. Discussion of results of studying the structural and parametric synthesis of HCNN on the example of determining the degree of activity of tuberculosis

The study of existing modern convolutional neural networks led to the separation of individual functional blocks from them. Each of these blocks was investigated separately for the possibility of use for the synthesis of hybrid convolutional networks. Criteria for the possibility of such use were identified, namely: functional properties, the ability to combine with other blocks and layers, the ability to parameterize, the ability to isolate learning. As a result, the following blocks were separated that meet the specified criteria: batch normalization unit, simplification unit, compression

and excitation unit, highway network block, residual unit, Inception block, attention block.

The structural-parametric synthesis of a hybrid convolutional neural network, which would consist simply of the selected unique blocks, showed a nonlinear increase in learning time losses. Therefore, the procedure for the synthesis of hybrid neural networks was proposed. At the first stage, the structural and parametric synthesis of basic CNN was performed. At the second stage, the structural and parametric synthesis of hybrid CNN was performed, which consisted of the singled-out unique blocks.

To achieve the accuracy of 85 % (Table 1) by different optimization algorithms, the difference between the number of epochs only by the optimizer and the proposed method was:

- for the algorithm of the fastest descent – 790 epochs, which corresponds to a win of 21.55 %;
- for the accelerated Nesterov gradient – 307 epochs, which corresponds to a win of 17.72 %;
- for Adagrad algorithm – 220 epochs, which corresponds to a win of 17.50 %;
- for RMSProp algorithm – 13 epochs, which corresponds to a win of 4.47 %;
- for Adadelta algorithm – 19 epochs, which corresponds to a win of 9.74 %;
- for Adam algorithm – 4 epochs, which corresponds to a win of 7.59 %.

Compared to the performance of the two-step algorithm against optimization algorithms (Table 2), it turned out that the winnings range from 8.21 % (in the case of the Adam algorithm) to 27.47 % (in the case of the algorithm of the fastest descent).

In comparison, the use of a two-level optimization algorithm using the algorithm of the fastest descent at the last stage and with the use of other two-step algorithms at the second stage of neural network configuration (Table 3)

demonstrated that the winnings range from 101.75 % (in the case of the algorithm of the fastest descent) to 5,805.54 % (in the case of the Adam algorithm).

Data collected during the verification of the proposed algorithmic support on the example of processing computed tomography of lung examination in order to detect tuberculosis in the presence of tuberculosis showed that the accuracy of solutions of individual convolutional classifiers without the proposed segmentation architecture (Table 4) ranges from 90.02 % (in the case of using Squeeze and Excitation-Residual Module as a classifier) to 91.74 % (in the case of using Inception-Attention Block as a classifier).

In the case of using a full convolutional majoritarian ensemble (Table 5), it amounted to 93.36 % on the test sample. While the accuracy of other ensembles (Table 8) is less. Thus, for CNN bagging, it is 91.39 %, for Bagging Decision Tree – 83.85 %, and for AdaBoost – 83.85 %.

Such a win for the proposed procedure of structural and parametric synthesis is explained by the fact that other architectural ensembles have deterministic topology, while the developed procedure makes it possible to change the topology (or rejecting unsuccessful topology) in case of failure to reach the accuracy criterion.

The advantages of the ensemble structure of neural networks are a significant improvement in the results of the solution to the set problem in comparison with any network that is part of the ensemble, on condition that the involvement of NN in the ensemble takes place according to the criteria of accuracy and diversity.

Under the proposed approach, the ensemble structure is used twice: when solving the problems of segmentation and classification, and, in the problem of classification, each NN that is part of the ensemble processes its set of features, which increases accuracy.

The disadvantage of using the ensemble structure is to increase computing costs due to the need to train more networks and implement the NN selection procedure, which are included in the ensemble. This study does not take into consideration the possibility of a small training sample and its quality. Eventually, other researchers may conduct research on the use of the transfer learning procedure in small sampling.

This study could be used in the development of new hybridization principles both in choosing the structure of NN and in setting parameters. In applied terms, this study could be used to build intelligent medical diagnostic systems to diagnose affected lung areas, namely tuberculosis. There are certain limitations for the development of such systems, namely, the effectiveness of the proposed approaches depends on the quality of the training sample, namely its length, which, in some cases, is difficult to achieve. The

situation is complicated in the case when the results of the research are not digitally translated, which, in turn, complicates the preparation of the training sample for their use in the configuration of the selected HNN architecture.

The results of our work could also be expanded as a result of the study of the properties of unique blocks and the possibilities of combining them in different configurations.

7. Conclusions

1. Unique blocks (modules) of modern convolutional networks, their functionality, and properties have been defined. This makes it possible to improve the quality of the network to detect significant features of the image by, for example, increasing the number of layers without the possibility of dropping for a local gradient. In addition, a reduction in computing costs is achieved by selecting a significant processing area. The use of such blocks in hybrid neural networks, instead of using the networks themselves, will increase the accuracy of solving the classification problem while reducing computing costs.

