Segmentation of Satellite Images of the Earth's surface using Neural Network Technologies

Stanislav Bagaev, Elena Medvedeva
Vyatka State University, VyatSU
Kirov, Russia
hedgehog_rus1@mail.ru, emedv@mail.ru

***Abstract*—The authors present an algorithm for segmentation of satellite images based on the U-net convolutional neural network. It includes the following stages: preparation of input data, modification of the structure of a convolutional neural network (CNN) taking into account the parameters of the studied images, training CNN on the formed training sample, and segmentation of test images. The authors suggest procedures for optimizing the CNN training time and increasing the accuracy of the selected classes of objects with a limited training sample of images and computing resources. The results of experimental studies that confirm the effectiveness of the method are presented.**

I. INTRODUCTION

Segmentation of high spatial resolution images obtained from satellites during Earth remote sensing (ERS) is an effective tool for agricultural and environmental monitoring of the Earth's surface in order to prevent emergencies [1].

There are various approaches to the segmentation of satellite images and their use can speed up the process of identifying areas of interest. As a rule, we use methods based on the selection of contours to detect artificial structures, roads and rivers on images [1,2]. And we use methods based on the extraction of texture features in order to select areas that do not have clearly defined boundaries and are devoid of essential details, such as forests, fields, settlements [3-11]. Many well-known segmentation methods [1-11] allow to select objects of interest of a limited class with a vector of features of small dimension. To improve the accuracy of identifying areas of interest, a combination of different segmentation methods or a larger set of features are used. It complicates data processing. At the same time, the growing volumes of ERS data require increased processing speeds and segmentation accuracy, especially with a large group of objects.

A promising approach to the segmentation of regions of interest in satellite images is the approach based on convolutional neural networks (CNN) [12-17].

The main requirement for solving the segmentation problem is to ensure high accuracy of the selected areas. The accuracy of the segmented areas is determined by the volume of the training sample, which should be as close as possible to real data.

If it is necessary to select objects related to another biome on satellite images, it becomes necessary to retrain the network on new images. Therefore, another important task is to optimize the training time of the neural network, as well as the possibility of solving this problem with limited computing resources.

The article deals with the multiclass problem of segmentation of multispectral satellite images based on CNN, which makes it possible to automate the process of identifying areas of interest with high accuracy.

The article has the following structure. Section II describes the basic segmentation algorithm based on the U-net CNN. Section III lists the procedures for changing the basic algorithm, which can improve the segmentation accuracy by increasing the accuracy of identifying objects of rarer classes with a limited training sample of images and computational resources. Section IV presents the results of experimental studies. Section V contains conclusions.

II. BASIC SEGMENTATION ALGORITHM

Many CNN structures proved themselves in various tasks have appeared recently. The U-net CNN has been chosen as the basis for satellite images, which provides good results in segmentation of objects in biomedical images when using a small amount of data. Moreover, the U-net network has good relative performance and often used for medical images [18, 22-24] and satellite images [25,26].

The basic architecture of the U-net consists of compressing and expanding paths and performs the following procedure: it downsizes the image with some transformations step by step and then reconstructs the predicted masks from the compressed image. In this research, we introduced the following changes to the architecture of the U-net network in order to perform the task of multiclass segmentation:

– the depth of the neural network was increased by adding new layers of compression and pooling to expand the generated feature map;

– the dimension of the input window was changed for image processing 112 × 112, as well as the dimensions of the output layer and the parameters of the remaining layers;

