I.J. Image, Graphics and Signal Processing, **2022, 6, 1-14**
Published Online on December 8, 2022 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2022.06.01

The Method of Semantic Image Segmentation Using Neural Networks

Ihor Tereikovskyi*
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
Email: terejkowski@ukr.net
ORCID iD: https://orcid.org/0000-0003-4621-9668
*Corresponding Author

Zhengbing Hu
School of Computer Science, Hubei University of Technology, Wuhan, China
Email: drzbhu@gmail.com
ORCID iD: https://orcid.org/0000-0002-6140-3351

Denys Chernyshev
Kyiv National University of Construction and Architecture, Kyiv, Ukraine
Email: taqm@ukr.net
ORCID iD: https://orcid.org/0000-0002-1946-9242

Liudmyla Tereikovska
Kyiv National University of Construction and Architecture, Kyiv, Ukraine
Email: tereikovskal@ukr.net
ORCID iD: https://orcid.org/0000-0002-8830-0790

Oleksandr Korystin
Scientifically Research Institute of the Ministry of Internal Affairs, Kyiv, Ukraine
Email: alex@korystin.pro
ORCID iD: https://orcid.org/0000-0001-9056-5475

Oleh Tereikovskyi
National Aviation University, Kyiv, Ukraine
Email: tereikovskyio@gmail.com
ORCID iD: https://orcid.org/0000-0001-5045-0163

Received: 12 August, 2022; Revised: 13 September, 2022; Accepted: 15 October, 2022; Published: 08 December, 2022

Abstract: Currently, the means of semantic segmentation of images, which are based on the use of neural networks, are increasingly being used in computer systems for various purposes. Despite significant progress in this industry, one of the most important unsolved problems is the task of adapting a neural network model to the conditions for selecting an object mask in an image. The features of such a task necessitate determining the type and parameters of convolutional neural networks underlying the encoder and decoder. As a result of the research, an appropriate method has been developed that allows adapting the neural network encoder and decoder to the following conditions of the segmentation problem: image size, number of color channels, acceptable minimum segmentation accuracy, acceptable maximum computational complexity of segmentation, the need to label segments, the need to select several segments, the need to select deformed , displaced and rotated objects, allowable maximum computational complexity of training a neural network model, allowable training time for a neural network model. The main stages of the method are related to the following procedures: determination of the list of image parameters to be registered; formation of training example parameters for the neural network model used for object selection; determination of the type of CNN encoder and decoder that are most effective under the conditions of the given task; formation of a representative educational sample; substantiation of the parameters that should be used to assess the accuracy of selection; calculation of the values of the design parameters of the CNN of the specified type for the encoder and decoder; assessment of the accuracy of selection and, if necessary, refinement of the architecture of the neural network model. The developed method was verified experimentally on examples of semantic segmentation of images containing objects such as a car. The obtained experimental results show that the application of the proposed method allows, avoiding complex long-term experiments,

to build a NN that, with a sufficiently short training period, ensures the achievement of image segmentation accuracy of about 0.8, which corresponds to the best systems of similar purpose. It is shown that it is advisable to correlate the ways of further research with the development of approaches to the use of special modules such as ResNet, Inception and mechanisms of the Partial convolution type used in modern types of deep neural networks to increase their computational efficiency in the encoder and decoder.

Index Terms: Semantic Segmentation Method, Convolutional Neural Network, Encoder, Decoder, Neural Network Model Efficiency, Segmentation Accuracy.

1. Introduction

In modern conditions, image recognition tools are widely distributed in computer systems for various purposes. Yes, similar tools are used for biometric authentication, determination of emotional state, environmental monitoring, detection of material defects, medical diagnostics. In most cases, the main result of recognition is the classification of objects contained in the image. For example, in biometric protection systems, the result of recognition can be the user's face, classified based on the image of the face, and in technical diagnostics systems, the determination of the technical condition of the device, classified based on the image of one or more parts. Regardless of the expected main result of recognition, one of the main stages of its implementation is the selection of one or more target objects in the input image. At the same time, in many cases, the selection of boundaries and areas of location of target objects is one of the main results of recognition. Note that, unlike the classification task, the task of selecting objects in an image involves assigning a label belonging to a certain class to each pixel of this image. In the literature [12, 17], the process of identifying boundaries and locations of target objects, which involves the use of artificial intelligence, is called semantic segmentation of images. Accordingly, it can be considered that the task of semantic segmentation is one of the varieties of the more general task of selecting target objects in images. The problem of semantic segmentation of images is complicated by the possible partial or complete overlap of target objects, the vagueness of their boundaries, the variety of sizes and placement. In addition, the variability of the recognition conditions increases due to a possible change in the source of image acquisition, which is used in most general-purpose computer systems using a video camera, which leads to a possible change in the main parameters of the raster image under test - size, resolution, and the number of color channels. Also, the negative impact on the recognition result can be caused by obstacles and the angle of the video registration. As a result of the described difficulties, means of semantic segmentation of images, based on classical methods such as "region expansion", "boundary analysis", "region fragmentation" are highly specialized and require significant modification even with minor changes in the application conditions. Thus, the solutions given in [2, 7] for tracking the image of a face in a video series are quite difficult to adapt for highlighting the contours of other objects, for example, for selecting the internal organs of a person based on the results of an ultrasound scan. The conclusion regarding the limitations of classical selection methods is confirmed by the results of the analysis of works [13, 19], which describe models for tracking images of objects using integrated filters of various types. At the same time, over the past few years, interest in neural network means of selection has grown, which is explained by the proven effectiveness of neural networks (NN) in solving such complex formalizable and multifactorial tasks.

2. Literature Review

The use of NN to select objects in images is described in scientific and practical works [3, 5, 14, 17]. The conducted analysis made it possible to build a typical structural diagram of the functioning of the neural network system of semantic segmentation of raster images shown in Fig. 1.

Also, the results of the literature analysis allow us to state that in the case of using neural network technologies, a neural network model (NM) is used, which consists of two blocks - an encoder and a decoder. The task of the encoder is to determine the multidimensional array of features of the initial image, and the task of the decoder is to obtain a processed image in which each of the pixels receives a marker of relation to one of the selected target objects or the background. That is, in such a NM, the image is first presented along the so-called narrowing path, which actually implements the selection of significant features of the image, and then along the expansion path, which ensures the labeling of image pixels. A typical schematic diagram of NM, intended for the selection of objects in the image, is shown in Fig. 2. Note that this figure illustrates the segmentation of a color ultrasound image of human lungs. At the entrance presented in Fig. 2 models provide an image in RGB format, and the output is a segmented image. In this case, the encoder and decoder are slightly modified convolutional neural networks (CNN).

