Image Semantic Segmentation Based on High-Resolution Networks for Monitoring Agricultural Vegetation

Valentin Ganchenko, Valery Starovoitov
United Institute of Informatics Problems,
Minsk, Belarus
e-mail: ganchenko@lsi.bas-net.by

Xiangtao Zheng
Xi'an Institute of Optics and Precision Mechanics,
Shaanxi, China
e-mail: xiangtaoz@gmail.com

Abstract – In the article, recognition of state of agricultural vegetation from aerial photographs at various spatial resolutions was considered. Proposed approach is based on a semantic segmentation using convolutional neural networks. Two variants of High-Resolution network architecture (HRNet) are described and used. These neural networks were trained and applied to aerial images of agricultural fields. In our experiments, accuracy of four land classes recognition (soil, healthy vegetation, diseased vegetation and other objects) was about 93-94%.

Keywords – convolutional neural network, semantic segmentation, aerial photograph, agricultural vegetation.

I. INTRODUCTION

The foundation of precision farming reliable and operative updated information of state of cultivated vegetation and soil. The most practical way to obtain such information is remote sensing. Remote sensing allows obtaining data as quickly as possible and with different spatial resolution. Such information can be the basis for vegetation monitoring and allows the user find areas affected by some diseases. Data for remote sensing of lands are usually recorded in two main representations – optical and spectrometric [1–7]. These approaches define different analysis algorithms and data acquisition equipment. In this paper, we have focused on optical data processing due to the less online availability of multispectral image data.

One of the most available remote sensing data acquisition tools in agriculture are unmanned aerial vehicles (UAVs). In comparison with satellites UAVs are cheaper and more efficient [8–9]. Usage of images with GPS data allows us to prepare and visualize data about large and small crop areas. The data can be stored in GIS database and used for decision making based on processing and analysis.

Usage of artificial convolutional neural networks (CNNs) is a popular approach in agricultural remote sensing data analysis. The networks may be used to solve various problems of precision farming [10]. For example, in [11–13] authors use CNNs to detect weeds with resulting accuracy obove 90%. An UAV is used for data collection and networks are used for semantic segmentation of images with object classification. The authors of [14] use residual CNN to detect flowers for the yield prediction task. This approach gives accuracy from 67 to 94%. It depends on the vegetation. In addition, a multilayer perceptron is used to predict yield, as shown by the authors of [15]. Their research is based on detection of growing fruits. In some case, more complicated neural architectures are required. CNNs can be used not only to assess the state of individual plants, but also to assess vegetation cover. For example, in [16] the authors analyzed Gaofen-2 satellite imagery to categorize vegetation with accuracy of 89-90%. Authors proposed two-leveled architecture. The first level consists of two convolutional kernels sets and is used for separation of farmlands and woodlands. The second one contains two coders for encoding of nonlinear features, which can be associated with vegetation category. Additionally, the authors of [17] used CNN-based semantic segmentation for thematic mapping of agricultural fields.

One of the neural architectures that can be applied for remote sensing data processing is called High-Resolution networks (HRNet). It may be used for semantic segmentation. In [18] an application of this architecture for urban image processing is described.

Our study is aimed at recognizing the affected areas of vegetation.It is based on detection of vegetation, the state of which has changed under the influence of disease. We tested two variants of the HRNet architecture for semantic segmentation of color images of remote sensing of crops.

The results presented in the paper is a part of our study of the usege of the HRNet architecture for semantic segmentation of aerial imagery based on SegNet and U-Net architectures [19]. An algorithm for digital color image processing of various spatial resolutions was obtained. The task of classification the identified disease was not set at this stage of the study.

II. PROBLEM FORMULATION

The goal of this study is search of a conversion $A: I_{orig} \rightarrow I_{result}$, which form a map image I_{result} from the original remote sensing image of an agricultural vegetation field I_{orig}. Each pixel of $I_{orig}(x, y)$ is a point in RGB space and every element in $I_{result}(x, y)$ contains the number of one of the four predefined classes ("soil", "healthy vegetation", "diseased vegetation", and "other objects").

