DEVELOPMENT OF AN IMAGE SEGMENTATION MODEL BASED ON A CONVOLUTIONAL NEURAL NETWORK

This paper has considered a model of image segmentation using convolutional neural networks and studied the process efficiency based on models involving training the deep layers of convolutional neural networks. There are objective difficulties associated with determining the optimal characteristics of neural networks, so there is an issue related to retraining the neural network. Eliminating retraining by determining the optimal number of epochs only would not suffice since it does not provide high accuracy.

The requirements for the set of images for training and model verification were defined. These requirements are best met by the image sets PASCAL VOC (United Kingdom) and NVIDIA-Aerial Drone (USA).

It has been established that AlexNet (Canada) is a trained model and could perform image segmentation while object recognition reliability is insufficient. Therefore, there is a need to improve the efficiency of image segmentation. It is advisable to use the AlexNet architecture to build a specialized model, which, by changing the parameters and retraining some layers, would allow for a better process of image segmentation.

Five models have been trained using the following parameters: learning speed, the number of epochs, optimization algorithm, the type of learning speed change, a gamma coefficient, a pre-trained model.

A convolutional neural network has been developed to improve the accuracy and efficiency of image segmentation. Optimal neural network training parameters have been determined: learning speed is 0.0001, the number of epochs is 50, a gamma coefficient is 0.1, etc. An increase in accuracy by 3 % was achieved, which makes it possible to assert the correctness of the choice of the architecture for the developed network and the selection of parameters. That allows this network to be used for practical tasks related to image segmentation, in particular for devices with limited computing resources

Tags: image processing, image segmentation, convolutional neural networks, unmanned aerial vehicle

Bogdan Knysh
PhD, Associate Professor
Department of Electronics
and Nanosystems*
E-mail: tutmos-3@i.ua
Yaroslav Kulyk
PhD, Associate Professor
Department of Automation and
Intelligent Information Technologies*
E-mail: Yaroslav_Kulik@i.ua
*Vinnytsia National Technical University
Khmelnytske highway, 95,
Vinnytsia, Ukraine, 21021

1. Introduction

Image processing is extremely important in modern science and practice, so it is constantly evolving and improving. Image processing could be used in many industries, namely precision farming (agromonitoring), safety systems, quality control systems, etc. These areas involve vision systems, robotic complexes, unmanned aerial vehicles (UAVs), video surveillance systems, web services and mobile applications for identification and search.

One type of image processing is segmentation, which is widely used in industry, art, medicine, space, process management, automation, and many other industries [1]. Segmenting images involves splitting the input image into areas that are not similar in terms of a certain criterion. The result is a set of regions that all together cover the entire input image.

There are a large number of image segmentation methods, among which the most common are methods based on image graph analysis, clustering, contour and threshold methods, as well as neural network methods. At the same time, these methods [2–4] work much faster with small image sizes and limited coloration.

Since segmentation precedes a higher image processing level, certain requirements apply to segmentation methods. In general, these requirements could be formulated as follows:

– maximum correspondence of the segmented area to the real object;

– high performance;

– resistance to errors;

– high accuracy.

Therefore, it is necessary to analyze the methods of image segmentation and choose the optimal one according to the above requirements, in particular, high accuracy. It is also worth considering the parameters that characterize these methods, a change in which has a direct impact on the accuracy, performance speed, and overall efficiency of the segmentation process.

A modern relevant direction of production is the development of precision farming, which is based on the results from agromonitoring, namely, on the images from UAV video cameras for the analysis of vegetation, assessment of the areas of damage to crops, forecasting yields, etc. An important criterion, in this case, is the UAV's ability to avoid collisions with close objects, determine the position in space, the direc-

tion and trajectory of a flight by acquiring input data in the form of segmented images.

The effectiveness of these systems is determined by the accuracy of segmentation, for the evaluation of which experimental research is required.

2. Literature review and problem statement

Paper [5] reports the results of image segmentation using clustering. The search for clusters in images based on similar characteristics is shown. This process is characterized by high speed, accuracy, and resistance to errors at the initial stage. However, unresolved issues remain related to a significant decrease in the effectiveness of the method when the image size increases. The reason for this may be objective difficulties associated with the large number of small elements in the image and the poor operation of the method in the noisy image, which makes the relevant studies impractical. Study [6] considers threshold methods of image segmentation, which are characterized by high performance and ease of implementation. However, unresolved issues related to rather low segmentation accuracy remain. This causes difficulties when limiting the color palette at the boundaries of image elements, which prevents the use of the method for full-color images. Paper [7] reports the results of image segmentation using contour methods. The resistance to changing the parameters of input images is shown. However, the problem of tearing up areas of the image remains unresolved. In addition, the methods are characterized by low performance. This could be caused by objective difficulties associated with using a wide palette of colors in real images. This limits the use of this method for photorealistic images. Study [8] reports the results of image segmentation using methods based on image graph analysis. It is shown that the obtained data do not depend on changes in the uniformity of colors and the size of the input images. However, these methods demonstrate poor performance and require a lot of memory. This could be caused by the difficulties of choosing a metric separately for each type of image and using a large number of graph elements to segment the image. This makes the cited studies costly in terms of the use of computing resources.