2. A two-step procedure for determining the structure and parameters of a hybrid neural network with the formation of its binary representation based on the use of a hybrid multicriteria genetic algorithm has been developed. Other modern gradient methods with the determination of their effectiveness were also used, which makes it possible to bypass the problem of getting into a local extremum when training the network, as well as to raise the accuracy of solving the classification problem.

3. The algorithm of formation of the ensemble of hybrid convolutional neural networks for solving the segmentation problem based on the use of the Ensemble-ADD method has been developed. As components of Ensemble-ADD, the results of such segmentators as Mask R-CNNADD, DeeplabV3+ADD, Deep Pyramid-Attention Module ADD were used. That created the final mask of segmentation, the algorithm for forming an ensemble of classifiers with the definition of its structure, criteria, the contribution of each classifier. This has made it possible to optimize the composition of the ensemble, reduce computational costs, and raise the accuracy of solving the classification problem.

4. The proposed algorithmic support was checked using the example of processing computed tomography images of lung examination in order to detect tuberculosis in the presence of tuberculosis. The test accuracy of solving the full ensemble of classifiers was 97.14 %, while the test accuracy of other ensembles was: Random Forest – 87.15 %, AdaBoost – 54.25 %, Bagging Decision Tree – 83.85 %, CNN bagging – 91.39 %.

References

1. Al-Marzouqi, H. (2009). Data Clustering Using a Modified Kuwahara Filter. 2009 International Joint Conference on Neural Networks. doi: <https://doi.org/10.1109/ijcnn.2009.5178658>
2. Chumachenko, O. I. (2017). Deep Learning Classifier Based on NEFCLASS and NEFPXOX Neural Networks. *Materialy mizhnarodnoi naukovo-praktychnoi konferentsiyi: Informatsiyi tekhnolohiyi ta kompiuterne modeliuвання*. Ivano-Frankivsk. Івано-Франківськ, 278–281. Available at: <http://itcm.comp-sc.if.ua/2017/Chumachenko.pdf>
3. Zgurovsky, M., Sineglazov, V., Chumachenko, E. (2021). Artificial Intelligence Systems Based on Hybrid Neural Networks. *Studies in Computational Intelligence*. doi: <https://doi.org/10.1007/978-3-030-48453-8>
4. Xie, L., Yuille, A. (2017). Genetic CNN. arxiv.org. Available at: <https://arxiv.org/pdf/1703.01513.pdf>