Input data for research are two sets of images: images of lone plants (for investigation of disease influence) and remote sensing imagery of experimental field of potato. Remote sensing images were photographed at heights of 5, 15, 50 and 100 meters [20–21]. The corners of the experimental field were marked by four square marks with sides in one meter and two intersecting black lines 20 cm wide (fig. 1). These labels can be used to calculate the spatial resolution of images.

Lonely plants were divided into three groups:
- infected with alternaria disease;
- infected with erwinia disease;
- control group of healthy plants.

Images of lone plants were taken daily at 8, 10, 12, 14 and 16 hours for 8 days in July.

In diseased leaves, chlorophyll is destroyed and the color of plants is changed. This fact was used in image analysis. In addition, the available images had a problem with sun's glare the on the leaves. These highlights add yellow color and these leaves can be recognized as affected by early disease.

978-1-7281-7628-4/20/$31.00 ©2020 IEEE
DOI 10.1109/SYNASC51798.2020.00050

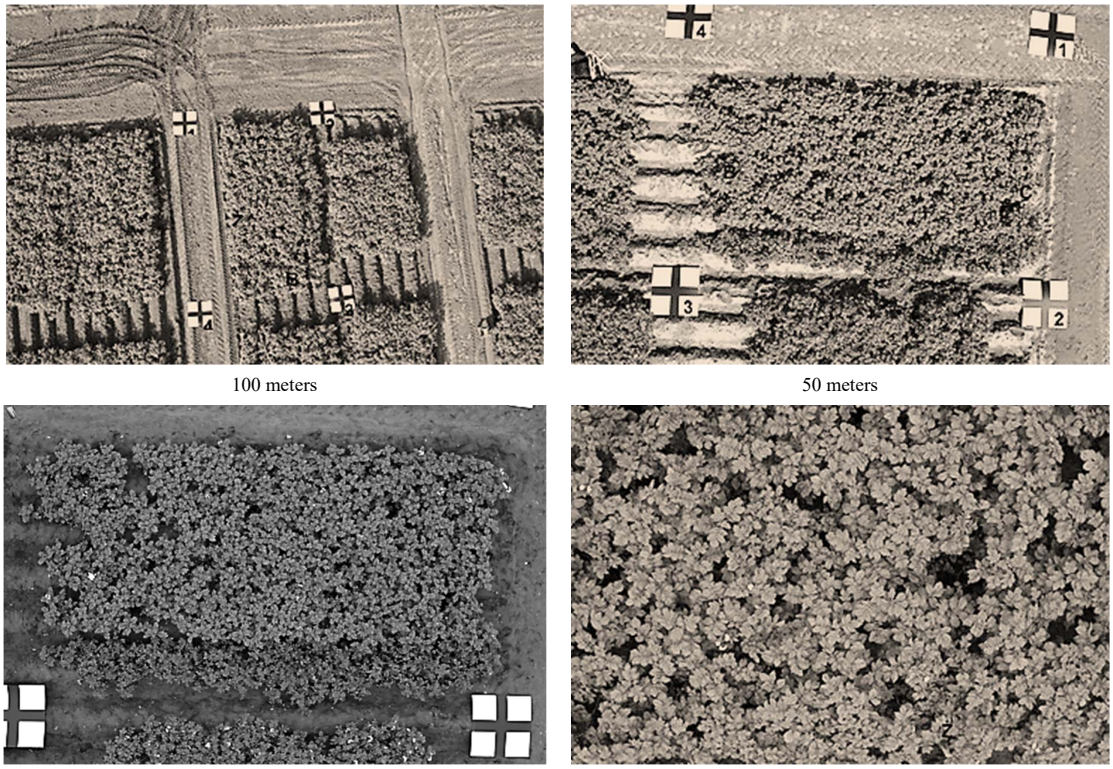


Fig. 1. Original remote sensing images samples

Analysis of color features of various images types shown significant difference between vegetation and soil images. However, color features of diseased and healthy leaves are not so significant and more appreciable in blue color channel. If color features would be represented in form of histogram for color channels than differences in shapes and peaks for vegetation and soil features would be visible for each channel. However, for healthy and diseased vegetation histograms would differ in shape only.

