jet INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2023, VOL. 69, NO. 1, PP. 41-46
Manuscript received December 2, 2022; revised January, 2023. DOI: 10.24425/ijet.2023.144329

A Neural Network Model for Object Mask Detection in Medical Images

I. Tereikovskyi, O. Korchenko, S. Bushuyev, O. Tereikovskyi, R. Ziubina, and O. Veselska

***Abstract*—In modern conditions in the field of medicine, raster image analysis systems are becoming more widespread, which allow automating the process of establishing a diagnosis based on the results of instrumental monitoring of a patient. One of the most important stages of such an analysis is the detection of the mask of the object to be recognized on the image. It is shown that under the conditions of a multivariate and multifactorial task of analyzing medical images, the most promising are neural network tools for extracting masks. It has also been determined that the known detection tools are highly specialized and not sufficiently adapted to the variability of the conditions of use, which necessitates the construction of an effective neural network model adapted to the definition of a mask on medical images. An approach is proposed to determine the most effective type of neural network model, which provides for expert evaluation of the effectiveness of acceptable types of models and conducting computer experiments to make a final decision. It is shown that to evaluate the effectiveness of a neural network model, it is possible to use the Intersection over Union and Dice Loss metrics. The proposed solutions were verified by isolating the brachial plexus of nerve fibers on grayscale images presented in the public Ultrasound Nerve Segmentation database. The expediency of using neural network models U-Net, YOLOv4 and PSPNet was determined by expert evaluation, and with the help of computer experiments, it was proved that U-Net is the most effective in terms of Intersection over Union and Dice Loss, which provides a detection accuracy of about 0.89. Also, the analysis of the results of the experiments showed the need to improve the mathematical apparatus, which is used to calculate the mask detection indicators.**

***Keywords*—model; neural network; object mask; medical images**

I. Introduction

CURRENTLY, raster image recognition systems are widely used in various fields of science and technology. The field of medical diagnostics is no exception. The capabilities of medical diagnostic devices and hardware-computer systems lead to a wide range of types of images. A range of processing and analysis methods are used for their interpretation, transformation and research. At the same time, one of the main tasks of processing the obtained medical images is the allocation of a mask of the target object on it, which is subject to further research. With a significant amount of received medical images, the complexity of their operational analysis necessitates the use of appropriate automation tools. At the same time, known extraction systems are highly specialized and insufficiently adapted to the variability of application conditions, which leads to the need to develop new solutions in this area.

II. Analysis of Literary Works in the Field of Bitmap Image Segmentation

Work [9] is devoted to the development of a low-resolution image processing sequence for the selection of objects specified by several parameters. Filtering of the halftone image is implemented using a low-pass Gaussian filter, and threshold division of the intensity histogram is used for shadow levelling. The actual selection of the boundaries of the object was carried out using the Kenny algorithm. It should be noted that the effectiveness of the presented processing method, and the main effectiveness of the stages of boundary selection and search for working objects, cause certain doubts, which are primarily related to the description of experiments on searching for sufficiently primitive objects on raster images and the application of the boundary selection algorithm, which does not ensure the separation of boundaries of objects that intersect with each other.

Article [15] is devoted to the development of a strategy for object segmentation in digital images. The features of segmentation of objects on digital microstructural images, characteristic of digital images of the metal surface obtained with a microscope, were studied. The concept of a segmentation strategy is proposed, which is based on the assumption of the semantic poverty of neighbourhood relations provided by the iconic level of the digital matrix of the image, the application of which leads to regular errors. It is assumed that the influence of the error and the segmentation goal is minimized when moving to a more complex semantic model of neighbourhood relations characteristic of the property of the object. The forms of syntactic neighbourhood relations defined in [14] were applied.

In [18], a method of segmentation of medical images showing identical objects is proposed. The identified segmentation method is based on the analysis of the homogeneity index of the array of input images, which is calculated based on the brightness of such images. It is declared that due to the use of this method, it is possible to qualitatively divide the test image into the background, constituent parts, and contours of the target objects. The specified image segments are characterized

This work was supported in part by Institute of Electronics and Information Technology of Lublin University of Technology.

I. Tereikovskyi is with Department of System Programming and Specialised Computer Systems of the National Technical University of Ukraine, Igor Sikorsky Kyiv Polytechnic Institute, Ukraine (e-mail: terejkowski@ukr.net).