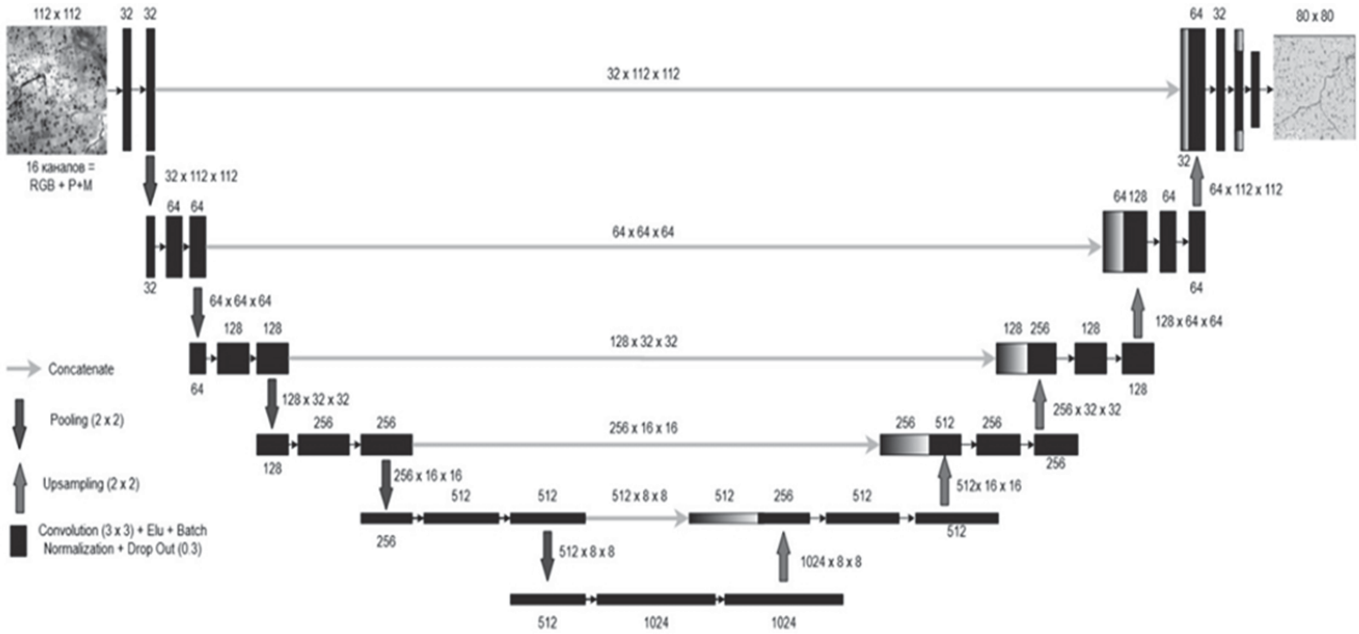
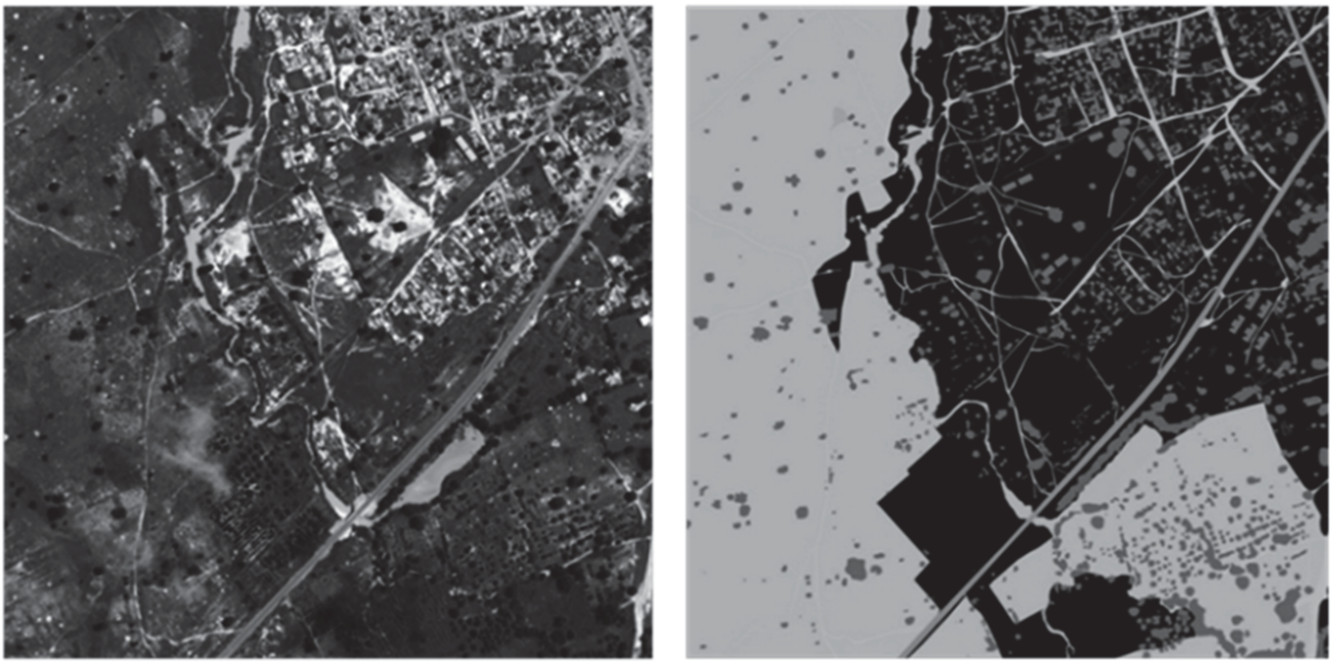