An option to overcome the above difficulties associated with insufficient accuracy, efficiency, and performance may be to use neural network-based image segmentation methods [9], specifically convolutional ones. This is the approach used in work [10], where neural networks for color images are used but training parameters are not analyzed. In addition, a similar principle is implemented in paper [11], where learning parameters are given without explaining the possibilities of use, in particular for segmentation tasks.

All this suggests that it is advisable to conduct a study to improve the effectiveness of neural network learning, which would significantly improve the accuracy of image segmentation.

3. The aim and objectives of the study

The aim of this study is to improve the architecture of a convolutional neural network to segment images and select the learning parameters of this network. That would make it possible to build a new neural network with improved accuracy to segment images, which could be used as a pre-trained neural network for other tasks.

To accomplish the aim, the following tasks have been set:
– to investigate neural network models based on the PASCAL VOC set;
– to evaluate the Voc-3 model based on the NVIDIA-Aerial Drone set.

4. Materials and methods to study image segmentation using convolutional neural networks

We studied image segmentation by applying appropriate methods based on convolutional neural networks, taking into consideration neural network learning parameters. To test the effectiveness of these methods, PASCAL VOC and NVIDIA-Aerial Drone sets were used, containing a large number of images with labeled groups of pixels and defined object classes. PASCAL VOC contains images where the desired class is clearly distinguished from among other pixels by color [12]. The NVIDIA-Aerial Drone contains images acquired from video cameras attached to UAV while being shot from a height of several hundred meters [13]. Our research was carried out using the DIGITS programming environment involving the Caffe environment, which is designed to deeply train the neural network, taking into consideration the speed and modularity in the development of the model. The combination of these environments makes it possible to quickly train neural networks with deep layers and is used for the tasks of classification, segmentation of images, and detection of objects on them. DIGITS contains a pre-trained AlexNet model, which is characterized by parameters adapted for segmentation (Table 1) and has a flexible architecture (Fig. 1).

Table 1

AlexNet model parameters

| Level	Kernel size	Filter size	Increment	Filling	Data size	Activation function
Input	–	–	–	–	227×227×3	–
Layer Conv1	96	11×11	4	–	55×55×96	ReLU
Layer Max Pool1	–	3×3	2	–	27×27×96	–
Layer Conv2	256	5×5	1	2	27×27×256	ReLU
Layer Max Pool2	–	3×3	2	–	13×13×256	–
Layer Conv3	384	3×3	1	1	227×13×384	ReLU
Layer Conv4	384	3×3	1	1	13×13×384	ReLU
Layer Conv5	256	3×3	1	1	13×13×256	ReLU
Layer Max Pool3	–	3×3	2	–	6×6×256	–
Dropout1	–	–	–	–	6×6×256	Softmax
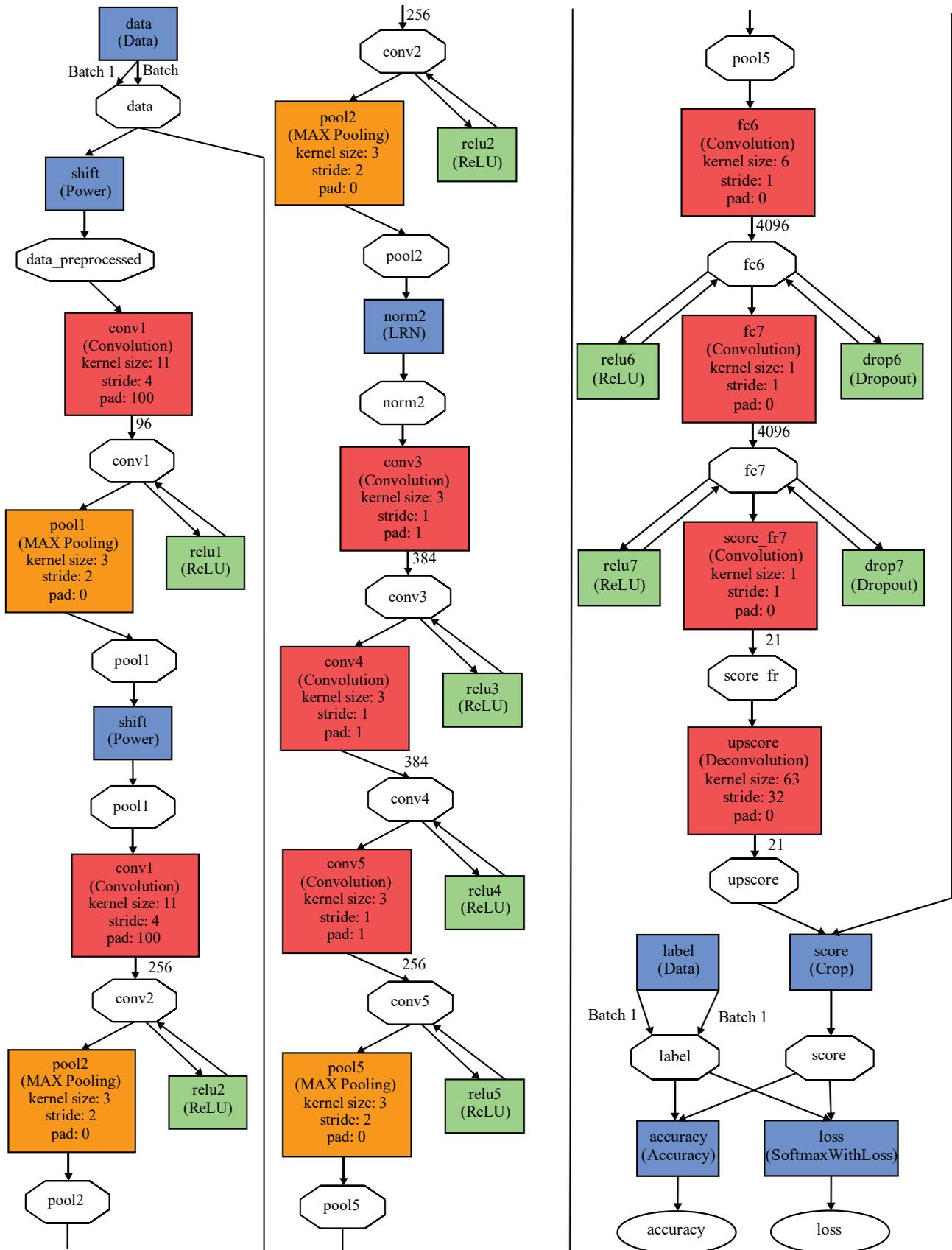
Fig. 1. AlexNet neural network model architecture