O. Korchenko, R. Ziubina and O. Veselska are with Department of of Computer Science and Automatics of the University of Bielsko-Biala, Bielsko-Biala, Poland (e-mail: rziubina@ath.bielsko.pl).

S. Bushuyev is with Department of project management Kyiv National University of Construction and Architecture, Ukraine (e-mail: sbushuyev@ukr.net).

O. Tereikovskyi is with Department of Information Technology Security of National Aviation University, Kyiv, Ukraine, (e-mail: tereikovskyio@gmail.com).

using an array of binary images. As evidenced by the results of the experiments given in [18], the method does not allow to ensure sufficient quality of segmentation of arbitrary images, which can be explained by the use of exclusively classical approaches to image processing based on explicit algorithmic classification rules using threshold values that are set using expert rules.

In [13], the justification of the multi-stage algorithm for selection, detection and evaluation of the parameters of images of aerial objects for automatic tracking systems is given. It is indicated that video cameras recording images in the visible and infrared range are the source of image acquisition. The algorithm is based on the principle that to choose between the hypothesis Y_1, which indicates the presence of the object, and the hypothesis Y_2, which indicates its absence, the value of the coefficient of the likelihood ratio is calculated, which is compared with a predetermined threshold.

In the work [19], which is devoted to the improvement of the computer vision of industrial robots, the justification of algorithms and methods of selection invariant to rotation, transfer and scaling of features of objects are given. The above developments are based on the methods of skeletonization, contour construction and image segmentation, which allow obtaining from the initial image first objects from open lines, then objects from closed lines, and as a result a so-called silhouette form, which consists of planar objects.

It should be noted that, in general, the proposed approaches to the segmentation of the initial images of the dissertation work [19] have no fundamental differences from the works [13, 18], and its scientific novelty mainly consists in the application of certain classical image processing algorithms and the use of certain processing coefficients. For example, the applied skeletonization algorithm can be presented in the form of a symbiosis of well-known image binarization and denoising algorithms.

In the article [2], the algorithm for the automatic selection of moving objects in the case of multispectral observation is considered. When developing the algorithm, mathematical models of images were used, which provide the possibility of taking into account geometric transformations caused by the movement of the image sensor. In addition, the possibility of levelling noise associated with geometric distortions caused by atmospheric disturbances is declared. In an analytical form, the decision-making rule for assigning a single point to the target object is written in the form:

$$u^*(b_1,b_2)=\begin{cases}1, if\ p\left(b_1,\frac{b_2}{h}=0\right)<\delta\\ 0, if\ p\left(b_1,\frac{b_2}{h}=0\right)\geq\delta\end{cases}, \quad (1)$$

where (b_1,b_2) is the decision-making result; δ is the threshold value; b_1,b_2 – pixel brightness in each of the observation channels; $p\left(b_1,\frac{b_2}{h}=0\right)$ is the conditional density of the brightness distribution of the target image.

To calculate the threshold value δ, specific expressions are used, which, as in works [2, 15, 19], include sets of coefficients specified using expert rules.

Note that the OpenCV library, which is used to implement computer vision tools, also implements services that can be classified as the selection of objects on a raster image. These services include boundary selection, K-means (clustering), watershed and threshold segmentation. In the case of clustering, the feature space of image objects is divided into clusters if the specified criterion of this object exceeds a predetermined value. All objects in the image are first distributed by clusters, and then by segments that unite the elements of one cluster. Combining into a segment is implemented by defining the homogeneity criterion. The effectiveness of the functioning of these services mainly depends on empirically determined threshold coefficients.

Thus, the analysis of modern literary sources related to the basic technologies for solving the problem by selecting objects on images allows us to reasonably assert that the features of these technologies, which consist in the use of empirically determined coefficients, significantly complicate their application in solving multivariate and multifactorial problems of segmentation of medical images. A promising way to correct these shortcomings is the use of neural network solutions in the methods of semantic segmentation of medical images [3, 12]. At the same time, known neural network systems for the segmentation of raster images are highly specialized and insufficiently adapted to the variability of application conditions [10, 17].

III. Formulation of the Research Problems

Development and research of a neural network model designed to determine the mask of an object in medical images.