Fig. 1. The structure of the modified CNN



a) b)

Fig. 2. An example of an image from the training sample (a) and markup (b)

– there were added Batch normalization layers after each compression and Dropout after each step of the compression and expansion branches, except for the output fully connected layer;

– as an activation function we used an exponential linear unit - the Elu function, which allowed to increase the efficiency of training, avoiding the problems associated with the appearance of Dead ReLU [20] .

Fig. 1 shows the structure of the modified CNN U-net.

Panchromatic and multispectral images of the Earth's surface were used as input data, taken in 16 ranges from the World View 3 satellite scanners used for the contest "DSTL Satellite Imagery Feature Detection challenge" [21]. The dataset consisted of training and test images. The training dataset used to train the network included 80 different images of areas of the Earth's surface. These images contained objects of the following classes:

- buildings and other artificial structures;
- asphalt and unsurfaced roads;
- trees;
- fields and agricultural land;
- reservoirs and rivers;
- motor transport.

These objects are of significant interest for various fields of activity and often become the desired objects in the tasks of processing satellite images.

At the first stage, there was preliminary training of the CNN. As a training sample, we used 20 images with a size of ~ 3300x3300 pixels, labeled with the appropriate classes. Fig.2 shows an example of an image from the training sample (a) and markup (b). For training, the input images were scaled to one size 3472 x 3472 pixels by padding with zeros at the edges of the image, which were then filled with mirror images of the original image. Then the images were divided into intersecting 112x112 fragments with an overlap of 16 pixels. Fragments of the image were collected in one large packet and fed to the network input. To expand the test sample, we used mirroring procedures vertically and horizontally, and rotating image fragments by an arbitrary angle.

Thus, the network was trained on 13448 samples corresponding to the established threshold values for the area of polygons of the selected classes. For all calculations, we used a personal computer with the following configuration:
NVIDIA GTX 1080Ti 8 GB graphics card;
Ryzen 2600x processor (4.0 GHz) ;
RAM size – 16 Gb;
SSD Samsung EVO 860 – 256 Gb.

To optimize and adjust the weights of the neural network, the built-in Adam optimizer was used. All layers' Dropout parameter was set to 0.2.

The convolutional neural network was trained on these samples for 250 epochs, which on the given computer configuration took about 27 hours.

For the validation process were used 5 images with a size of ~ 3300x3300 pixels. Validation was carried out on 1012 randomly selected frames 112x112 from the test multipolygon.

To assess the outcomes of the neural network, in particular to assess the accuracy of the selection of the required classes on test images, we used the Jaccard Index:

$$J = \frac{TP}{TP + FP + FN},$$

where TP is true positive decision; TN — true negative decision; FN — false negative decision.

The Jaccard Index was averaged separately for each class of objects and for the entire image.

To assess the outcomes of the CNN, we used standard estimates of accuracy and estimates of total losses. Fig. 3 and 4 present the graphs of the respective ratings.