The AlexNet architecture consists of five convolutional layers, among which Pooling and normalization layers are located, and three fully interconnected layers. Moreover, their parameters may change in the learning process. An image with a 227×227×3 RGB palette is fed to the input. The filter size of the first layer is 11×11. The 96 kernel is

used in increments of 4. The output activation function is Softmax, and in layers – ReLU, which makes it possible to improve performance by 6 times. AlexNet has a total of 62.3 million parameters [14]. Compared to similar models [10], AlexNet has significantly lower memory size requirements (by 10 times) with an increase in accuracy by more than 70 %.

Thus, the AlexNet architecture was used as the base one to build a specialized FCN-AlexNet model by adding a fully-connected convolution layer by making the following changes to DIGITS:

– the training and verification networks should be combined into a single neural network;

– specialized layers to retrieve image data for SBDDSeg-DataLayer and VOCSegDataLayer segmentation must be replaced by simple layers to retrieve data from LMDB;

– the power layer must be added to the shear layer of the input data;

– one needs to add a bilinear filtration of the weight of neurons to the upscore layer;

– one needs to add a layer of accuracy to evaluate the effectiveness of the model on the test set of images;

– training and verification processes should be normalized by entering the batch size=1 parameter [15].

The choice of the FCN-AlexNet model optimization algorithm is determined by the features of image segmentation, for which it is necessary to have a good convergence of the algorithm, and for practical use – high performance. The comparison of algorithms [16] shows that for the task of image segmentation, the best performance results are demonstrated by Adam (an increase of 10–50 %) and a stochastic gradient descent (increase by 5–20 %) depending on the metric. In addition, these algorithms show good convergence, especially the stochastic gradient descent.

The main chosen indicators of the effectiveness of neural network training, which were determined during our study, were accuracy and errors. Accuracy is calculated as a percentage of correctly defined classes (or pixels belonging to a particular class) in the image relative to all classes (or all pixels). The error is calculated as a percentage of incorrectly recognized classes (or pixels belonging to a particular class) in the image relative to all classes (or all pixels). To assess the effectiveness of neural network training, optimal neural network parameters are determined. These parameters are the duration of training (the number of epochs), optimization algorithm (an adaptive instant estimation (Adam), stochastic gradient descent (SGD)), the type of change in learning speed, a gamma coefficient, learning speed (an increment value). Combinations of parameters in the process of training 4 models are summarized in Table 2.

The error was investigated on the training and verification sample, which are the above sets of marked segmented images (PASCAL VOC and NVIDIA-Aerial Drone). The PASCAL VOC set contains 20 categories of objects, which include 1,464 images for training, and 1,449 for verification. For the NVIDIA-Aerial Drone set, the images were distributed at a ratio of 80 % per training sample and 20 % for the test using a cross-check to assess the accuracy of the model.

The error was investigated on the training and verification sample, which are the above sets of marked segmented images (PASCAL VOC and NVIDIA-Aerial Drone). The PASCAL VOC set contains 20 categories of objects, which include 1,464 images for training, and 1,449 for verification.

tion. For the NVIDIA-Aerial Drone set, the images were distributed at a ratio of 80 % per training sample and 20 % for the test using a cross-check to assess the accuracy of the model.

Table 2

Combinations of parameters in the training process

Model	Training duration (the number of epochs)	Optimization algorithm	Based on mode	The type of learning speed, gamma coefficient	Learning speed
Voc-1	25	Adam	FCN-AlexNet	Exponential, 0.99	0.0001
Voc-2	30	SGD	FCN-AlexNet	Exponential, 0.95	0.0001
Voc-3	50	SGD	FCN-AlexNet	Stepped (increment size=33), 0.1	0.0001
Voc-4	50	SGD	FCN-AlexNet	Exponential, 0.97	0.00001

The error value on the training and, in particular, on the verification sample should gradually decrease. That would indicate the correctness of neural network training and the lack of retraining, that is, the adequacy of the model. Validation of training results could be defined as a gradual decrease in the error in the test sample. The number of epochs of training is selected from the condition of obtaining the highest accuracy on the test sample in the absence of significant fluctuations in numerical values. The criterion for increasing the epochs of learning is a gradual increase in accuracy on the test sample. The beginning of the drop in accuracy on the test sample is a criterion for retraining, the absence of which is a condition for validation of the model.