IV. Development and Research of a Neural Network Model

As a starting point of the development, the approach given in [20, 21] for determining the most effective neural network model was used, which is generally described as follows:

$$N_{ef}=max\left(E_{N_1},E_{N_2},\dots E_{N_K}\right), N_k\in\{N\}_K, \quad (2)$$

where N_{ef} is the most effective type of neural network model, E_{N_K} is the efficiency of the k-th type of neural network model, K is the number of acceptable types of neural network models, N_k is the k-th type of neural network model, $\{N\}_K$ is the set of acceptable types of neural network models.

At the same time, to formalize the procedure for evaluating the effectiveness of the kth type of neural network model, it is possible to use the results of works [10, 22], which provide the rationale for such a procedure using expert evaluation methods.

When determining the set of acceptable types of neural network models, it is taken into account that the vast majority of tested neural network models of semantic segmentation use an encoder and a decoder built based on a convolutional neural network of one of the modern types [1, 3]. The basic structural diagram of such a model is shown in fig. 1.

The complexity of developing and testing original convolutional neural networks that can be used as the basis of an encoder and decoder determines the feasibility of forming the set $\{N\}_K$ based on neural network models that have proven their effectiveness in related tasks of semantic segmentation.

Based on the results of [5, 6], acceptable types of neural network models include U-Net, YOLOv4, and PSPNet, the parameters of which are given in works [4, 7, 16]. Note that the U-Net encoder and decoder are slightly modified versions of the VGGtype convolutional neural network.

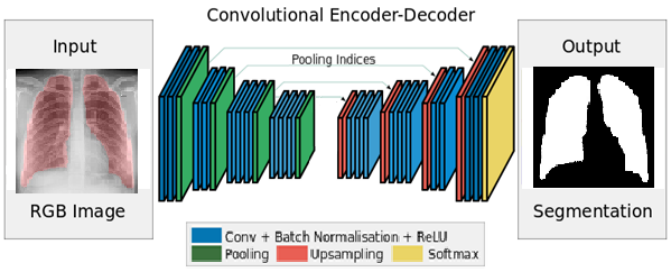


Fig. 1. Structural diagram of the neural network model designed for semantic segmentation of images

The modification mainly consists in removing the fully connected and output layers of neurons from the VGG network. The YOLOv4 neural network model is based on the GoogLeNet architecture, which made it possible to use the so-called Inception modules in the model, which provides the possibility of reducing the resource consumption of the model without a noticeable loss of accuracy and learning speed. Another feature of YOLOv4 is the use of PAN (Path Aggregation Networks) technology, which uses feature pyramids to combine feature maps of the lower and upper levels of the neural network model. The PSPNet neural network model is based on the ResNet architecture and ensures that contextual information is taken into account when selecting an object mask [7, 11]. For this, the model uses an intermediate pool for the classification and comparison of pyramids of selected features.

According to the results of [8, 23, 25], the Accuracy and Loss metrics, which are traditionally used to evaluate the effectiveness of the type of neural network model, were used to estimate E_{N_K}. At the same time, the specificity of the task of determining the mask of an object on a raster medical image necessitates the use of specific expressions to assess the accuracy of the localization of a fragment of a raster image.

Based on the results [10, 16], IoU was used as the Accuracy indicator, and Dice Loss was used as the Loss indicator, calculated using expressions (3, 4).

$$IoU(A,B) = |A \cap B|/|A \cup B|, \quad (3)$$

$$DiceLoss(A,B) = 2|A \cap B|/|A| + |B|, \quad (4)$$

where A, B are the areas to be compared.

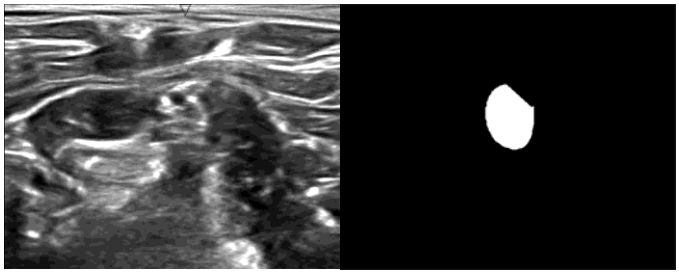
The use of the IoU metric is explained by its approbation in the problem of bitmap images segmentation, and the use of the Dice Loss metric is explained by its stability when training a neural network model on an unbalanced sample.