The graph in Fig. 3 shows the convergence of the neural network to unity in accuracy, which indicates the correctness of the network architecture. After the 50th epochs, training slowed down a lot due to the limited number of training images. The classes most often found in the images are well defined by the trained network, which leads to an increase in the overall segmentation accuracy and a decrease in the correction of the neural network weights in each iteration. At the same time, the selection of more rare classes, such as buildings and unsurfaced roads, becomes less accurate and practically does not improve in the future.

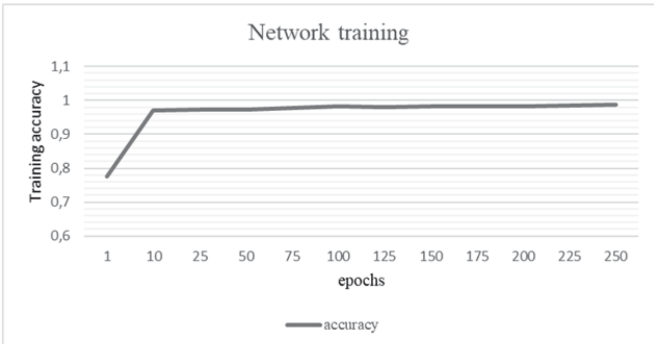


Fig. 3. Training accuracy

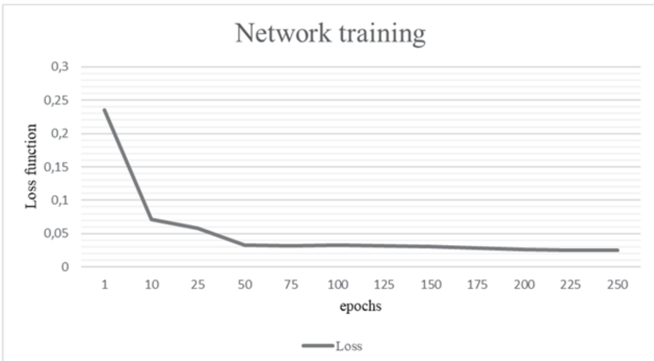


Fig. 4. Training losses

Fig. 5 shows a graph of CNN validation on test images.

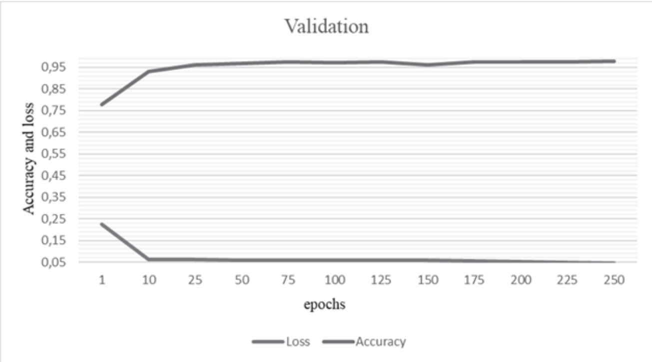


Fig. 5. Validation of weights in images

Table I presents the results of assessing the accuracy of finding objects belonging to different classes, as well as the thresholds for making decisions about the presence of such objects in the image.

TABLE I. RESULTS OF ESTIMATING THE ACCURACY OF THE DISTINGUISHED CLASSES BY THE BASIC ALGORITHM

Class	Jaccard Index	Binarization threshold
Buildings	0.72	0.3
Structures	0.65	0.2
Roads	0.56	0.1
Unsurfaced roads	0.69	0.3
Trees	0.97	0.2
Fields	0.92	0.9
Rivers	0.76	0.2
Reservoirs	0.82	0.9

The results shown in Table I indicate the high accuracy of the identified classes of objects. The most accurate is the selection of objects belonging to the classes "Trees" and "Fields" due to the predominance of representatives of these classes in the test sample during training. Objects of the class "Buildings" and "Structures" have a lower selection accuracy due to the high similarity of objects of this class.

The processing time for one set of satellite images (10 images) up to 3600x3600 in size not from the training sample and the test set was less than 10 seconds. The overall accuracy estimated by the Jaccard Index for different satellite images was in the range 0.76 - 0.86.