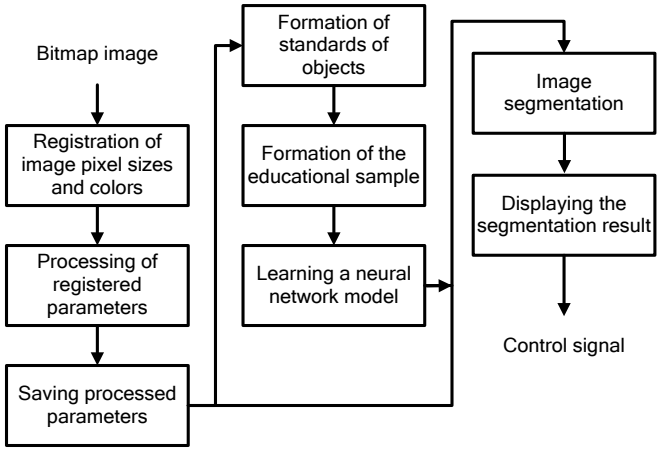

Fig. 1. A typical structural diagram of the functioning of a neural network system for the semantic segmentation of bitmap images

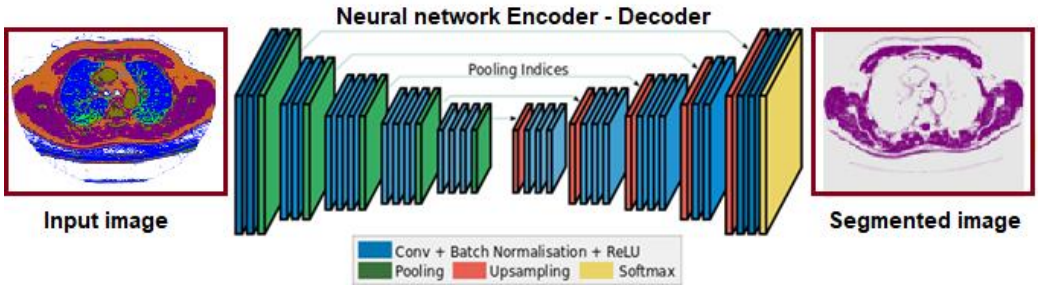

Fig. 2. A typical diagram of a neural network model designed to select objects in images.

The modification consists in the fact that not full-size CNNs are used, but only CNN blocks, which are designed to determine significant features. That is, such CNNs include only convolution and scaling layers, and fully connected layers are removed. Another feature of the NM circuit shown in Fig. 2, the symmetry of the convolution and scaling layers of the encoder and decoder appears. In the general case, such symmetry may be absent. As evidenced by the data [16, 25], the accuracy and computational resource intensity of means of selection of different types of objects in images based on neural network approaches significantly exceeds other selection technologies. At the same time, due to the novelty of this approach and the rather rapid development of neural network technologies, a number of individual problems remain unsolved today, the solution of which will allow increasing the efficiency of selection. According to the conclusions [6, 22] and the analysis results [1, 4, 18], one of these tasks is the adaptation of the NM type and parameters to the conditions of object selection in the image. In turn, the solution to this problem should be preceded by the definition and formalization of the specified selection conditions, the development of an apparatus for assessing the accuracy and computational resource intensity of the model. In addition, the criteria for determining the effectiveness of NM need to be clarified. Thus, the goal of this study is to develop an effective method of semantic segmentation of images using neural networks, which provides sufficient accuracy under variable application conditions. By analogy with well-known solutions in the field of application of neural networks for processing biometric parameters [11, 24], the development of such a method involves building a model of semantic segmentation of images, a model of a neural network encoder, and a model of a neural network decoder.

3. A Model of Semantic Image Segmentation

According to the results of [2, 10, 19], the model of semantic image segmentation based on NN can be written using an expression of the form:

$$Im_{in} \overset{Pr}{\rightarrow} Im_{pr} \overset{NC}{\rightarrow} Fm \overset{ND}{\rightarrow} Im_{out}, \quad (1)$$

Where, Im_{in} – initial raster image; $\overset{Pr}{\rightarrow}$ – image preprocessing operator; Im_{pr} – pre-processed image; $\overset{NC}{\rightarrow}$ – neural network image coding operator; Fm is a tuple of feature matrices obtained as a result of the operation of neural network coding of the image; $\overset{ND}{\rightarrow}$ – neural network image decoding operator; Im_{out} – segmented image.

In turn, Im_{in}, Im_{pr} and Im_{out} represent a matrix of the form:

$$Im_x = \begin{Vmatrix} pt_{1,1}(x) & \dots & pt_{n,1}(x) & \dots & pt_{N,1}(x) \\ \dots & \dots & \dots & \dots & \dots \\ pt_{1,m}(x) & \dots & pt_{n,m}(x) & \dots & pt_{N,m}(x) \\ \dots & \dots & \dots & \dots & \dots \\ pt_{1,M}(x) & \dots & pt_{n,M}(x) & \dots & pt_{N,M}(x) \end{Vmatrix}, \quad (2)$$

Where, $pt_{n,m}(x)$ – pixel color description with coordinates (n, m); N – horizontal image size; M – vertical image size; x is the stage of image processing, $x \in \{in, pr, out\}$.

The set of feature matrices is defined as follows:

$$Fm = \{fm_1, \dots, fm_k, \dots, fm_K\}, \quad (3)$$

Where, fm_k – the feature matrix corresponding to the k-th subsampling map in the last layer of the CNN, which is used as an encoder; K is the number of subsampling cards in the last layer of the CNN coder.

It should be noted that in the case when the last layer of the CNN encoder is the convolution layer, then in expression (2.6) fm_k corresponds to the k-th convolution map, and K is the number of convolution maps. In turn, fm_k can be written in the form:

$$fm_k = \begin{Vmatrix} \alpha_{1,1} & \dots & \alpha_{L,1} \\ \dots & \dots & \dots \\ \alpha_{J,1} & \dots & \alpha_{L,J} \end{Vmatrix}_k, \quad (4)$$

Where, $\alpha_{j,l}$ – the value of the feature at the point with coordinates *j*, *l* for the k-th subsampling/convolution map; *L* is the horizontal size of the k-th map; *J* is the vertical size of the k-th map.

At the same time, in the vast majority of well-known CNNs, maps of the same size are used in one subsampling/convolution layer. It should also be noted that expressions (3, 4) describe the output of the CNN-based encoder in the case of single-channel image processing.

Since the results of the literature analysis indicate that the prospects for increasing the effectiveness of neural network tools for semantic segmentation due to the improvement of image preprocessing modules are practically exhausted today, and the main components of such tools are the encoder and decoder, it is obvious that their improvement will ensure the possibility of increasing effectiveness of the specified means. At the same time, by analogy with [14, 17, 20], the process of improving the neural network models of the encoder and decoder at the stage of their design can be defined as follows:

$$E \to max, \quad (5)$$

$$E = \sum_{i=1}^{I} \alpha_i w_i, \;\; \alpha_i \in \{\alpha\}, w_i \in \{w\}, \quad (6)$$

Where *E* is the efficiency function of segmentation tools; *I* – the number of efficiency parameters k_i– the value of the i-th efficiency parameter; α_i – weight coefficient of the ith efficiency parameter; $\{\alpha\}$ is a set of weighting coefficients of efficiency parameters; $\{w\}$ is a set of efficiency parameters.

Based on [2, 11, 15], a list of ten performance evaluation parameters was obtained, which is given in the table 1.