We have observed that these color features cannot be used directly due to the mixing of objects of different classes. This mixing results in a number of color distortions of the features. These distortions significantly decrease similarity reduce the similarity of the color characteristics of object classes. Therefore, color features are not enough for classification, and in addition to color, it is necessary to analyze information about the structure. HRNet can be used to combine color and structural features.

III. Data preparing for training and validation

Training and validation sets were formed from the original high-resolution images. Our network consists of layers containing matrices of 256×256 elements, so the images were cut into intersecting fragments of the same size in pixels. The training set was formed from such fragments and their corresponding class labels, and was also supplemented with reflections and rotations of all fragments. Expert-created class masks can be represented as images containing the corresponding color values of the classes: soil or 0, healthy vegetation or 1, diseased vegetation or 2 and other objects or 3.

IV. HRNet without streams connections

We have proposed a HRNet-based CNN architecture without inter-thread connections [22, 23] (we call it A1 in the text below). It is presented in Figure 2. It segments images into four types of areas: soil, healthy vegetation, diseased vegetation, and other objects. In Fig. 2 MaxPooling layers halve the input matrix size, but UpSampling doubles the input matrix size. The following network details were empirically defined:

- Input layer size: $256 \times 256 \times 3$ (as color image).
- Output convolutional layer: activation function – sigmoid, output layer size – $256 \times 256 \times 4$.

The reasons for choosing these architecture parameters: with a decrease in the number of filters and / or convolutional layers, we got a lower accuracy and metrics F1 (described below); as the number of filters and / or convolutional layers were increased, the overall accuracy was higher, but the F_1 scores and accuracy for some classes decreased due to imbalanced data.

The categorical cross-entropy was chosen as the loss function [24].

Details of the training phase are:

- Training algorithm: Adam [25] with parameters: $\mathrm{lr} = 0.0001, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 \times 10^{-8}$, $decay = 0$ [26].
- Training set size: 10000 images.
- The size of the validation set: 2000 images.
- Accuracy for the validation set: 93.84 %.

V. HRNET WITH STREAMS CONNECTIONS

The second CNN architecture called below A_2 is based on HRNet with connections between streams with different resolution (Figure 3). The segmenter is a CNN that divides an image into four types of areas: soil, healthy vegetation, diseased vegetation, and other objects.

This architecture differs from HRNet without streams connections in that there are additional connections between high-to-low resolution convolution streams. halve the size of the input matrix, and UpSampling doubles the size of the input matrix. The following network parameters were selected empirically:

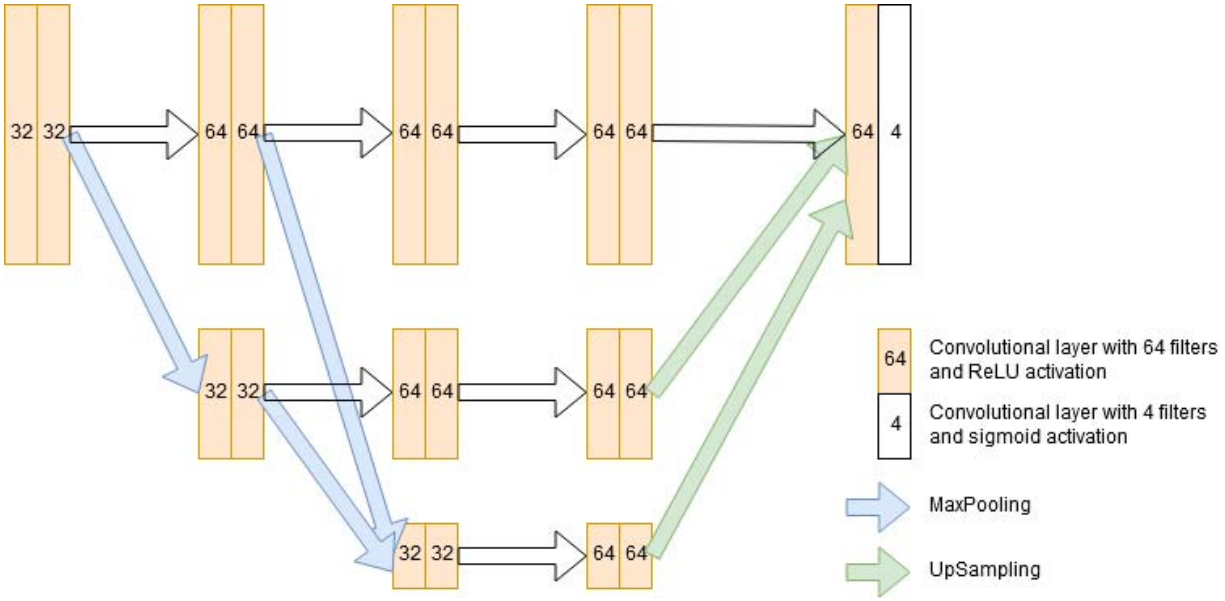


Fig. 2. Implemented HRNet architecture without connections between streams

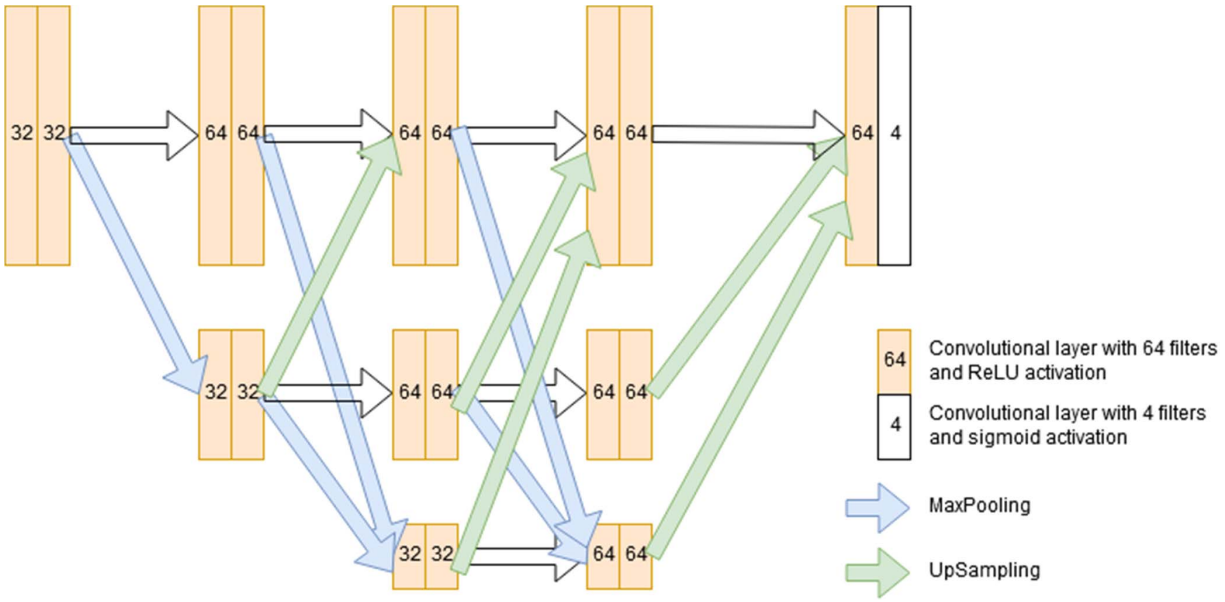


Fig. 3. Implemented HRNet architecture without connections between streams

- The input layer size is $256 \times 256 \times 3$ (as a color image).