III. IMPROVED SEGMENTATION ALGORITHM

The developed algorithm is an effective tool for the segmentation of satellite images. However, the long network training time is a significant drawback in its real use. For any neural network to work effectively, the training dataset should be as close as possible to the data with which the algorithm will work in reality. And when working with data related to a completely different biome, it becomes necessary to train the network using new examples. As a result, optimization of the neural network training time becomes one of the most important tasks. At the same time, it is necessary to ensure the convergence of the neural network and maintain the overall accuracy of calculations, avoiding retraining.

One method to achieve this result is to carefully select the Dropout layer parameter. This layer allows to control the number of active connections between neurons of the layers where it is located. A lower value of the Dropout parameter leads to a low efficiency of this layer. A higher value leads to a decrease in the overall accuracy of the algorithm operation due to too few active neurons to distinguish class feature maps and violation of the convergence of the network. The study shows that the optimal value for the developed algorithm is 0.3.

The choice of a training algorithm can significantly reduce the network training time. In this research we tested the following algorithms: adaptive gradient algorithm (AdaGrad), natural gradient descent (kSGD), adaptive method for estimating moments (Adam), adaptive method for estimating moments with Nesterov's loop (NAdam). The results of testing the algorithms are presented in Table II. Based on testing, the NAdam method was chosen as the training algorithm. For the same number of epochs, this method with long training (200 epochs) made it possible to achieve the highest accuracy of image processing, and with a short training period (50 epochs) it's better to use the Adam method.

TABLE II. RESULTS OF ESTIMATING THE DEPENDENCY OF ACCURACY OF IMAGE PROCESSING FROM LEARNING METHOD (LEARNING RATE IS SET TO 0.01)

Epochs	Learning method			
	AdaGrad	Adam	Nadam	kSGD
50	0.64	0.66	0.64	0.55
100	0.72	0.74	0.76	0.68
200	0.75	0.76	0.79	0.74

Other important parameters that affect the training rate of a network are the number of epochs and the size of the packets. The influence of batch size of the training rate of different Learning methods is shown in Table III. The increase of batch size during training allows to achieve higher training efficiency at same training time.

Although an increase in the packet leads to an increase in the number of test images arriving at the input of the network in each iteration and, as a consequence, to a significant increase in computing power. It leads to the impossibility of using the algorithm on weak computers due to the limitation in the amount of available memory. An increase in the number of epochs does not require an increase in computing power, but it significantly increases the total training time.

TABLE III. RESULTS OF ESTIMATING THE EFFICIENCY OF LEARNING PROCESS FROM USED BATCH SIZE DURING TRAINING (LEARNING RATE IS SET TO 0.01, EPOCHS 50)

Batch size	**Learning method**			
	AdaGrad	**Adam**	**Nadam**	**kSGD**
16	0.654	0.674	0.673	0.664
32	0.643	0.681	0.676	0.653
64	0.638	0.675	0.675	0.652

Another important parameter is the network training rate factor. A low rate factor allows to stably reduce the parameters of training losses, but requires a significant number of iterations and greatly slows down the training process. Too big value does not give the network convergence, or leads to retraining, and as a result does not allow to achieve high accuracy, but allows to train many times faster network. The most optimal solution is to use an adaptive training rate that reduces the rate depending on the number of epochs passed.

Table IV presents the results of the effectiveness of various coefficients of network training when using the NAdam method.

TABLE IV. TRAINING FACTORS (NADAM OPTIMIZER)

Initial network assesment (Jaccard), %	**Training rate**	**Losses**	**Accuracy**	**Jaccard Index**	**Epochs**
0	0.1	0.033	0.9805	0.66	50
	0.01	0.034	0.9711	0.62	50
	0.001	0.028	0.9653	0.54	50
52	0.1	0,032	0.9800	0.70	50
	0.01	0.033	0.9811	0.74	50
	0.001	0.027	0.9793	0.73	50

The data obtained confirm that at the beginning of training the neural network, it is advisable to use high training coefficients, gradually decreasing the coefficient with an increase in the number of passed epochs.

The considered procedures for changing the basic algorithm make it possible to reduce the training time of the developed CNN by increasing the overall training rate and, therefore, the need to use a smaller number of epochs to achieve a similar result while maintaining the overall accuracy of object classification.