Table 1. List of parameters for evaluating the effectiveness of neural network tools for semantic segmentation of images

Marking	Characteristics of the parameter
w_1	Segmentation accuracy
w_2	Computational complexity of segmentation
w_3	Ability to label image segments
w_4	The possibility of selecting several segments corresponding to different objects
w_5	The possibility of selecting several segments corresponding to objects that partially overlap each other
w_6	Ability to select deformed objects
w_7	Ability to select shifted objects
w_8	Ability to select returned objects
w_9	Computational complexity of neural network model training
w_{10}	Term of learning neural network model

The resulting list of performance parameters is preliminary and may be expanded in the future. The value of each of the efficiency parameters and the value of their weighting factors can be determined using expert evaluation methods. It should be noted that the given efficiency parameters allow you to evaluate the efficiency of the selection tool

regardless of the conditions of the task. In the basic case, their value can be estimated on a binary scale of 0 or 1. The value of the parameter is set equal to 1, if the tool provides the corresponding service. Otherwise, the parameter is considered equal to 0. It should be noted that the use of expressions (5, 6) is appropriate for the preliminary determination of the prospects for the use of neural network tools at the design stage, when the conditions of the selection problem cannot be clearly described with the help of numerical values. In the case of evaluating the effectiveness of developed experimental samples of neural network image segmentation tools, it is advisable to use recognition segmentation accuracy and computational resource intensity of neural network tools as performance indicators. This allows you to formalize the improvement process using a species expression:

$$\begin{cases} A \to max \\ \Theta \le \Delta \end{cases}, \tag{7}$$

Where A is the accuracy of segmentation, Θ is the amount of computing resources during image segmentation.

Peculiarities of the task of image segmentation make it necessary to use indicators reflecting the similarity of geometric objects to assess accuracy. According to the results of the analysis [2, 21, 25], the use of the Jacquard coefficient is provided:

$$J = \frac{|N \cap M|}{|N \cap M| + |N - M| + |B - \mathrm{N}|}, \tag{8}$$

Where N, M are areas to be compared.

Under the conditions of using one-hot coding of the expected output signal and the same size of the input and segmented image, expression (8) is detailed as follows:

$$J = \frac{\sum_{i=1}^{I} n_i m_i}{\sum_{i=1}^{I} n_i + \sum_{i=1}^{I} m_i - \sum_{i=1}^{I} (n_i - m_i)}, \tag{9}$$

Where I is the number of points describing the expected output signal of the neural network model; n_i is the value characteristic of the i-th pixel of the segmented image; m_i is the value characteristic of the i-th pixel of the expected output signal.

4. A Model of a Neural Network Coder

Data from literary sources [2, 3, 5, 14, 17] indicate that in neural network tools designed for the selection of objects on raster images, it is advisable to use a coder developed on the basis of CNN type LeNet-5, VGG, ResNet and GoogLeNet. The use of the specified types of CNNs is explained by their high efficiency, provenness and the availability of available tools for computer implementation. As shown in fig. 3, the structure of a classical neural network coder is a partially limited CNN of one of the listed types, from which the fully connected layers of neurons and the output layer have been removed.

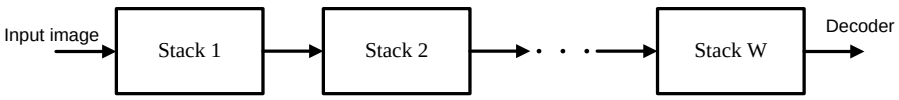

Fig. 3. Representation of the stack structure of a classical neural network encoder.

The functionality of the CNN-based encoder is described using an expression of the form

$$F_C\left(\|R\|_{H,L,K}\right) = \left\{\|\mathrm{c}_1\|_{X,Y}, \|\mathrm{c}_2\|_{X,Y}, \dots, \|\mathrm{c}_N\|_{X,Y}\right\}, \tag{10}$$

Where F_C is the input image encoding function; $\|R\|_{H,L,K}$ is a three-dimensional array of values, the elements of which are correlated with the input image to be segmented; K is the number of color channels of the input image; H, L are the dimensions of the input image.

Taking into account the LeNet-type CNN characteristics given in [4, 11] and the features of the structure of a classic neural network encoder, the expression characterizing the dependence of the accuracy indicators and computational resource capacity of the encoder on its design parameters can be written in the form:

$$\Theta, A = F(S, \|b\|, \|m\|, C, H, \|k\|), \tag{11}$$

Where S is the number of stacks; $\|b\|$ is an array containing the sizes of convolution kernels; b_w is the size of the convolution kernel for the wth stack; $\|m\|$ is an array containing scale factor values; $m_w$is scale factor value for the wth

stack; $\|k\|$ is an array containing the number of convolution maps; k_w is the number of convolution cards for the wth stack; *C* is the number of color channels to be processed; *H* is the size of the input image.

Taking into account (7, 11), an expression is obtained, which is the basis of the procedure for determining the constructive parameters of a neural network coder based on a CNN of the LeNet type:

$$\begin{cases} A(S, \|b\|, \|m\|, C, H, \|k\|) \rightarrow max \\ \Theta(S, \|b\|, \|m\|, C, H, \|k\|) \leq \Delta_\theta \end{cases}, \quad (12)$$

Note that when evaluating the encoder efficiency using expression (12), the service capabilities of NM are not taken into account. At the same time, it is possible to determine the value of the structural parameters used in expression (12), based on the principles given in [11, 22]:

1. The number of stacks should be equated to the number of levels of recognition by the expert of target objects in the image to be analyzed;
2. The number of convolutional maps in a certain convolutional layer is set equal to the number of significant features recognized at the same level by an expert.

5. A Model of a Neural Network Decoder

The functioning of the decoder, which consists in restoring the segmented image, can be described using an expression of the form:

$$F_D(\{\|c_1\|_{X,Y}, \|c_2\|_{X,Y}, \ldots, \|c_N\|_{X,Y}\}) = \|r\|_{h,l,k}, \quad (13)$$

Where F_D is the segmented image decoding function; $\|c_n\|_{X,Y}$ is an array of values corresponding to the nth feature map of the last layer of the neural network encoder; N is the number of feature maps in the last layer of the neural network coder; *X*,*Y* – the size of the feature map in the last layer of the neural network coder; $\|r\|_{h,l,k}$ is a three-dimensional array of values corresponding to the original image; *k* is the number of color channels of the original image; *h*, *l* are the dimensions of the original image.

According to the results [2, 3, 17], the development of a neural network decoder model intended for use in semantic image segmentation tools can be based on:

1. An approach based on one-stage resampling of the original image.
2. An approach based on multi-stage resampling of the original image.
3. An approach based on multi-stage symmetric resampling of the source image with integration of symmetric sets of weighting coefficients.

When building a decoder based on the first approach, the peculiarities of its functioning are that the multidimensional array of features obtained as a result of encoding the input image is transformed into an array designed to describe an image with a selected object using classic image scaling procedures. One of these procedures is the bilinear interpolation procedure [1, 10, 19]. An illustration of the application of a decoder based on one-stage resampling for semantic segmentation of medical images is shown in Fig. 4.

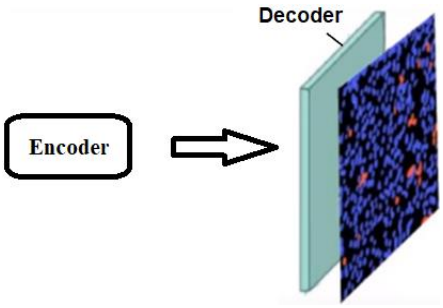

Fig. 4. Illustration of the application of a decoder based on one-stage resampling.