5. Results of studying image segmentation using convolutional neural networks

5. 1. Investigating neural network models for image segmentation from the PASCAL VOC set

Our study was carried out on pixelwise marked images from the PASCAL VOC set. Because the sizes of the images in the set are different, if one needs to check the model on a new image, one must perform pixel-by-pixel markup of segmented areas. In addition, accuracy calculation occurs in soft real time, which requires a high-performance neural network with limited memory to ensure high accuracy. The neural network input image comes with an RGB (256 color palette) of up to 10 MB in size, and the output is obtained in PNG (lossless) format. The batch size is 32 with the number of threads equal to 4. Pixel markup (if any) must match the Lightning Memory-Mapped Database annotation format. Models are imported in prototxt format. Segmentation time must not exceed 50 ms for a Full HD image size.

The model performance check is illustrated by the charts that were constructed automatically in the DIGITS programming environment based on the specified parameters given in Table 2. The Caffe environment was used for hardware acceleration of training. The number of accuracy values is equal to the number of epochs of learning. The final accuracy of the model is the accuracy of learning at the last epoch.

Fig. 2 shows the Voc-1 model performance check chart.

Fig. 2 shows that on the training sample an increase in the number of stages of learning (epochs) leads to a decrease in the error in absolute value from 3 to 1.5–2.5, and, after epoch 10, it stabilizes in the range of 1.3–2.7. Accuracy on the test sample reaches 72 % and almost does not increase after the first epoch.

Fig. 3 shows a chart of checking the effectiveness of the Voc-2 model.

Fig. 3 shows that on the training sample an increase in the number of stages of learning (epochs) leads to a decrease in the error in absolute value from 3 to 0.7–0.2, and, after epoch 1 and thereafter, it practically does not change. The accuracy on the test sample reaches 80 % and smoothly increases after the first epoch to a value of 82 %.

Fig. 4 shows the Voc-3 performance check chart.

Fig. 4 shows that on the training sample an increase in the number of stages of learning (epochs) leads to a decrease in the error in absolute value from 3 to 0.6–0.2, and, after epoch 10, it stabilizes in the range of 0.4–0.2. Accuracy on the test sample reaches 78 %, and smoothly increases after the first epoch to a value of 83 %.

Fig. 5 shows the Voc-4 model performance test chart.

Fig. 5 shows that on the training sample an increase in the number of stages of learning (epochs) leads to a decrease in the error in absolute value from 3 to 1.7–0.3, and, after epoch 10, it stabilizes in the range of 0.7–0.3. Accuracy on the test sample reaches 75 % and smoothly increases after the first epoch to a value of 81 %.

Our study results are the findings from the efficiency check of the 4 models given in Table 3.
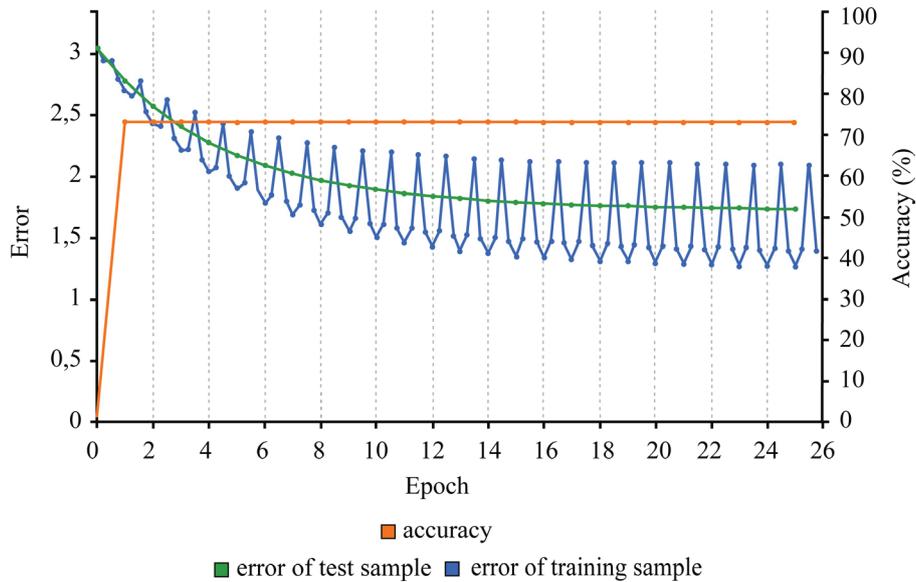