- In the output convolutional layer activation function is sigmoid, the output layer size is $256 \times 256 \times 4$.

Reasons for choosing these architecture parameters: with a decrease in the number of filters and / or convolutional layers we got a decrease in the accuracy and metric F_1; as the number of filters and/or convolutional layers increased, the overall accuracy was increased, but the F_1 scores and accuracy for some classes were decreased due to imbalanced data.

The categorical cross-entropy also was chosen as the loss function.

Details of the training phase are the same:

- Training algorithm: Adam with parameters: $\mathrm{lr} = 0.0001, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 \times 10^{-8}, \ decay = 0$.
- Training set size: 10000 images.
- The size of the validation set: 2000 images.

We obtained accuracy for the validation set of 93.9%.

VI. Output data structure

These networks return on output matrix with sizes 256×256×4 elements. The first two dimensions correspond to sizes of the input image slice, and the third one to number of sought classes: "soil", "healthy vegetation", "diseased vegetation" and "other objects". Values of this matrix contain probabilities of input image pixel belonging to corresponding class. Additionally, these values can be normalized. As result, we can obtain a fuzzy value of classes membership for corresponding pixels.

VII. Segmentation algorithm

Shortly our segmentation algorithm (transformation $A: I_{orig} \rightarrow I_{result}$) can be represented as follows:

1. Load an original color image I_{orig} .
2. Divide I_{orig} into parts $O_i(I_{orig})$ with size 256×256 pixels. For each part do:
   2.1. Copy the part as a color image $B_i(I_{orig})$.
   2.2. Transform $B_i(I_{orig})$ by segmenter $A \in \{A_1, A_2\}$ into matrix $Segm_A$ with the size of $256 \times 256 \times 4$.
   2.3. Calculate the class index for every pixel (x, y) of the image $B_i(I_{orig})$: $x \in [0,255], y \in [0,255]$:
   $$index = argmax([A(x,y)]),$$
   where $Segm_A(x,y)$ is a 4-component vector containing the probabilities of belonging to the predefined classes in the original image $O_i(I_{orig})$.
   2.4. Generate a pseudo-color output image ($I_{result}(O_I)$). Choose the color for every pixel (x, y) as follows: black for soil, dark-gray for healthy vegetation, light-gray for diseased vegetation, and white for others.
3. Store the obtained I_{result} as the segmentation map.

VIII. Testing

The segmenters described above were tested on a validation set. For each class, the segmentation accuracy was calculated, as well as for input image as a whole. The obtained estimates of accuracy in percent are presented in Table. 1.

Table I. Segmenters test results

Classes	Accuracy, %	
	HRNet A1	***HRNet A2***
Soil	86.1	88.87
Healthy vegetation	98.25	97.04
Diseased vegetation	64.36	72.16
Other objects	87.06	87.97
Average	**93.84**	**93.9**

The imbalance in different classes requires additional assessments of the classification results, which are presented by the confusion matrices in Table 2. The values in the matrices are specified as the ratio of the number of pixels belonging to a class to the total number of pixels of all classes in the sample. The slight difference obtained between the two architectures described above requires further research. Maybe the features in some stream have a more significant effect than connections between streams. In this case, it would be more efficient to use a separate stream.

For a comprehensive assessment of quality of the classification, we calculated following scores: precision, recall and F_1-score, where TP means True Positives count, FP – False Positives count, FN – False Negatives count [27]:

$$Precision = \frac{TP}{TP+FP}, \ Recall = \frac{TP}{TP+FN},$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall},$$

Table II. Two confusion matrixes

Predicted classes	Real classes			
	HRNet A1			
	Soil	***Healthy***	***Diseased***	***Others***
Soil	11.08	0.62	0.24	0.02
Healthy	1.57	77.06	2.70	0.04
Diseased	0.21	0.73	5.30	0.00
Other objects	0.02	0.03	0.00	0.41
	HRNet A2			
Soil	11.44	1.00	0.28	0.02
Healthy	1.16	76.11	2.02	0.03
Diseased	0.26	1.29	5.94	0.00
Other objects	0.01	0.03	0.00	0.41

The values of these indicators are collected in Table 3.