The approach based on the multi-stage resampling of the original image consists in the fact that the multidimensional array of features obtained at the output is transformed into an array designed to describe the image with a selected object by CNN using CNN feature extraction modules. It should be noted that, in general, the type and parameters of the ANN used to build the decoder may not match the type and parameters of the CNN used to build the encoder. Thus, the NM used to extract an object in an image when using a decoder based on multi-stage resampling can

be asymmetric. An illustration of the operation of a decoder built on the basis of multi-stage resampling is shown in Fig. 5.

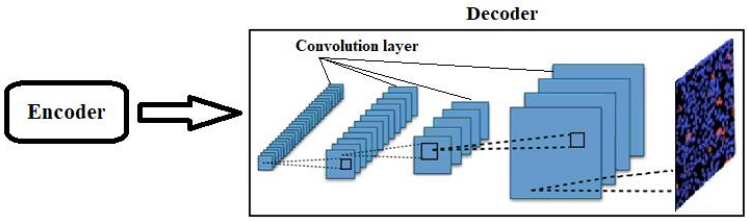

Fig. 5. Illustration of the application of a decoder based on multi-stage resampling.

The approach based on multi-stage symmetric resampling assumes that the same CNN with a mirrored structure is used to develop the encoder and decoder. That is, the structural CNNs of the encoder in reverse order correspond to the structure of the CNNs of the decoder, however, scaling layers are not used in the decoder. At the same time, some convolution layers of the encoder and decoder are interconnected. For example, in Fig. 6 shows NM for semantic image segmentation with a decoder based on multi-stage symmetric resampling. Transformation of the information shown in fig. 6 models are implemented in two stages. At the first stage, the encoder compresses the image to a set of features defined in layer C14. At the second stage, with the help of the decoder, the extension of the set of features contained in layer C14 to the processed image with the selected target object is implemented. Structurally, the encoder consists of 4 stacks, each of which includes two layers of convolution maps with a 3×3 kernel size and one subsampling layer, which reduces the image size by a factor of 2.

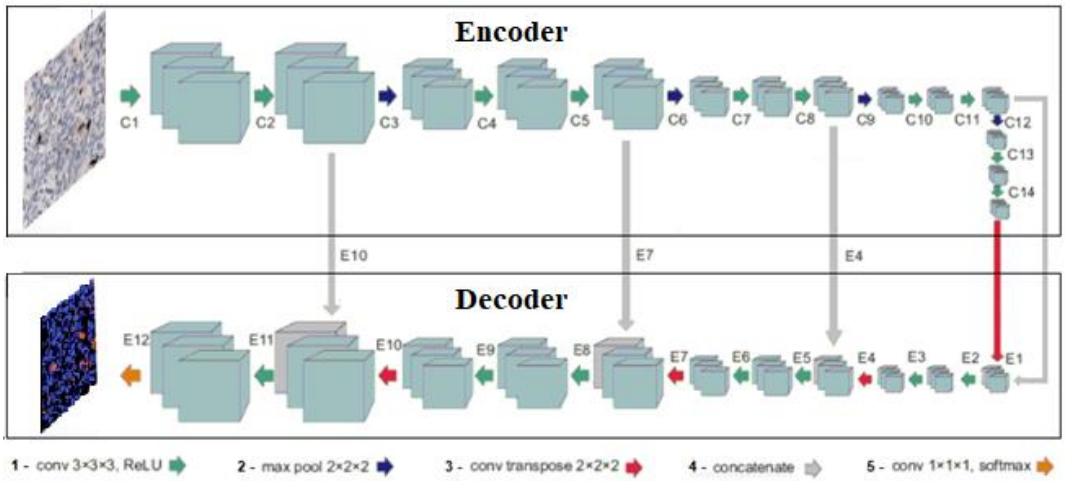

Fig. 6. Illustration of the application of a decoder based on multi-stage symmetric resampling.

Each subsequent stack contains twice as many trait cards as the previous one. The structure of the decoder, which also consists of 4 stacks, has certain differences. Each of the stacks includes a resampling layer, a convolution layer with a 2×2 kernel, an integration block with a corresponding encoder convolution map, and two convolution layers with a 3×3 kernel. The last layer uses a 1×1 convolution kernel to map the 64-dimensional feature vector to the pixels of the processed source image. The blocks included in the structure shown in fig. 6 of the decoder, have the Ex type designation, where x is the block number.

Like the encoder, the effectiveness of the decoder directly depends on the type of CNN on which it is based. In addition, both the accuracy and the computational resource intensity of the decoder largely depend on the approach on the basis of which it is built. Thus, a variation of the decoder construction based on one of the three approaches given and the possibility of using different types of CNN in it potentially allows the decoder to be adapted to the conditions of the task of object selection. The corresponding adaptation procedure can be described using an expression of the form:

$$\begin{cases} A(S, \|b\|, \|m\|, C, H, \|k\|, T) \rightarrow max \\ \Theta(S, \|b\|, \|m\|, C, H, \|k\|, T) \leq \Delta_{\theta} \end{cases}, \quad (14)$$

Where T is the type of decoder construction approach.

When constructing a decoder based on multistage resampling or based on multistage symmetric resampling, it is possible to determine the approximate values of the design parameters used in expression (14), based on the principles given in the description of the encoder model.

6. Method of Semantic Image Segmentation

According to the results of the analysis of the neural network object selection technologies and the developed models of the encoder and decoder, to describe the features of the method of semantic segmentation of images, it is recognized as appropriate to use the expression of the species

$$\langle\{u_{req}\},\{u_{con}\},\{NN_d\},\{u_{CNN}\},\{d\},\{\alpha\},\{w\},\rangle \rightarrow \langle CNN_{type}^{enc}, CNN_{type}^{dec},\{CNN^{enc}\},\{CNN^{dec}\}\rangle, \quad (15)$$

Where $\{u_{req}\}$ is a set of registration parameters; $\{u_{con}\}$ - set of requirements for recognition results; $\{u_{obj}\}$ - set containing a description of selected objects; $\{NN_d\}$ - set of available CNN types; $\{u_{cnn}\}$ - set of parameters of available CNN types; $\{d\}$ is a set of expert data that can be used to build a coder and decoder model; $\{\alpha\}$ - a set of efficiency parameter coefficients used in expression (6); $\{w\}$ is a set of efficiency parameters used in expression (6) and listed in the table. 1; $CNN_{type}^{enc}, CNN_{type}^{dec}$ - type of CNN used as a basis for building the encoder and decoder, respectively; $\{CNN^{enc}\},\{CNN^{dec}\}$ - parameters of CNN type $CNN_{type}^{enc}, CNN_{type}^{dec}$, respectively.

Note that $\{u_{req}\}$ largely depends on the requirements for registration format (g_{rf}), resolution (g_{rs}) and input image size (g_{sz}). In addition, based on the results [16, 21, 23], it was determined the need to apply in the method of semantic image segmentation stages aimed at solving tasks related to the implementation of the training sample formation procedure and the selection accuracy assessment procedure. Thus, the list of main tasks that should be associated with the stages of the semantic image segmentation method consists of:

- Definition of the list of image parameters to be registered.
- Formation of training example parameters for NM used for feature extraction.
- Determination of the type of CNN that is the most effective in the conditions of the task of selection.
- Formation of the training sample, the volume of which should be sufficient under the conditions of the task.
- Justification of the parameters to be used to assess the accuracy of selection.
- Calculation of the values of the design parameters of a CNN of a certain type.
- Evaluation of the accuracy of selection and, if necessary, refinement of the NM architecture.