The developed neural network models U-Net, YOLOv4 and PSPNet are implemented using the Python programming language and verified on examples of monochrome images of the brachial plexus of nerve fibres presented in a freely available database available for download at the link https://www.kaggle.com/ competitions/ultrasound-nerve-segmentation/data.

The specified database was formed in order to develop an automated procedure for determining the location of the catheter for giving pain medication to the patient.

These images were obtained using ultrasound diagnostics. Each image is saved in a separate TIF file. In total, the database contains 16,808 TIF files. The nerve plexus shown in a separate image corresponds to a separate patient and is annotated by experts using a mask that corresponds to the region of the given plexus. An example of a neural plexus and a corresponding mask is shown in Fig. 1.

The size of the original images is 580x420 pixels, the resolution is 96 dpi, and the colour depth is 8 bits. Masks are binary images with a size of 580x420 pixels. The database used to form the training, validation, and the test sample is specified. According to the recommendations [21, 24], the volume of the training sample is 3994 examples, the validation sample is 845, and the test sample is 846 examples.



Original image Mask

Fig. 2. An example of the annotation of the nervous plexus

The graphs of the accuracy indicators of the segmentation of nerve plexuses built on the basis of the results of experiments conducted using the U-Net, PSPNet and YOLOv4 neural network models are shown in fig. 3-5.

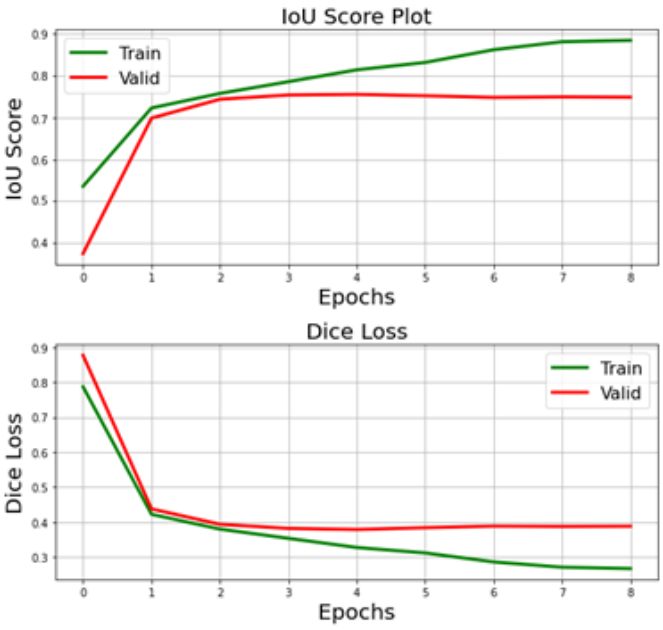


Fig. 3. Graphs of accuracy indicators of nerve plexus segmentation using the U-Net neural network model

As evidenced by the analysis of the graphs shown in Fig. 2-5, the accuracy of the semantic segmentation of medical images using the considered neural network models is approximately the same at the later stages of training. At the same time, we can find out sharp changes in the segmentation accuracy indicators of the validation data in the learning process for the PSPNet network and the YOLOv4 network, which indicates a certain instability of the learning of these neural network models.

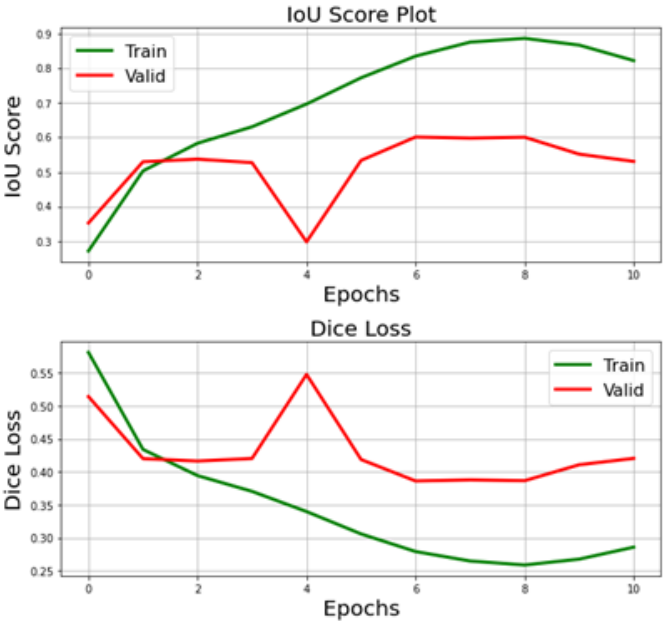


Fig. 4. Graphs of accuracy indicators of nerve plexus segmentation using the PSPNet neural network model