Most of the classification errors occurred in regions containing elements from several classes. For example, near borders of soil and vegetation without significant disease damage. Also, for high resolution images, often a small area of one class may be surrounded by another class. Plant shadows can also introduce errors.

Table III. Precision, recall and F1 for A_1 and A_2

Classes	HRNet A1		
	Precision	***Recall***	***F1***
Soil	0.93	0.86	0.89
Healthy	0.95	0.98	0.96
Diseased	0.85	0.64	0.73
Others	0.91	0.87	0.89
	HRNet A2		
Soil	0.90	0.89	0.89
Healthy	0.96	0.97	0.96
Diseased	0.79	0.72	0.76
Others	0.90	0.88	0.89

Estimates of classification errors for individual classes are presented in Table 4. As can be seen from this table, in the case of identifying soil as vegetation (along boundaries of vegetation and soil), many errors arise. However, the most significant problem is associated with identification of affected vegetation as healthy at site when the disease is in initial stage.

Figures 4 and 5 present an original image and the masks of the found classes.

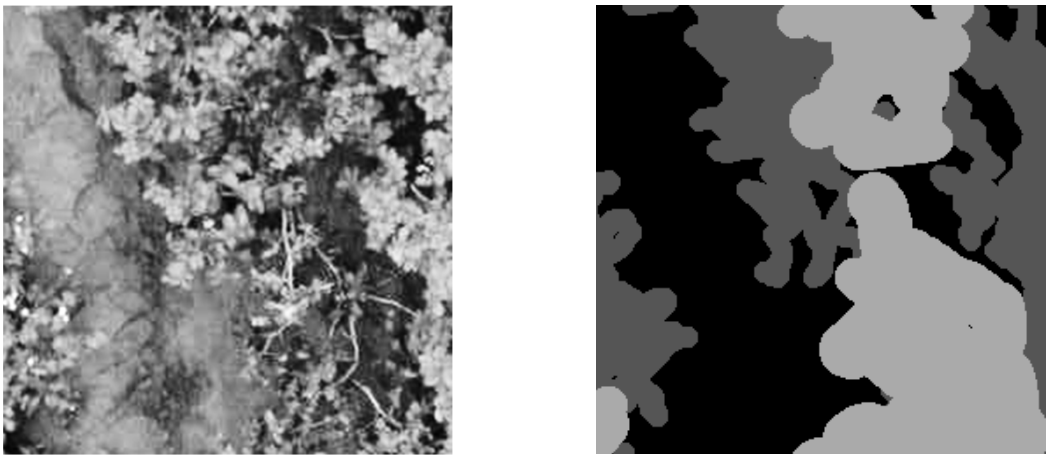
Table 1. Error estimation

Predicted classes	Error, %			
	HRNet A_1			
	Soil	***Healthy***	***Diseased***	***Others***
Soil	–	0.78	2.9	4.75
Healthy	12.18	–	32.74	8.18
Diseased	1.6	0.94	–	0.01
Other objects	0.12	0.03	0	–
	HRNet A_2			
Soil	–	1,28	3,35	5,31
Healthy	8,98	–	24,5	6,71
Diseased	2,03	1,65	–	0,01
Other objects	0,11	0,04	0	–