The generalized scheme of the method, built taking into account the need to use an iterative approach when solving the formulated tasks, is shown in Fig. 7. We will give a detailed description of the stages of the method.

Stage 1. Determination of the list of registration parameters. At the input of the stage, which is the centralized input of this method, a tuple of values presented in the left part of expression (15) is supplied. At the first stage, the components of the set $\{X_{NN}\}$ containing the registration parameters of the input image are determined. The determination is based on the comparison of the array $\|Im_{in}\|$ with the requirements of g_{rf}, g_{rs}, g_{sz}. The output of the first stage is $\{X_{NN}\}$.

Stage 2. Determination of the effective type of CNN. The stage is focused on determining the types of CNNs that are the most effective when building an encoder and when building a decoder, which are included in the neural network model of object selection. The input of the stage is $\{X_{NN}\}$, a set of available CNN types ($\{NN_d\}$) and the list of parameters for evaluating the effectiveness of means of selecting objects on raster images, which is given in the table. 1. The execution of the stage is divided into five steps.

Step 2.1. Determination of values of performance evaluation parameters. At this step, for each type of CNN included in the set $\{NN_d\}$, the values of the components $\{w\}$ are determined. The definition is implemented using expert evaluation methods taking into account $\{X_{NN}\}$. In the first approximation, the processing of expert data regarding the determination of the w_i value can be implemented using an expression of the type:

$$w_i = \frac{1}{Q}\sum_{q=1}^{Q} w_{i,q}, \quad (16)$$

Where $w_{i,q}$ is the value of the i-th efficiency parameter set by the q-th expert; Q is the number of experts.

The output of the step is the matrix $\|w\|$ containing the values of efficiency parameters for each component $\{NN_d\}$.

Step 2.2. Determination of the values of weight coefficients of efficiency parameters. The definition is implemented separately for the encoder and decoder using expert evaluation methods using expression (16). The output of the step is the matrices $\|\alpha_{enc}\|$ and $\|\alpha_{dec}\|$, which contain the values of the weighting coefficients of the efficiency evaluation parameters for the encoder and decoder, respectively.

Step 2.3. Calculation of efficiency function values. The calculation is carried out for each component $\{NN_d\}$ separately for the encoder and decoder. Expression (6) is used. The output of the step is the sets $\{E_{enc}\}$ and $\{E_{dec}\}$ containing the values of the efficiency function for the encoder and decoder. At the same time, the i-th element contains the value of the efficiency function for the i-th type of CNN, which is included in the set $\{NN_d\}$.

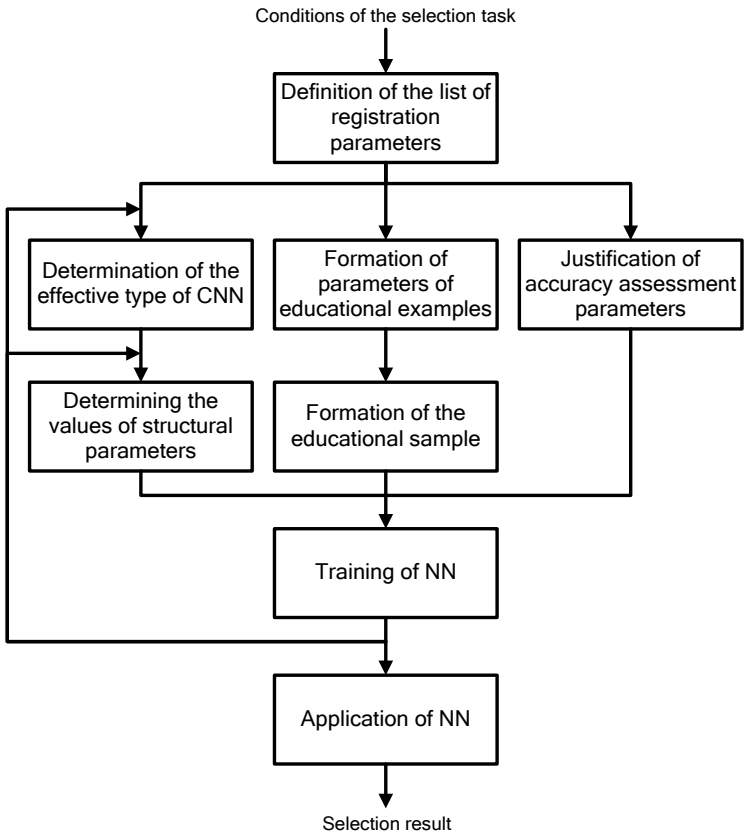

Fig. 7. Generalized scheme of the method of semantic segmentation of images using neural networks.

Step 2.4. Determination of the most efficient type of CNN encoder. Expression (5) is used for this purpose. In the first approximation, the limiting efficiency indicator may not be taken into account. There is a possible case when for several types of CNN coder that meet the condition (5), the values of the efficiency function are the same. Therefore, the output of the step is a set containing the most efficient types of CNN coder $\{N_{main}^{enc}\}$.

Step 2.5. Determination of the most efficient type of CNN decoder. The execution of this step is similar to the execution of step 2.4. The output of the step is the set containing the most efficient types of CNN decoder $\{N_{main}^{dec}\}$.

The output of the stage is $\{N_{main}^{enc}\}$ and $\{N_{main}^{dec}\}$.

Stage 3. Formation of the parameters of the training example. $\{X_{NN}\}$, g_{rf}, g_{rs} and g_{sz} are applied to the input of the stage. In the general case, the stage is focused on forming a $\langle\{X_{NN}\},\{Y_{NN}\}\rangle$ containing a list of input and input parameters of a separate training example. During the implementation of the stage, the use of proven image processing methods is provided, taking into account possible changes in the size, color and resolution of the input image. It is assumed that the expected output signal of the decoder is formed using the one-hot coding procedure. The output is $\langle\{X_{NN}\},\{Y_{NN}\}\rangle$.

Stage 4. Justification of accuracy assessment parameters. The implementation of this stage is related to the definition of the list of parameters that should be used to assess the accuracy of the selection of objects in the conditions of the given task. $\langle\{X_{NN}\},\{Y_{NN}\}\rangle$ and a list of accuracy assessment parameters are given to the input of the stage. In the basic version, the Jacquard coefficient given by expression (9) is used as a segmentation accuracy assessment parameter.

Stage 5. Determination of the values of design parameters. $\{N_{main}^{enc}\}$, $\{N_{main}^{dec}\}$, $\langle\{X_{NN}\},\{Y_{NN}\}\rangle$ and the list of design parameters of CNN defined in expression (11) are given to the stage input.

Step 5.1. Determination of constructive parameters of the coder. The execution of the step consists in delineating the range of values of the constructive parameters of the encoder, which, according to preliminary estimates, ensure the fulfillment of condition (12). At the same time, solutions obtained in the process of developing a model of a neural network coder are used.

Step 5.2. Calculation of design parameters of the decoder. The execution of the step is similar to the previous one, taking into account the condition (14).

The output of the stage is $\{P_{cod}\}$ and $\{P_{dec}\}$ – sets containing a defined range of values of the design parameters of the encoder and decoder.