Fig. 2. Charts of change in accuracy, error on the test sample, and error on the training sample depending on the epoch for the Voc-1 model
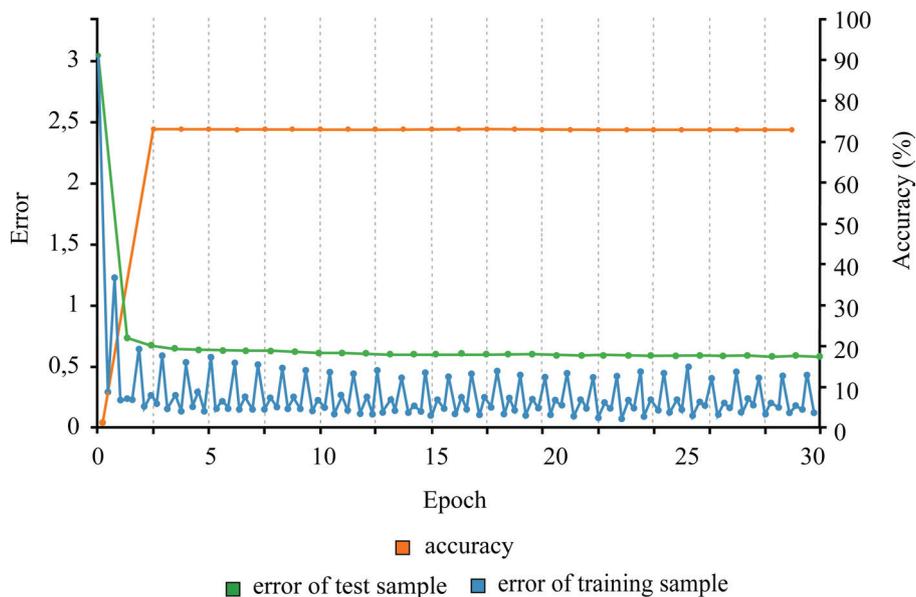

Fig. 3. Charts of change in accuracy, error on the test sample, and error on the training sample depending on the epoch for the Voc-2 model
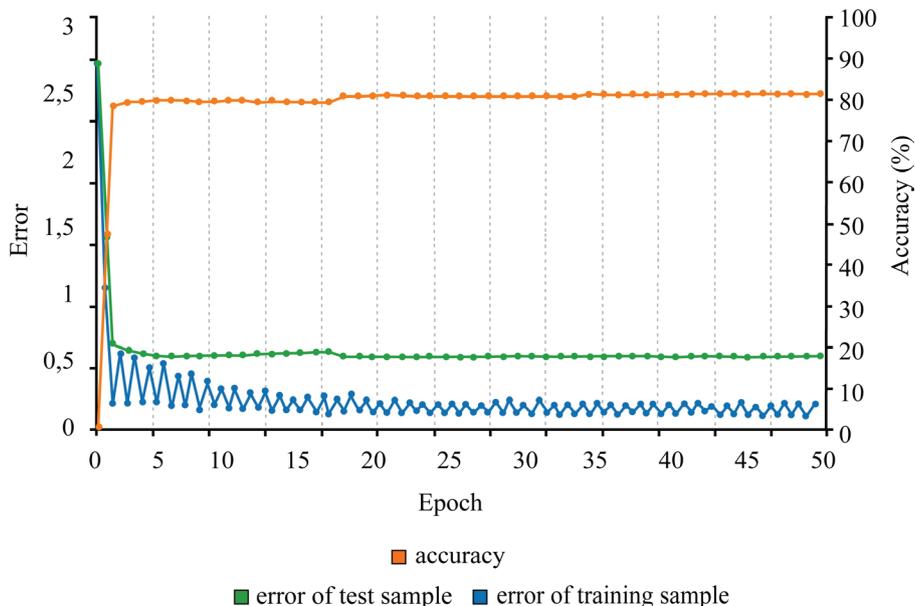
Fig. 4. Charts of change in accuracy, error on the test sample,
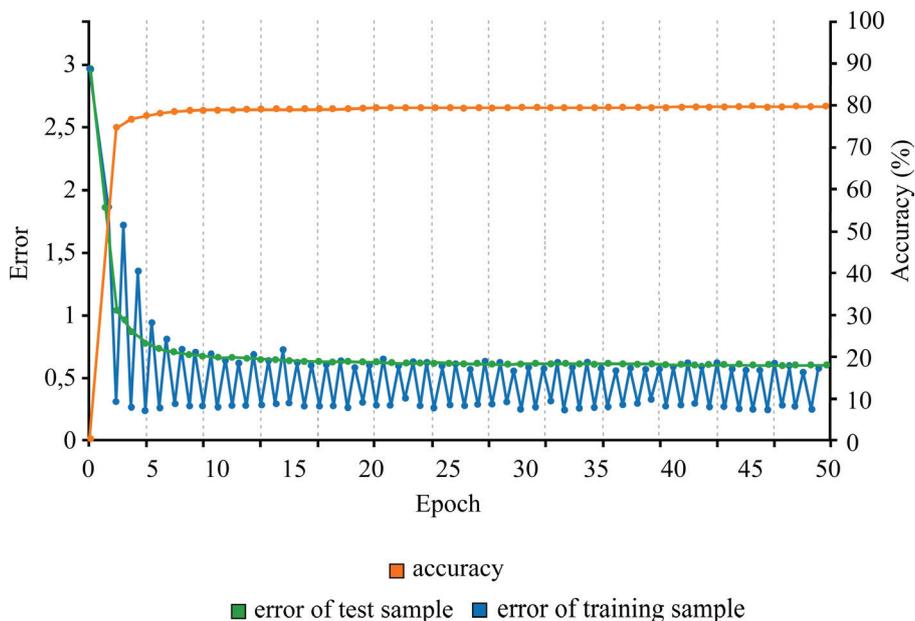and error on the training sample depending on the epoch for the Voc-3 model

Fig. 5. Charts of change in accuracy, error on the test sample,
and error on the training sample depending on the epoch for the Voc-4 model

Table 3

Results of studying the effectiveness of models with
different parameters

Model title	Epoch with the best result/the number of epochs	Accu-racy
Voc-1	25/25	72 %
Voc-2	30/30	82 %
Voc-3	50/50	83 %
Voc-4	50/50	81 %

Table 3 shows that the Voc-3 model demonstrates the greatest accuracy of 83 % at a learning speed of 0.0001 based on SGD with a step-by-step technique for changing the speed of learning. The smallest accuracy value is 72 % for the Voc-1 model, which uses an adaptive instant estimation algorithm.