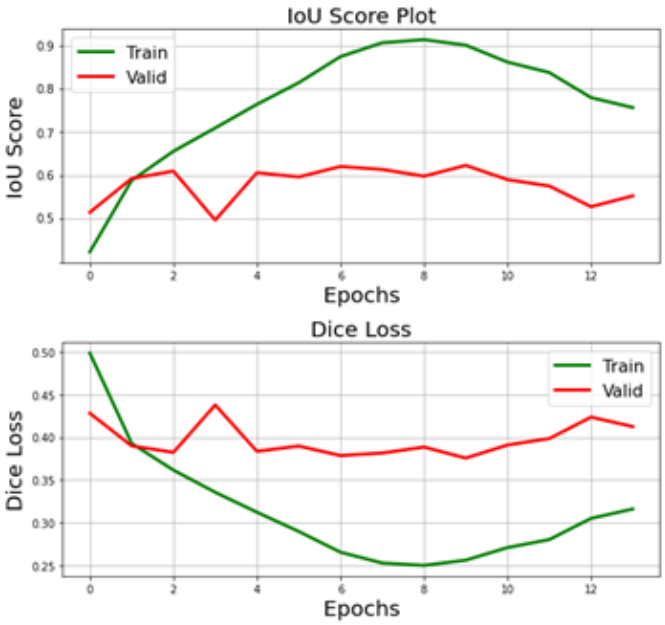
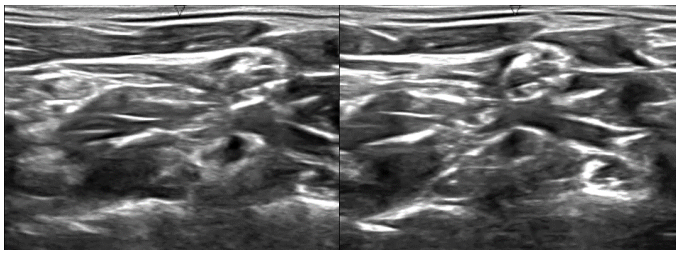


Fig. 5. Graphs of accuracy indicators of nerve plexus segmentation using the YOLOv4 neural network model

For a better interpretation of the obtained accuracy indicators, expert comparison of the results of image segmentation with the help of the considered neural network models with each other and with the mask presented in the training sample was carried out. The expert comparison procedure is illustrated with the help of fig. 6-10. In fig. 6 shows three examples of different test images (A, B, C) to be segmented. In fig. 7 shows the expected output (mask) for each of the test images. In fig. 8 shows the segmentation results of the test image A performed by the U-Net, YOLOv4 and PSPNet networks, in Fig. 9 and fig. 10 - segmentation results of test image B and image C, respectively. It should be noted that the preparation of data for expert comparison and the comparison procedure itself is quite a complex and time-consuming process. Therefore, due to limited resources, the expert comparison concerned only 100 images of the training sample.



A B

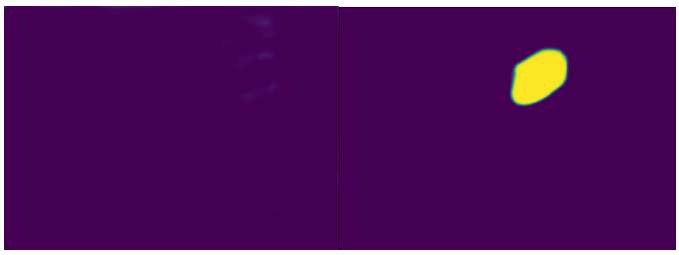
Fig. 6. Test images of the nerve brachial plexus