Stage 6. Formation of the training sample. $\{X_{NN}\}$, $\{Y_{NN}\}$ and an array of images $\{Im\}$ are given to the input of the stage. It is assumed that the training sample can be formed with the help of expert data and with the use of available databases. Taking into account practical experience, it was determined that the minimum volume of the training sample

should be at least 1000 training examples. If necessary, it is possible to supplement the educational sample by augmenting educational examples. The output of the stage is a set of training examples ($\{\psi\}_N$).

Stage 7. Learning NN. The defined architectural parameters of NN ($\{N_{main}^{enc}\}$,$\{N_{main}^{dec}\}$, $\{P_{cod}\}$, $\{P_{dec}\}$), examples of the training sample ($\{\psi\}_N$), a set of possible parameters for estimating the selection accuracy ($\{\Delta\}$) and the required level of selection accuracy on each type of training sample ($\{\Delta_d\}$). Stage 7 is divided into three steps.

Step 7.1. Implementation of NN training. It is assumed that the training takes place on the basis of the algorithm of backpropagation of the error, taking into account the possible change in the way of providing training examples and the change in the method of optimizing the learning speed. The result of learning is a set of weight coefficients of synaptic connections ($\{W\}$).

Step 7.2. Calculation of selection accuracy indicators. The indicated indicators are calculated on the training, test and validation samples. The result of the step is the set $\{\Delta_m\}$.

Step 7.3. Assessment of selection accuracy. The indicators of selection accuracy obtained in the previous step are compared with the corresponding indicators from $\{\Delta_d\}$.

If the accuracy is insufficient, it is assumed that the type and design parameters of the NN are specified, which is implemented when performing the second and fifth stages of this method. If the selection accuracy is satisfactory, then the transition to the eighth stage of the method takes place. The output of the stage is a set of refined NN architectural parameters and parameter values that allow to assess the accuracy of object selection.

Stage 8. Application of NN. At the input of the stage, the architectural parameters of the trained NN, determined as a result of the implementation of the seventh stage of this method, and the image to be analyzed are submitted. As a result of the implementation of the stage, an output signal is formed, consisting of a marked image and parameter values of the selected object (if necessary). Together with the output data of the seventh stage, these parameters are the result of the execution of this method.

7. Experimental Studies

In order to verify the proposed method, computer experiments were carried out, which were implemented with the help of specially developed software, which was based on the proposed encoder and decoder models. The software is written in the Python programming language using the TensorFlow library. In the process of computer experiments, the effectiveness of the proposed method for semantic segmentation of color images was investigated. Masks of the car type object are provided on these images. The Carvana Image Masking Challenge database, which is freely available at the link https://www.kaggle.com/c/carvana-image-masking-challenge, was used as a source for forming the training sample of the neural network. This database contains 5088 high-contrast images recorded in RGB format. The size of a single image is 256×256 pixels. In accordance with the features of recording labeled images in the Carvana Image Masking Challenge database, the selection task is classified as a semantic image segmentation task class. The main results of other stages of the proposed method of semantic segmentation are as follows. Taking into account the size and color format of the original and segmented images, the expediency of submitting the unprocessed original image of the plexus of nerve fibers to the input of the encoder is determined. Accordingly, the set of registered parameters corresponds to a three-channel image with a size of 256x256 pixels. When determining the most effective type of CNN, taking into account the capabilities of the instrumentation, it was determined that it is appropriate to include CNNs of the LeNet, VGG, AlexNet, GoogLeNet, and ResNet types in $\{NN_d\}$. The value of the efficiency function for these types of NN when they are used as the encoder and decoder base is given in table. 2. The possibility of building a decoder based on single-stage, multi-stage and multi-stage symmetric resampling is considered. In the table 2 column E_{dec1} corresponds to multi-stage resampling, and column E_{dec2} corresponds to multi-stage symmetric resampling. For one-stage resampling, E_{dec}=0,75.

Table 2. The value of the efficiency function for admissible types of CNN

CNN type	E_{enc}	E_{dec1}	Ed_{ec2}
LeNet	0,55	0,65	0,7
VGG	0,75	0,7	0,75
AlexNet	0,5	0,4	0,5
GoogLeNet	0,6	0,5	0,55
ResNet	0,6	0,5	0,55

The value of the efficiency function for each of the specified types is determined by expert evaluation taking into account expressions (6, 16). Using expression (5), it is determined that the most efficient encoder is based on the VGG-type ZNM, and the most efficient type of decoder is the decoder based on one-stage resampling.

The input and output parameters of the training examples of the neural network model correspond to the original and segmented image shown in Fig. 11. Design parameters of VGG were adapted to the analysis of halftone images of size 256x256 by increasing the input field of CNN. The training sample is formed on the basis of the described database of images and is divided into training, validation and test samples. The training sample of NN is supplemented by the

augmentation of training examples. Of these, 8,000 examples were used as training data, and 1,088 examples were used for validation and testing. According to the recommendations [11, 21], the volume of the training sample is 8,000, and the total volume of the validation and test sample is 1,680 examples. The Jacquard coefficient was used to assess the accuracy of the NN. Graphs of the accuracy of the selection of the car mask on the raster image, implemented using the constructed NN, are shown in Fig. 8.

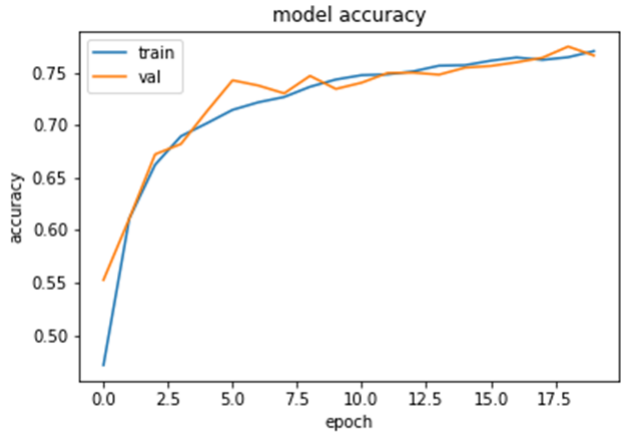

Fig. 8. Graphs of segmentation accuracy using NN with an encoder based on VGG and a decoder based on one-stage resampling.

For comparison, experiments related to the selection of a car mask using NN with other architectural parameters were carried out. Thus, in fig. 9 and Fig. 10 shows graphs of object selection accuracy using NN, in which encoder and decoder blocks are built on the basis of VGG. At the same time, the accuracy graphs shown in fig. 9, correspond to a decoder with multi-stage resampling (see Fig. 5), and the accuracy graphs shown in Fig. 10 – decoder with multi-stage symmetrical resampling (see Fig. 6).

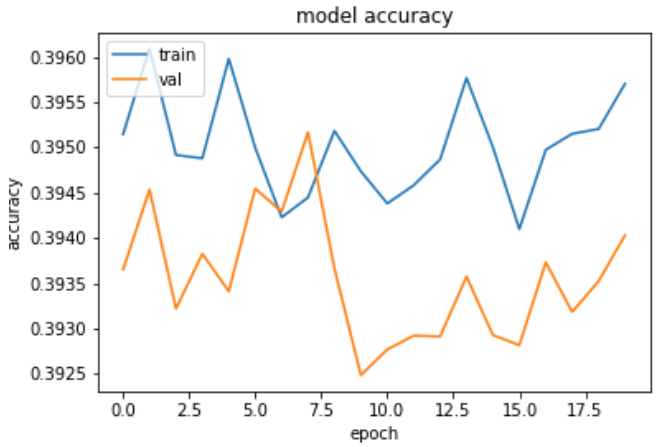

Fig. 9. Graphs of segmentation accuracy using NNM with VGG-based encoder and multi-stage resampling decoder.