Fig. 6 shows the image segmentation of animal photography in the DIGITS programming environment using the Voc-3 model for training at different epochs. At the same time, the Caffe environment was used for hardware acceleration of training.

In Fig. 6, when performing object recognition, we highlighted 2 areas: ■ – "animal", ■ – "background".

Fig. 6. Segmenting the image of a cyclist using the trained Voc-3 model for different number of training epochs: a — 3 epochs; b — 12 epochs; c — 30 epochs; d — 50 epochs

5. 2. Evaluation of the trained Voc-3 model for segmenting images from the NVIDIA-Aerial Drone set

In practice, segmentation is part of environmental monitoring involving UAV that requires high segmentation accuracy for control and orientation in space. UAV camcorders acquire a contrasting image, unlike images in the PASCAL VOC set. Therefore, the model with the highest accuracy Voc-3 must be additionally trained on another set of images acquired from a UAV camcorder. This task uses the NVIDIA-Aerial Drone set. At the same time, the model must demonstrate high performance at limited memory to ensure high accuracy in real time.

A set of NVIDIA-Aerial Drone images was used to test the effectiveness of the Voc-3 model trained with DIGITS. The Caffe environment was used for hardware acceleration of training. During our research, we used the values of parameters defined as optimal in the study (Table 2):

– training speed, 0.0001;
– training duration (the number of epochs), 50;
– optimization algorithm, SGD;

– the type of change in the speed of learning, stepped;
– a gamma coefficient, 0.1;
– a pre-trained model, fully-connected convolutional neural network FCN-AlexNet.

Thus, the Drone-1 model was built, the results of testing which are shown in Fig. 7.
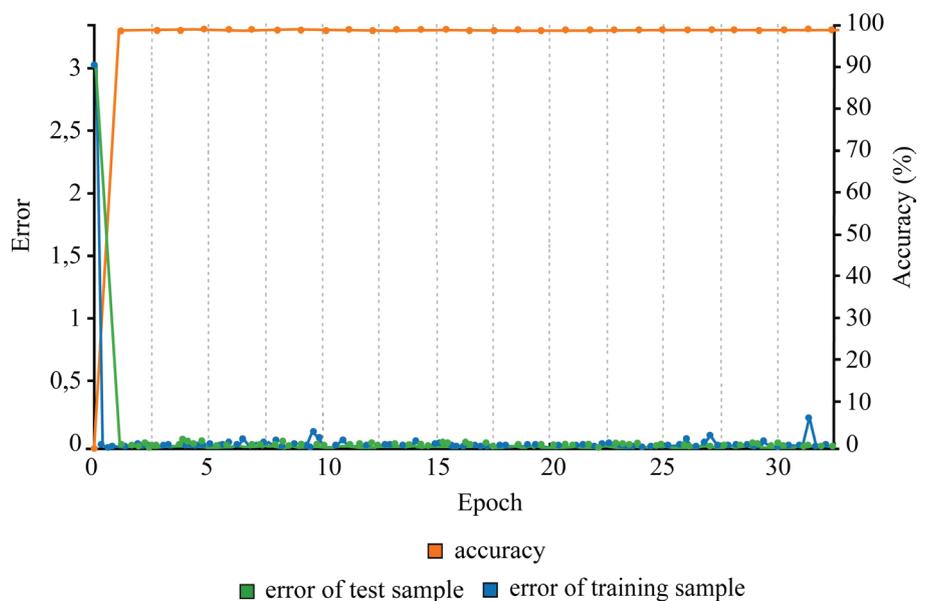

Fig. 7. Charts of change in accuracy, error on the test sample, and error on the training sample depending on the epoch for the Drone-1 model

Fig. 7 shows that the error in absolute value for the training sample of images becomes close to 0 after the first epoch of training, except for episodic cases in certain training epochs. On the test sample of images, the error after the first epoch also becomes close to 0, and the accuracy of the model becomes close to 100 %, and it practically does not change. After the end of epoch 30, the accuracy value is 99 %.

Fig. 8 shows the segmentation of an image acquired from a UAV camcorder for the Drone-1 model in the DIGITS programming environment.
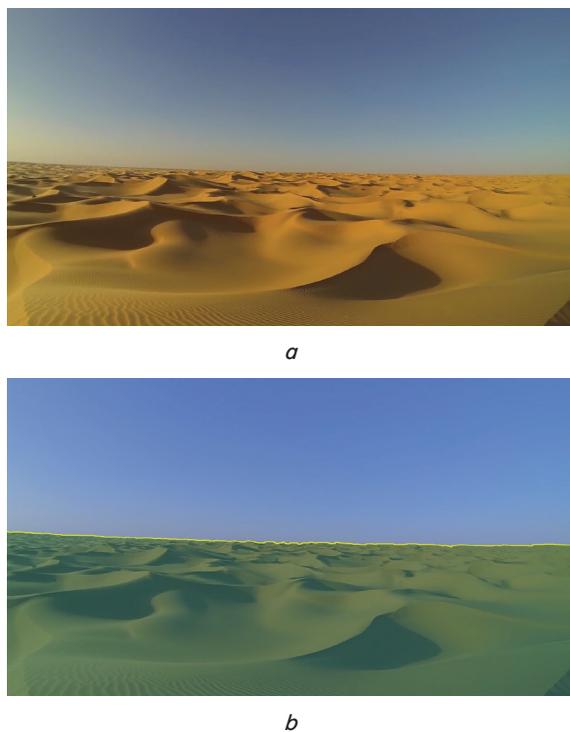

a

b

Fig. 8. Segmentation of an image acquired from NVIDIA-Aerial Drone UAV camcorder for Drone-1 model: *a* — initial image; *b* — completed image segmentation

In Fig. 8, when performing object recognition, 2 areas are highlighted: ■ – "earth", ■ – "sky".