The increase of the overall accuracy of the segmentation requires increasing the accuracy of the selection of objects of rarer classes. If it is impossible to increase hardware resources and the lack of images with a uniform statistical distribution of classes in the training sample, a reasonable approach is to use separately trained networks with each specific class, with the subsequent combination of the results. This is due to the fact that when using images with an uneven statistical distribution of classes when training a network, the neural network is oriented towards the selection of the most common classes. Therefore, training a separate copy of the neural network for each individual class allows to achieve a higher accuracy of the final segmentation accuracy.

With this approach, the total training time of the network increases n times (n is the number of allocated classes), but the need for overall quality (in terms of the distribution of object classes in images) and the volume of required training images decreases.

IV. EXPERIMENTAL RESULTS

Table V shows the results of evaluating the accuracy of the selected classes of objects by the improved algorithm. The results were obtained under the following conditions: NAdam optimizer (lr 0.1 to 50 epochs, 0.01 to 100, then 0.001), Dropout (0.3).

TABLE V. RESULTS OF EVALUATING THE ACCURACY OF THE SELECTED CLASSES OF OBJECTS BY THE IMPROVED ALGORITHM

Class	**Jaccard Index**	**Training epochs**
Buildings	0.84	184
Structures	0.82	184
Roads	0.79	184
Unsurfaced roads	0.75	219
Trees	0.95	200
Feilds	0.96	200
Rivers	0.89	234
Reservoirs	0.84	243

Fig. 6 shows:

a) the original image fed to the network input;

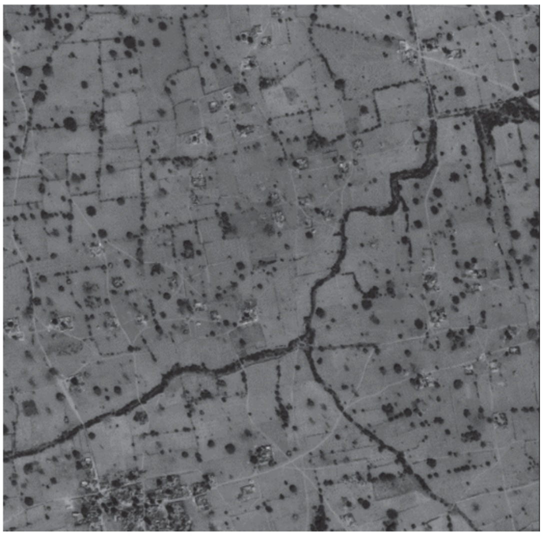
b) the result of segmentation by the basic algorithm;

c) the result of segmentation by the improved algorithm.

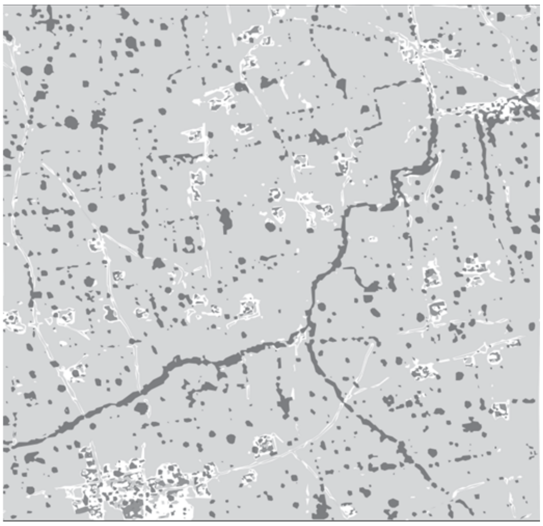
A comparison of the results given in Tables I and III and in Fig. 6 shows a gain in the accuracy of the selected classes of objects. The largest gains are observed for objects of similar classes (unsurfaced and asphalt roads, buildings and structures), which ranges from 9 to 41%. The overall accuracy estimate by the Jaccard Index for different satellite images was 0.86.

Limitations and directions for further research can be summarized as follows.