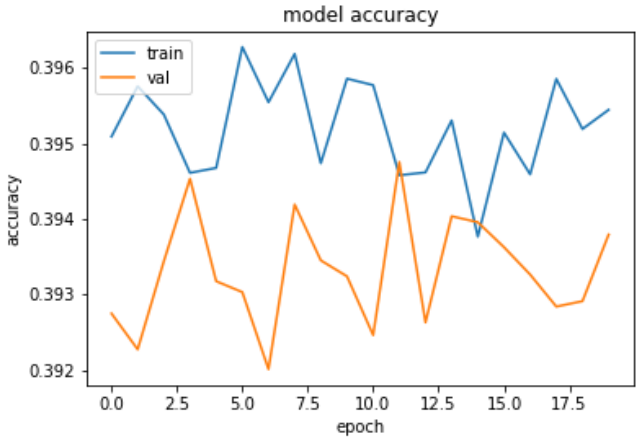

Fig. 10. Graphs of segmentation accuracy using NNM with VGG-based encoder and multi-stage symmetric resampling decoder.

Analysis of the graphs shown in fig. 8-10, allows us to state that when using the proposed NM, the segmentation accuracy is approximately 0.8, which, according to the results of the experiments, is more than 2 times higher than the selection accuracy that can be achieved using other NNs. Further improvement of accuracy, which can be realized due to modification of VGG parameters, requires additional theoretical research. Also it is possible to increase the accuracy of selection due to the use of the most modern types of CNN in the construction of the encoder and decoder. To illustrate the intended use of the developed system, an object (car) was selected on the images that were not used in the NN training process. The result of such selection is shown in fig. 11. In fig. 11, the original image is shown on the left side, and the area of the location of the object (car) is circled on the right side. Analysis of fig. 11 shows a satisfactory selection result.

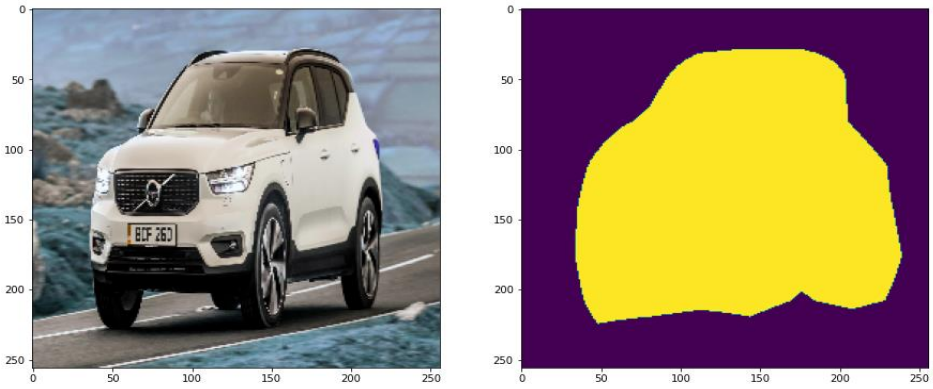

Fig. 11. An example of semantic segmentation.

Thus, the results of the experimental experiments show that the application of the proposed method allows you to build a NN, which, with a sufficiently short training period, ensures the achievement of image segmentation accuracy according to the Accuracy indicator of about 0.8. At the same time, it was possible to avoid conducting complex experiments aimed at determining the effective architecture of NN, which makes it possible to minimize the amount of computing resources aimed at building NN. Thus, the proposed method allows, while reducing the computational resources associated with the construction of NN, to achieve segmentation accuracy that is comparable to the accuracy of the best known systems of a similar purpose [2, 14, 17].

8. Conclusion and Future Work

A. Conclusion

Based on the literary analysis of modern works in the field of semantic segmentation of images, it is shown that the most promising direction of their improvement is the use of neural network technologies, the effectiveness of which is largely determined by the level of their adaptability to the conditions of the task. At the same time, researchers currently pay little attention to the issue of theoretical approaches to the adaptation of neural network model parameters to the most significant conditions of the task of semantic segmentation of images, which entails the need to conduct appropriate experiments. As a result of the conducted research, a method of semantic segmentation of images using CNN was developed, which involves the adaptation of the neural network encoder and neural network decoder to the image size, the number of color channels, the permissible minimum segmentation accuracy, the permissible maximum computational complexity of the implementation of the segmentation process, the need to label image segments, the need to select several segments corresponding to different objects, the need to select several segments corresponding to objects that partially overlap each other, the need to select deformed objects, the need to select shifted objects, the need to select rotated objects, the permissible maximum computational the complexity of learning a neural network model, the permissible term of learning a neural network model. The developed method was verified experimentally on examples of semantic segmentation of images containing objects such as a car. The obtained experimental results show that the application of the proposed method allows to avoid complicated long-term experiments to build a NN, which, with a sufficiently short training period, ensures the achievement of image segmentation accuracy of about 0.8, which corresponds to the best systems of a similar purpose. Summing up, it can be said that this article makes a significant contribution to the development of neural network systems for semantic segmentation of raster images and can serve as a basis for other researchers in this area.

B. Future Work

In the studies, the results of which are presented in this article, the main focus was on determining the most efficient type of CNN for the encoder and decoder. An approach is also considered for determining the parameters of such a CNN for the case when this network consists exclusively of convolutional layers and subsample layers. At the same time, modern proven CNNs include special modules and mechanisms that improve their efficiency in solving

many problems [8, 9, 23]. For example, the introduction of the ResNet module made it possible to increase the CNN depth by leveling the gradient drop effect, the introduction of the Inception module allows us to reduce the number of weight coefficients and process objects of different sizes, and the use of the partial convolution mechanism allows us to obtain additional information about the input image. It should be noted that the use of these modules and mechanisms is not always advisable. For example, it can be assumed that when extracting an object mask on a small grayscale image, there is no need to use an encoder and decoder based on a convolutional neural network with a large number of hidden layers. Accordingly, there is no need to use modules and mechanisms designed to level the effect of the gradient drop. At the same time, there is such a need for semantic segmentation of large color images. Thus, the ways of further research can be correlated with the use in the encoder and decoder of modern modules and mechanisms adapted to the significant conditions of the problem of semantic segmentation.

References

[1] Abraham J., Paul V. "An imperceptible spatial domain color image watermarking scheme". Journal of King Saud University – Computer and Information Sciences. 2019. Vol. 31 (1), pp. 125-133.

[2] Adithya U., Nagaraju C., "Object Motion Direction Detection and Tracking for Automatic Video Surveillance", International Journal of Education and Management Engineering (IJEME), Vol.11, No.2, pp. 32-39, 2021. DOI: 10.5815/ijeme.2021.02.04.

[3] Badrinarayanan V., Kendall A., Cipolla R. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". URL: http://arxiv.org/abs/1511.0051 (accessed October 12, 2022).