5. Real, E., Aggarwal, A., Huang, Y., Le, Q. V. (2019). Regularized Evolution for Image Classifier Architecture Search. Proceedings of the AAAI Conference on Artificial Intelligence, 33, 4780–4789. doi: <https://doi.org/10.1609/aaai.v33i01.33014780>
6. Saxena, S., Verbeek, J. (2016). Convolutional Neural Networks. Advances in Neural Information Processing Systems (NIPS). Barcelona.
7. Baker, B., Gupta, O., Naik, N., Raskar, R. (2016). Designing Neural Network Architectures using Reinforcement Learning. arXiv.org. Available at: <https://arxiv.org/pdf/1611.02167.pdf>
8. Brock, A., Lim, T., Ritchie, J. M., Weston, N. J. (2018). SmaSH: One-shot model architecture search through hypernetworks. Paper presented at 6th International Conference on Learning Representations 2018. Vancouver. Available at: <https://arxiv.org/pdf/1708.05344.pdf>
9. Zoph, B., Vasudevan, V., Shlens, J., Le, Q. V. (2018). Learning Transferable Architectures for Scalable Image Recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. doi: <https://doi.org/10.1109/cvpr.2018.00907>
10. Liu, H., Simonyan, K., Vinyals, O., Fernando, C., Kavukcuoglu, K. (2018). Hierarchical representations for efficient architecture search. arXiv.org. Available at: <https://arxiv.org/pdf/1711.00436.pdf>
11. Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., Shen, D. (2015). Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. NeuroImage, 108, 214–224. doi: <https://doi.org/10.1016/j.neuroimage.2014.12.061>
12. Yap, M. H., Pons, G., Marti, J., Ganau, S., Sentis, M., Zwiggelaar, R. et. al. (2018). Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. IEEE Journal of Biomedical and Health Informatics, 22 (4), 1218–1226. doi: <https://doi.org/10.1109/jbhi.2017.2731873>
13. Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 234–241. doi: https://doi.org/10.1007/978-3-319-24574-4_28
14. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.-A. (2017). Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. IEEE Transactions on Medical Imaging, 36 (4), 994–1004. doi: <https://doi.org/10.1109/tmi.2016.2642839>
15. Bi, L., Kim, J., Ahn, E., Kumar, A., Fulham, M., Feng, D. (2017). Dermoscopic Image Segmentation via Multistage Fully Convolutional Networks. IEEE Transactions on Biomedical Engineering, 64 (9), 2065–2074. doi: <https://doi.org/10.1109/tbme.2017.2712771>
16. Yuan, Y., Chao, M., Lo, Y.-C. (2017). Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance. IEEE Transactions on Medical Imaging, 36 (9), 1876–1886. doi: <https://doi.org/10.1109/tmi.2017.2695227>
17. Goyal, M., Yap, M., Hassanpour, S. (2020). Multi-class Semantic Segmentation of Skin Lesions via Fully Convolutional Networks. Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies. doi: <https://doi.org/10.5220/0009380302900295>
18. Vesal, S., Malakarjun Patil, S., Ravikumar, N., Maier, A. K. (2018). A Multi-task Framework for Skin Lesion Detection and Segmentation. OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, 285–293. doi: https://doi.org/10.1007/978-3-030-01201-4_31
19. Yap, M. H., Goyal, M., Ng, J., Oakley, A. (2019). Skin lesion boundary segmentation with fully automated deep extreme cut methods. Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging. doi: <https://doi.org/10.1117/12.2513015>
20. Soudani, A., Barhoumi, W. (2019). An image-based segmentation recommender using crowdsourcing and transfer learning for skin lesion extraction. Expert Systems with Applications, 118, 400–410. doi: <https://doi.org/10.1016/j.eswa.2018.10.029>
21. De Brebisson, A., Montana, G. (2015). Deep neural networks for anatomical brain segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). doi: <https://doi.org/10.1109/cvprw.2015.7301312>
22. LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation, 1 (4), 541–551. doi: <https://doi.org/10.1162/neco.1989.1.4.541>
23. Lee, H., Grosse, R., Ranganath, R., Ng, A. Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09. doi: <https://doi.org/10.1145/1553374.1553453>
24. Khan, A., Sohail, A., Zahoora, U., Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review, 53 (8), 5455–5516. doi: <https://doi.org/10.1007/s10462-020-09825-6>
25. Chumachenko, H. I., Levitsky, O. Y. (2011). Development of image processing algorithm for diagnostics. Electronics and Control Systems, 1 (27), 57–65. doi: <https://doi.org/10.18372/1990-5548.27.258>
26. Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: <https://doi.org/10.1109/cvpr.2017.195>
27. Xie, S., Girshick, R., Dollar, P., Tu, Z., He, K. (2017). Aggregated Residual Transformations for Deep Neural Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: <https://doi.org/10.1109/cvpr.2017.634>

28. Zhang, X., Li, Z., Loy, C. C., Lin, D. (2017). PolyNet: A Pursuit of Structural Diversity in Very Deep Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: <https://doi.org/10.1109/cvpr.2017.415>
29. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H. et. al. (2017). Residual Attention Network for Image Classification. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: <https://doi.org/10.1109/cvpr.2017.683>
30. Woo, S., Park, J., Lee, J.-Y., Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. Lecture Notes in Computer Science, 3–19. doi: https://doi.org/10.1007/978-3-030-01234-2_1
31. Hu, J., Shen, L., Sun, G. (2018). Squeeze-and-Excitation Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. doi: <https://doi.org/10.1109/cvpr.2018.00745>
32. Ioffe, S., Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning, 37, 448–456.
33. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: <https://doi.org/10.1109/cvpr.2016.90>
34. Simonyan, K., Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. Published as a conference paper at ICLR 2015. Available at: <https://arxiv.org/pdf/1409.1556.pdf>
35. Bodyanskiy, E. V., Rudenko, O. G. (2004). *Iskusstvennye neyronnye seti: arhitektury, obuchenie, primeneniya*. Kharkiv: Teletekh, 369.
36. Golovko, V. A. (2001). *Neyronnye seti. Obuchenie, organizatsiya i primeneniye. Kniga 4*. Moscow: IPRZHR, 256.
37. Kohonen Self-Organizing Maps. Available at: <http://mnemstudio.org/neural-networks-kohonen-self-organizing-maps.htm>
38. Martinez-Munoz, G., Hernandez-Lobato, D., Suarez, A. (2009). An Analysis of Ensemble Pruning Techniques Based on Ordered Aggregation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31 (2), 245–259. doi: <https://doi.org/10.1109/tpami.2008.78>
39. Goyal, M., Oakley, A., Bansal, P., Dancy, D., Yap, M. H. (2020). Skin Lesion Segmentation in Dermoscopic Images With Ensemble Deep Learning Methods. IEEE Access, 8, 4171–4181. doi: <https://doi.org/10.1109/access.2019.2960504>
40. Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. International Journal of Machine Learning Technology, 2 (1), 37–63.