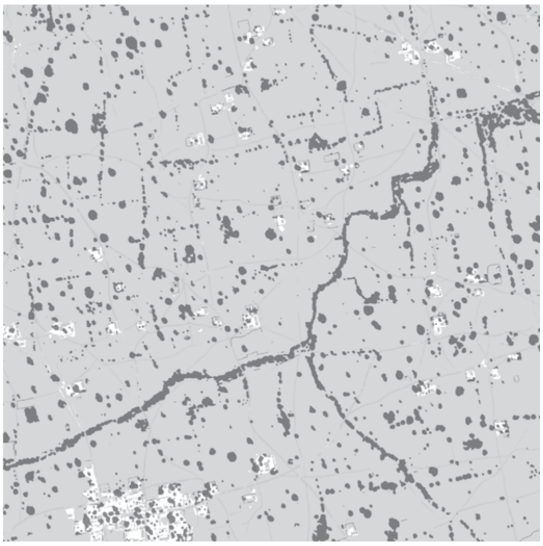
While the model can be expanded with additional layers, to enable learning a richer set of relevant feature maps, this stands technically challenging due to the increasing requirements for GPU RAM memory for such alterations. It



a)



b)



c)

Fig. 6. Satellite image segmentation results

was also impossible to increase the batch size during trainig or to increase input spatial dimensions (for example using frames of 144x144 as input) cause of lack of RAM.

The size of the filter kernels can also be varied to increase the filter size, but can be difficult to optimize, since it connection to the receptive field size and therefore also the minimum input size for the network, which in the end also depends from the available RAM.

It should be noted that large areas of a homogeneous surface present some problem for the algorithm. Due to the fact that segmentation occurs by dividing one large image into many smaller crop areas, when processing homogeneous areas, sometimes there are cases of incorrect selection of the area, due to the lack of areas characteristic of the joints of two classes, which leads to the fact that the algorithm cannot determine the class of a homogeneous area and selects only the boundary value along the edge of the crop, as belonging to this class. Further improvement of accuracy can be achieved either by more careful adjustment of the algorithm for each of the classes (for example, the ability to change the size of the input images.

The Data set used in training and validation, was also an important factor in the final results. The heterogeneity of class distribution both in single image and between different images of the Data set leads to reconfiguring the network for more frequent classes, which makes training of a single network for multiclass segmentation very difficult This leads to necessity of training special networks for each class. The more direct way will be thorough preparation of training set frames with similar class distribution for training. Both ways lead to the loss of time efficiency when dealing with need of retraining the network for new task, which is not the desirable result in most cases. Some alteration of the network maybe can lead to the less dependants on the prepared DataSet static.

V. Conclusion

The research proposes basic and improved algorithms for segmentation of satellite images based on the U-net CNN. Algorithms allow to classify objects into ten classes.

The estimation of the accuracy of the segmented areas using the Jaccard Index by the basic algorithm was 0.84. Average training time for each epoch on a test configuration was 20 minutes. The processing time for one image was about 10 seconds.

The improved algorithm made it possible to increase the segmentation accuracy by increasing the accuracy of selecting objects of rarer classes with a limited training sample of images and computing resources. For objects of different classes, the gain in the accuracy of object selection ranged from 1 to 41%. The processing time of one image, taking into account the separate recognition of each class and the combination of the results, was about 100 sec.

REFERENCES

[1] R. A. Schovengerdt, *Remote sensing. Models and methods of image processing*. M .: Technosphere, 2010.

[2] R. Gonzalez and R. Woods, *Digital Image Processing*. M.: Technosphere, 2012.

[3] E. Medvedeva and A. Evdokimova, "Detection of Texture Objects on Multichannel Images", *in 24th Conf.of Open Innovations Association (FRUCT)*, 2019, pp. 249 – 254

[4] E. Medvedeva, I. Trubin and E. Kurbatova, "Methods of filtration and texture segmentation of multicomponent images" *Intelligent Systems Reference Library.* Springer International Publishing Switzerland, vol. 135, 2018, pp. 97-118.

[5] M. Li, S. Zang, B.Zhang, S. Li and C.Wu, "A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information", *European Journal of Remote Sensing*, vol. 47, 2014, pp. 389-411.