[4] Cun Y. Le, et al. "Learning Hierarchical Features for Scene Labeling". URL: http://yann.lecun.com/exdb/publis/pdf/farabet-pami-13.pdf (accessed October 11, 2022).

[5] Dmitry A. "Segmentation Object Strategy on Digital Image". Journal of Siberian Federal University. Engineering & Technologies. 2018. № 11(2), pp. 213-220.

[6] Dychka I., Chernyshev D., Tereikovskyi I., Tereikovska L., Pogorelov V. "Malware Detection Using Artificial Neural Networks". Advances in Intelligent Systems and Computing, 2020. Vol. 938, pp. 3-12.

[7] Cherrat, Rachid Alaoui, Hassane Bouzahir. "Score Fusion of Finger Vein and Face for Human Recognition Based on Convolutional Neural Network Model". International Journal of Computing, 2020. 19(1), pp. 11-19.

[8] Forrest N. Iandola, et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size". ArXiv1602.07360 Cs. (2016). http://arxiv.org/abs/ 1602.07360 (accessed September 17, 2022).

[9] Andrew Howard, et al. "Searching for MobileNetV3". ArXiv1905.02244 Cs. (2019). http://arxiv.org/abs/1905.02244 (accessed October 20, 2022).

[10] Hu Z., Tereikovskyi I., Chernyshev D., Tereikovska L., Tereikovskyi O., Wang D. "Procedure for Processing Biometric Parameters Based on Wavelet Transformations". International Journal of Modern Education and Computer Science. 2021. Vol. 13, No 2, pp. 11-22.

[11] Hu Z., Tereykovskiy I., Zorin Y., Tereykovska L., Zhibek A. Optimization of convolutional neural network structure for biometric authentication by face geometry. Advances in Intelligent Systems and Computing. 2019. Vol. 754, pp. 567-577.

[12] Jun Shen. "Motion detection in color image sequence and shadow elimination". Visual Communications and Image Processing. 2014. Vol. 5308, pp. 731-740.

[13] Kong T., et al. "FoveaBox: Beyound Anchor-Based Object Detection", IEEE Trans. Image Process. 29 (2020), pp. 7389–7398.

[14] Liu, X.-P., Li, G., Liu, L., Wang, Z. "Improved YOLOV3 target recognition algorithm based on adaptive eged optimization". Microelectron. Comput. 2019. Vol. 36, pp. 59–64.

[15] Prilianti, K. R., Anam, S., Brotosudarmo, T. H. P., Suryanto, A. "Non-destructive Photosynthetic Pigments Prediction using Multispectral Imagery and 2D-CNN". International Journal of Computing. 2021. 20(3), pp. 391-399.

[16] Reja, S. A., Rahman, M. M. "Sports Recognition using Convolutional Neural Network with Optimization Techniques from Images and Live Streams". International Journal of Computing, 2021. 20(2), pp. 276-285.

[17] Ronneberger O., Fischer P., Brox T. "U-Net: Convolutional Networks for Biomedical Image Segmentation". Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015. Vol.9351, pp. 234-241.

[18] Senocak A. et al. "Part-based player identification using deep convolutional representation and multi-scale pooling". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 1732-1739.

[19] Shkurat O. et al. "Image Segmentation Method Based on Statistical Parameters of Homogeneous Data Set". Advances in Intelligent Systems and Computing. 2020. Vol. 902, pp. 271–281.

[20] Simonyan K., Zisserman A. "Very deep convolutional networks for large-scale image recognition". ArXiv1409.1556 Cs. (2019). http://arxiv.org/abs/1409.1556 (accessed October 11, 2022).

[21] Taqi A., Awad A., Al-Azzo F., Milanova M. "The impact of multi-optimizers and data augmentation on TensorFlow convolutional neural network performance". Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). 2018, pp. 140-145.

[22] Toliupa S., Tereikovskiy I., Dychka I., Tereikovska L., Trush A. "The Method of Using Production Rules in Neural Network Recognition of Emotions by Facial Geometry". 3rd International Conference on Advanced Information and Communications Technologies. 2019, pp. 323-327.

[23] Wu C., Wen W., Afzal T., Zhang Y., Chen Y. "A compact DNN: Approaching GoogLeNet-Level accuracy of classification and domain adaptation". In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

[24] Yudin O., Toliupa S., Korchenko O., Tereikovska L., Tereikovskyi I., Tereikovskyi O. "Determination of Signs of Information and Psychological Influence in the Tone of Sound Sequences". IEEE 2nd International Conference on Advanced Trends in Information Theory. 2020, pp. 276-280.

[25] Zhang S. et al. "Single-Shot Refinement Neural Network for Object Detection". ArXiv 1711.06897 Cs. (2018). http://arxiv.org/abs/ 1711.06897 (accessed October 16, 2022).

Authors' Profiles

Ihor Tereikovskyi graduated from National Aviation University, Kyiv, Ukraine. Currently, he is a Doctor of Science, professor at Faculty of Applied Mathematics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine. He has currently published more than 200 publications. His research interests are information security, neural network systems for cyber-attacks recognition, recognition of voice signals.

Zhengbing Hu: Prof., Deputy Director, International Center of Informatics and Computer Science, Faculty of Applied Mathematics, National Technical University of Ukraine "Kyiv Polytechnic Institute", Ukraine. Adjunct Professor, School of Computer Science, Hubei University of Technology, China. Visiting Prof., DSc Candidate in National Aviation University (Ukraine) from 2019. Major research interests: Computer Science and Technology Applications, Artificial Intelligence, Network Security, Communications, Data Processing, Cloud Computing, Education Technology.

Denys Chernyshev is Doctor of Technical Sciences, First vice-rector of Kyiv National University of Construction and Architecture, Kyiv, Ukraine. He has currently published more than 200 publications. His research interests are information security, the construction of distance education systems.

Liudmyla Tereikovska graduated from State Academy of Light Industry of Ukraine, Kyiv. Currently, she is a PhD, associate professor at Kyiv National University of Construction and Architecture, Ukraine. She has currently published more than 100 publications. Her research interests are data mining, development of neural network systems, recognition of voice signals, the construction of distance education systems, recognition of cyber-attacks.

Oleksandr Korystin: DSc, PhD, Professor. In 2009 he received DSc degree in information law from NAIA. In 2014 he received Professor degree. Honored Academic of Science and Technology of Ukraine. Chief Research Scientist of the Criminological Research Laboratory of the State Scientific Research Institute of the Ministry of Internal Affairs of Ukraine. In 2014–2016 – Rector of the Odesa State University of Internal Affairs. Member of the Expert Council of the Ministry of Education and Science of Ukraine on legal sciences. Research interests: criminology; economic security; cybersecurity; intelligence; methodology of strategical (SWOT-analysis; risks assessment); counteraction to the hybrid threat.

Oleh Tereikovskyi is postgraduate student of the National Aviation University, Kyiv, Ukraine. He has currently published more than 20 publications. His research interests are information security, neural network biometric authentication systems.

How to cite this paper: Ihor Tereikovskyi, Zhengbing Hu, Denys Chernyshev, Liudmyla Tereikovska, Oleksandr Korystin, Oleh Tereikovskyi, "The Method of Semantic Image Segmentation Using Neural Networks", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.14, No.6, pp. 1-14, 2022. DOI:10.5815/ijigsp.2022.06.01