The available performance values make it possible to compare the Drone-1 model with others [2–7], however, it makes sense to compare with models close in architecture that are built during training on a similar basis. Therefore, the comparison was carried out according to the accuracy criterion with some well-known AlexNet-based models developed according to similar parameters, trained on the basis of images acquired from UAV cameras. FireCAMP2 SLIC, VEDAI, NZAM/ONERA Christchurch, ISPRS Potsdam were chosen as such models [17, 18]. The results of the assessment of the accuracy of models are given in Table 4.

Table 4

Results of estimating model accuracy

Model title	Accuracy, %
FIRECAMP2 SLIC	95
VEDAI	96
NZAM/ONERA Christchurch	97
ISPRS Potsdam	97
Drone-1	98

Table 4 shows that the greatest accuracy, namely 98 %, is demonstrated by the Drone-1 model, as others were trained based on images acquired from UAV cameras that were not part of the NVIDIA-Aerial Drone set.

The adequacy, reliability, and convergence of the Drone-1 model were compared to others [17, 18]. To this end, we segmented 100 pixel-hand-marked images from the Aeroscapes set [19]. Calculations were performed in the Jupiter Notebook environment in the Python language. Averaged results are summarized in Table 5.

Table 5

Test results to assess the reliability, adequacy, and convergence of the model

Model title	Accuracy, %	Share of images with incorrectly marked classes, %	Share of incorrectly marked pixels, %
FIRECAMP2 SLIC	84	6	12
VEDAI	82	8	17
NZAM/ONERA Christchurch	81	10	20
ISPRS Potsdam	85	4	16
Drone-1	87	5	13

Table 5 shows that the greatest accuracy, namely 87 %, is demonstrated by the developed Drone-1 model. This indicates the high authenticity of the Drone-1 model.

Objects in the images for validation are marked with specific classes, according to which the image is segmented by a model that performs class designation for selected areas. In the experimental verification of models, our results were incorrectly positive (the presence of a certain class in the image in its absence) and falsely negative (the absence of a certain class in the image in its presence) when designating classes. Table 5 shows that the share of images with incorrectly marked classes is 5 %. Thus, the share of images with properly marked classes is 95 %, which indicates the high adequacy of the model.

Segmentation involves taking into consideration the proportion of incorrectly marked pixels. In experimental model checking, the results of pixel marking were incorrectly positive (marking a pixel by a certain class in the absence of belonging in it) and incorrectly negative (non-marking of the pixel by a certain class when belonging in it). Table 5 shows that the percentage of incorrectly marked pixels is 13 %. Thus, the proportion of correctly marked pixels is 87 %, which indicates high convergence of the model.

6. Discussion of results of studying image segmentation using convolutional neural networks

The results of our study show that the trained Drone-1 model demonstrates high accuracy of image segmentation (Fig. 7). This is due to the choice of optimal parameters of the neural network, as well as the introduction of a convolutional layer into the standard neural network architecture. Drone-1 is based on Voc-3. This model demonstrated the best accuracy values (Table 3) in the learning

process based on the PASCAL VOC set (Fig. 4). That is due to the choice of a stepped function to change the speed of learning (Table 1). Fig. 6 visualizes the result obtained, which confirms the data acquired. Therefore, Voc-3 was chosen for additional training based on the NVIDIA-Aerial Drone set with the same parameters. The Drone-1 model created in this way could be used to segment real images (Fig. 8), the high accuracy of which determines the efficiency of UAV control system. The results of comparing the accuracy of image segmentation (Fig. 8) for Drone-1 and other similar models are given in Table 4. Drone-1 demonstrates, when compared to others, the high values of the accuracy of the segmentation of images that were not in the training and test samples, indicating the adequacy of this model and the lack of retraining.

The accuracy and performance of the developed Drone-1 neural network model accept higher values than similar values derived in [17], by 1–2 % and 20–50 %, respectively. At the same time, the segmentation process does not require significant computing resources at the stage of using the model. Compared to [18], a given model has a 3 % higher image segmentation accuracy. This is achieved by combining training and test networks, replacing specialized layers with simple ones, adding the power and accuracy layers, the bilinear filtering of the weight of neurons, and the batch size parameter. The reliability, authenticity, and convergence of the developed Drone-1 neural network model are comparable with other models [17, 18] while is not inferior to them.

Since neural network model training was conducted for images from the NVIDIA-Aerial Drone set, high segmentation accuracy values are typical of images acquired from the drone's camera, usually due to the high contrast of pixel groups. For other types of images, accuracy probably will not be as high. That requires additional research.