[6] F. Borne, "Texture-based classification for characterizing regions on remote sensing images", *Journal of Applied Remote Sensing*, vol. 11 (3), 2017.

[7] J. Yuan and D.L. Wang, "Remote sensing image segmentation by combining spectral and texture features", *in IEEE Transactions on geoscience and remote sensing*, vol. 52 (1), 2014, pp. 16-24.

[8] Y. Han, H. Kim, J. Choi and Y. Kim, "A shape-size index extraction for classification of high resolution multispectral satellite images", *International Journal of Remote Sensing*, vol. 33, no. 6, 2012. pp. 1682–1700.

[9] G. Fu, H. Zhao, C. Li and L. Shi, "Segmentation for High-Resolution Optical Remote Sensing Imagery Using Improved Quadtree and Region Adjacency Graph Technique", *Remote Sensing*, no. 5, 2013, pp. 3259–3279.

[10] E. Kurbatova and E. Medvedeva, "A method for texture segmentation of the multidimensional images" *in 2018 7th Mediterranean Conference on Embedded Computing, Proceedings*, 2018, pp. 1-4.

[11] E. V. Medvedeva, E. E. Kurbatova and A. A. Okulova, "Textural segmentation of noisy images of the Earth's surface", *Modern problems of remote sensing of the Earth from space*, vol. 14, no. 7, 2017, pp. 20–28.

[12] A. Garcia-Pedrero, C. Gonzalo-Martin and M. Lillo-Saavedra, "A machine learning approach for agricultural parcel delineation through agglomerative segmentation", *International journal of remote sensing*, vol. 38(7), 2017, pp. 1809-1819.

[13] V. Iglovikov, S. Mushinskiy and V. Osin, "Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition", https://arxiv.org/pdf/1706.06169.pdf.

[14] G.M. Musyoka, "Automatic delineation of small holder agricultural field boundaries using fully convolutional networks", University of Twente, 2018.

[15] L. Jonathan, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

[16] F. Zhang, B. Du, L. Zhang and M. Xu, "Weakly Supervised Learning Based on Coupled Convolutional Neural Networks for Aircraft Detection", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, 2016, pp. 5553–5563.

[17] Y.Zhang, F. Fei, L. Zhang, "Large patch convolutional neural networks for the scene classification of high spatial resolution imagery", *Journal of Applied Remote Sensing*, vol. 10, no. 2, 2016, pp. 025006–025006.

[18] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", *in International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp.234-241.

[19] R. Olaf, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation" *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*. Springer International Publishing, 2015.

[20] Activation Functions Explained - GELU, SELU, ELU, ReLU and more// URL: https://mlfromscratch.com/activation-functions-explained/#/.

[21] Dstl Satellite Imagery Feature Detection// URL: kaggle.com/c/dstl-satellite-imagery-feature-detection/.

[22] C. E. Akbaş, M. Kozubek, "Condensed U-Net (CU-Net): An Improved U-Net Architecture for Cell Segmentation Powered by 4x4 Max-Pooling Layers", *in IEEE 17th International Symposium on Biomedical Imaging*, Iowa: IEEE, 2020, pp. 446-450.

[23] S. Bell, A. Zsom, J. Conley, D. Spade, "Automated identification of multinucleated germ cells with U-Net", PLOS ONE 15: e0229967. 10.1371/journal.pone.0229967.

[24] T. Overton and A. Tucker, "DO-U-Net for Segmentation and Counting: Applications to Satellite and Medical Images", *in book: Advances in Intelligent Data Analysis XVIII*, 2020, pp.391-403.

[25] Y. Guo, X. Cao, B. Liu and M. Gao, "Cloud Detection for Satellite Imagery Using Attention-Based U-Net Convolutional Neural Network", Symmetry 2020, 12(6), 1056. 10.3390/sym12061056.

[26] P. Gudzius,O. Kurasova, E, Filatovas, "Optimal U-Net Architecture for Object Recognition Problems in Multispectral Satellite Imagery", *in IEEE/ACS 16th International Conference on Computer Systems and Applications*, 2019. 1-2. 10.1109/AICCSA47632.2019.9035305.