A B

Fig. 7. Expected output (mask) for test images

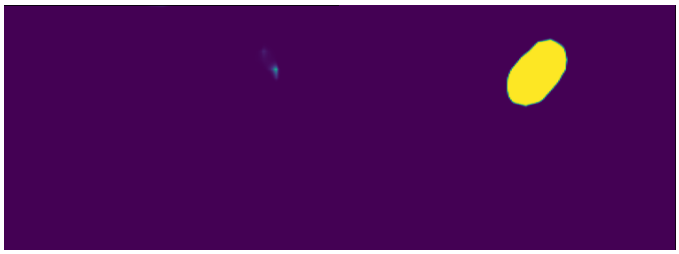
A B

Fig. 8. Results of mask detection for test images A and B performed by U-Net



A B

Fig. 9. Results of mask detection for test images A and B performed by YOLOv4



A B

Fig. 10. Results of mask detection for test images A and B performed by PSPNet

The data of the expert comparison of the results of the mask determination on the test images indicate that in most cases the quality of the segmentation by the U-Net network is sufficient and significantly exceeds the quality of the segmentation by the YOLOv4 and PSPNet networks. In addition, the results obtained generally correspond to the data of sources [7, 16], which prove the high efficiency of the U-Net neural network model in segmenting small halftone images in the case of a small training sample. Such a statement, on the one hand, indicates the expediency of using the U-Net network, and on the other hand, it contradicts the values of the accuracy indicators displayed on the graphs in Fig. 3-5. This indicates the need to improve the mathematical support used to calculate the accuracy indicators of image segmentation using neural network models, which determines the ways of further research in the direction of semantic segmentation of medical images using neural networks. It is also advisable to correlate the ways of further research with the adaptation of the parameters of the neural network model of the most effective type to the conditions of the task. In addition, it is advisable to explore the possibility of introducing an attention mechanism into the encoder and decoder, which can increase the potential for detecting object masks not only on static images, but also in the video sequence.

V. CONCLUSIONS

The analysis of scientific and practical works in the field of developing systems for extracting object masks on bitmap images indicates that the most promising direction for increasing their efficiency is the use of neural network tools adapted to the expected conditions of use. At the same time, the developers of well-known neural network solutions do not pay enough attention to the issue of forming theoretical approaches to adapting the parameters of a neural network model to the most significant conditions of the task of semantic image segmentation. An approach is proposed to determine the most effective type of neural network model, which provides for expert evaluation of the effectiveness of acceptable types of models and conducting computer experiments to make a final decision. As a result of the research, it was determined that among the tests in the tasks of segmentation of raster images, the views of neuronetwork models for viewing the mask on medical raster images of a small size, the U-Net model is the most effective. The use of this neural network model ensures the accuracy of mask selection at the level of 0.89. At the same time, the necessity of improving mathematical support, which is used to calculate accuracy indicators of image segmentation using neural network models, is determined. Also, it is advisable to correlate the ways of further research with the implementation of the neural network model of the attention mechanism in the encoder and decoder, which will allow to increase the efficiency of the selection of object masks in the video sequence.

REFERENCES

[1] U. Adithya, C. Nagaraju, "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. https://doi.org/10.5815/ijeme.2021.02.04

[2] B. Alpatov, P. Babayan. “Selection of moving objects in a sequence of multispectral images in the presence of geometrically distorted ones.” Herald of RGRTU. 2008. Issue 23. P. 18-25.

[3] V. Badrinarayanan, A. Kendall, R. Cipolla. “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.” https://doi.org/10.48550/arXiv.1511.00561

[4] A. Bochkovskiy, C. Wang, H. Liao. “YOLOv4: Optimal Speed and Accuracy.” https://doi.org/10.48550/arXiv.2004.10934

[5] I. Deepa, A. Sharma, “Multi-Module Convolutional Neural Network Based Optimal Face Recognition with Minibatch Optimization”, International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.14, No.3, pp. 32-46, 2022. https://doi.org/10.5815/ijigsp.2022.03.04

[6] D. Diwakar, D. Raj, “Recent Object Detection Techniques: A Survey”, International Journal of Image, Graphics and Signal Processing (IJIGSP), Vol.14, No.2, pp. 47-60, 2022. https://doi.org/10.5815/ijigsp.2022.02.05

[7] Z. Hengshuang, S. Jianping, Q. Xiaojuan, W. Xiaogang, J. Jiaya. “Pyramid Scene Parsing Network.” https://doi.org/10.48550/arXiv.1612.01105.