The shortcomings include the long process of training the neural network and the cost of resources at the initial stage of training. Overcoming this disadvantage is possible by the parallelization of computing on GPU, the creation of new compact architectures, and the emergence of more pretrained neural models.

This model could be advanced by further increasing the accuracy, performance speed, as well as decreasing computing resources. That would require complex mathematical modeling, taking into consideration the subject area of application and the development of software modules for a particular system.

7. Conclusions

1. The Voc-1, Voc-2, Voc-3, Voc-4 models of neural networks based on the PASCAL VOC set were investigated. It is established that the highest accuracy is demonstrated by the Voc-3 model, 83 %, at a learning speed of 0.0001, based on SGD with a stepwise method of changing the learning speed. The smallest accuracy value is 72 % for the Voc-1 model, which uses an adaptive instant estimation algorithm. This means that SGD copes better than Adam since Voc-2 and Voc-4 have significantly higher model accuracy indicators, 82 % and 81 %, respectively.

2. The accuracy values of the Drone-1 model, created on the basis of the previously trained Voc-3 model for the segmentation of images from the NVIDIA-Aerial Drone set with the parameters determined on the basis of the study, were determined. These values are quite high because the error in absolute value for the training sample of images becomes close to 0 after the first epoch of training. On the test sample of images, the error after the first epoch also becomes close to 0, and the accuracy of the model becomes close to 100 %; it practically does not change. After the end of epoch 30, the accuracy value is 99 %. The derived accuracy values make it possible to assert the correctness of the choice of network architecture and the selection of parameters. This allows this model to be used for practical tasks of image segmentation, for example, when estimating a fire size, analyzing field vegetation, categorizing crops, etc.

References

1. Bilinskiy, Y. Y., Knysh, B. P., Kulyk, Y. A. (2017). Quality estimation methodology of filter performance for suppression noise in the mathcad package. Herald of Khmelnytskyi national university, 3, 125–130. Available at: https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/23238/47857.pdf?sequence=2&isAllowed=y

2. Kurugollu, F., Sankur, B., Harmanci, A. E. (2001). Color image segmentation using histogram multithresholding and fusion. Image and Vision Computing, 19 (13), 915–928. doi: https://doi.org/10.1016/s0262-8856(01)00052-x

3. Wang, H., Oliensis, J. (2010). Generalizing edge detection to contour detection for image segmentation. Computer Vision and Image Understanding, 114 (7), 731–744. doi: https://doi.org/10.1016/j.cviu.2010.02.001

4. Felzenszwalb, P. F., Huttenlocher, D. P. (2004). Efficient Graph-Based Image Segmentation. International Journal of Computer Vision, 59 (2), 167–181. doi: https://doi.org/10.1023/b:visi.0000022288.19776.77

5. Chitade, A. Z., Katiyar, S. K. (2010). Colour based image segmentation using k-means clustering. International Journal of Engineering Science and Technology, 2 (10), 5319–5325. Available at: https://www.researchgate.net/publication/50361273_Color_based_image_segmentation_using_K-means_clustering

6. Bishop, C. M. (1995). Neural Networks for Pattern Recognition. Clarendon Press, 482. Available at: http://people.sabanciuniv.edu/berrin/cs512/lectures/Book-Bishop-Neural%20Networks%20for%20Pattern%20Recognition.pdf

7. Chaudhuri, D., Agrawal, A. (2010). Split-and-merge Procedure for Image Segmentation using Bimodality Detection Approach. Defence Science Journal, 60 (3), 290–301. doi: https://doi.org/10.14429/dsj.60.356

8. Keuchel, J., Schnorr, C. (2003). Efficient Graph Cuts for Unsupervised Image Segmentation using Probabilistic Sampling and SVD-based Approximation. University of Mannheim. Available at: https://madoc.bib.uni-mannheim.de/1805/1/2003_9.pdf

9. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L. (2015). Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv.org. Available at: https://arxiv.org/pdf/1412.7062.pdf

10. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems, 1097–1105. Available at: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

11. Sytnyk, V. F. (2004). Systemy pidtrymky pryiniattia rishen. Kyiv: KNEU, 614. Available at: http://kist.ntu.edu.ua/textPhD/sppr1.pdf

12. Shelhamer, E., Long, J., Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (4), 640–651. doi: https://doi.org/10.1109/tpami.2016.2572683

13. Semantic Segmentation with SegNet. Available at: https://github.com/dusty-nv/jetson-inference/blob/master/docs/segnet-dataset.md

14. Saxena, S. (2021). Introduction to The Architecture of Alexnet. Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2021/03/introduction-to-the-architecture-of-alexnet/

15. How to do Semantic Segmentation using Deep learning. Available at: https://nanonets.com/blog/how-to-do-semantic-segmentation-using-deep-learning

16. Kingma, D. P., Ba, J. (2015). Adam: a method for stochastic optimization. arXiv.org. Available at: https://arxiv.org/pdf/1412.6980.pdf

17. Solovey, D. (2017). Terrain Classification from High Resolution Aerial. Images Using Deep Learning. Available at: https://fenix.tecnico.ulisboa.pt/downloadFile/1689244997257919/Resumo_174132.pdf

18. Audebert, N., Le Saux, B., Lefèvre, S. (2017). Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images. Remote Sensing, 9 (4), 368. doi: https://doi.org/10.3390/rs9040368

19. Aeroscapes. Aerial Semantic Segmentation Benchmark. Available at: https://github.com/ishann/aeroscapes