[8] V. Hoai Viet, H. Nhat Duy, “Object Tracking: An Experimental and Comprehensive Study on Vehicle Object in Video”, International Journal of Image, Graphics and Signal Processing (IJIGSP), Vol.14, No.1, pp. 64-81, 2022. https://doi.org/10.5815/ijigsp.2022.01.06

[9] A. Horytov, S. Yakovchenko. “Selection of parametrically defined objects on a low-resolution image.” Management, computing and informatics. 2017. No. 2. P.88-90. https://doi.org/10.21293/1818-0442-2017-20-2-88-90

[10] Z. Hu, I. Tereykovskiy, Y. Zorin, L. Tereykovska, A. Zhibek “Optimization of convolutional neural network structure for biometric authentication by face geometry.” Advances in Intelligent Systems and Computing. 2019. Vol. 754. P. 567-577. https://doi.org/10.1007/978-3-319-91008-6_57

[11] T. Kong, F. Sun, H. Liu, Y. Jiang, L. Li, J. Shi, FoveaBox: Beyound Anchor-Based Object Detection, IEEE Trans. Image Process. 29 (2020) 7389–7398. https://doi.org/10.1109/TIP.2020.3002345

[12] Y. LeCun et al. “Learning Hierarchical Features for Scene Labeling.” URL: http://yann.lecun.com/exdb/publis/pdf/farabet-pami-13.pdf (access date: 02/02/2017). https://doi.org/10.1109/TPAMI.2012.231

[13] Muraviev “Models and algorithms of image processing and analysis for systems of automatic tracking of aerial objects.” author's review. diss. for the application of scientific degrees of candidate of technical sciences: special. 05.13.01 - system analysis, management and processing. Ryazan 2010. 17 p.

[14] P. Oniskiv, Y. Lytvynenko “Analysis of image segmentation methods”. Theoretical and applied aspects of radio engineering, instrument engineering and computer technologies: materials of IV all-Ukrainian. science and technology conf. 2019. P.48-49.

[15] D. Perfil`ev “Segmentation Object Strategy on Digital Image”. Journal of Siberian Federal University. Engineering & Technologies. 2018. No. 11(2). R. 213-220. https://doi.org/10.17516/1999-494X-0024

[16] O. Ronneberger, P. Fischer, T. Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” https://doi.org/10.48550/arXiv.1505.04597

[17] J. Shen, “Motion detection in colour image sequence and shadow elimination.” Visual Communications and Image Processing. 2014. Vol. 5308. P. 731-740. https://doi.org/10.1117/12.525653

[18] O. Shkurat “Methods and information technology of processing archival medical images.” dissertation. ... candidate technical Sciences: 05.13.06. K., 2020. 211 p.

[19] N. Stulov “Algorithms for the selection of basic features and methods of formation invariant to rotation, transfer, and rescaling of features of objects.” autoref. diss. for the application of scientific degrees of candidate of technical sciences: special. 05.13.01 - system analysis, management and processing. Vladimir 2006. 16 p.

[20] I. Tereikovskyi, I. Subach, O. Tereikovskyi, L. Tereikovska, S.Toliupa, V. Nakonechnyi “Parameter Definition for Multilayer Perceptron Intended for Speaker Identification.” IEEE International Conference on Advanced Trends in Information Theory. Kyiv, Ukraine. 2019. P. 227-231. https://doi.org/ 10.1109/ATIT49449.2019.9030504

[21] S. Toliupa, Y. Kulakov, I. Tereikovskyi, O. Tereikovskyi, L. Tereikovska, V. Nakonechnyi “Keyboard Dynamic Analysis by Alexnet Type Neural Network.” IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering. 2020. P. 416-420. https://doi.org/10.1109/TCSET49122.2020.235466

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

Advanced Information and Communications Technologies (AICT), 2019, pp. 323-327, https://doi.org/10.1109/AIACT.2019.8847847

[23] H. Wang, X. Wang, L. Yu and F. Zhong, "Design of Mean Shift Tracking Algorithm Based on Target Position Prediction," 2019 IEEE International Conference on Mechatronics and Automation (ICMA), 2019, pp. 1114-1119, https://doi.org/10.1109/ICMA.2019.8816295

[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. https://doi.org/10.1109/ATIT50783.2020.9349302

[25] S. Zhang, L. Wen, X. Bian, Z. Lei, S.Z. Li, Single-Shot Refinement Neural Network for Object Detection, in: 2018: pp. 4203-4212. https://doi.org/10.48550/arXiv.1711.06897