CONCLUSIONS

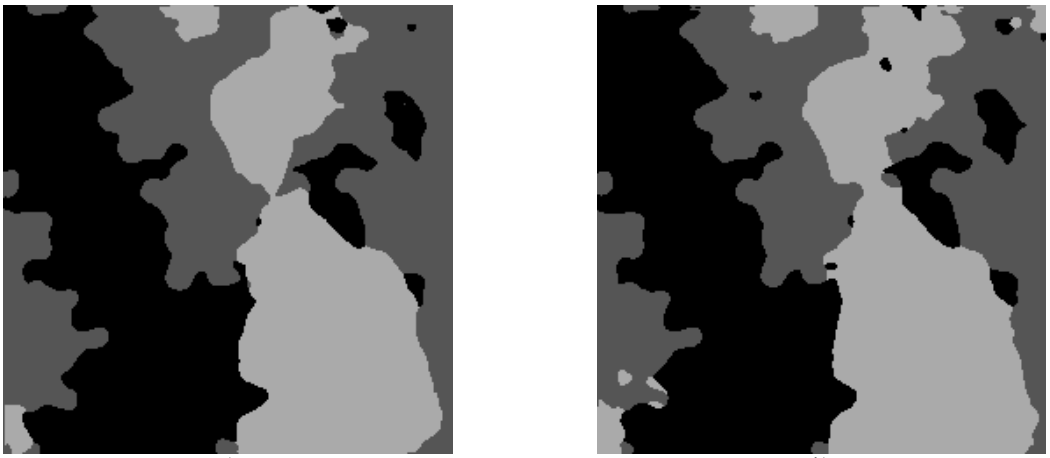
The goal of our research is classification of remote sensing imagery into four classes: “soil”, “healthy vegetation”, “diseased vegetation” and “other objects". We proposed and compared implementation of two variants of neural network for this task. The soft is based on the Keras library. The networks are based on HRNet architecture. We achieved the overall accuracy about 93-94%. However, it should be noticed significant errors in areas with early disease stage. This fact may limit practical usability of the obtained results.

One of directions of further researches is misclassification error decreasing.



a) b)

Fig. 4. Example of an original aerial image (a) and its ground truth segmented map (b)



a) b)

Fig. 5. Map from HRNet A1 (a) and map HRNet A2 (b)

REFERENCES

[1] B.I. Belyayev, L.V. Katkovskiy *Optical remote sensing*. Minsk: BSU, 2006. [In russian]

[2] R. A. Schowengerdt *Remote Sensing. Models and Methods for Image Processing*, 3rd Edition/ Academic Press, 2007.

[3] K. Chao, Y.R. Chen. M.S. Kim “Machine vision technology for agricultural applications”, *Elsevier science transactions on computers and electronics in agriculture*, vol. 36, pp. 173-191, 2002.

[4] N. Kumar [et al.] “Do leaf surface' characteristics affect agrobaeterium infection in tea camellia sinensis (1.)”, *J. Biosci*, vol. 29, No 3, pp. 309-317, 2004.

[5] L. Wu [et al.] "Identification of weed, corn using BP network based on wavelet features and fractal dimension", *Scientific Research and Essay*, vol. 4 (11), pp. 1194-1400, Nov 2009.

[6] Zh. Qin, M. Zhang "Detection of rice sheath blight for in-season disease management using multispectral remote sensing", *International Journal of Applied Earth Observation and Geoinformation*, vol. 7, pp. 115-148, 2005.

[7] S. Aksoy, H.G. Akcay, T. Wassenaar "Automatic mapping of linear woody vegetation features in agricultural landscapes using very high-resolution imagery", *IEEE Transactions on Geoscience and Remote Sensing*, No 48 (1, 2), pp. 511-522, Jan 2010.

[8] H. S. Abdullahi, O. M. Zubair "Advances of image processing in Precision Agriculture: Using deep learning convolution neural network for soil nutrient classification", *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*, vol. 4 Issue 8, pp. 7981-7987, Aug 2017.

[9] D. Wright, V. Rasmussen, R. Ramsey, D. Baker, and J. Ellsworth "Canopy Reflectance Estimation of Wheat Nitrogen Content for Grain Protein Management" *GIScience Remote Sens*, vol. 41, No. 4, pp. 287–300, 2004.

[10] K. A. Mate G Pooja, Singh R Kavita, "Feature Extraction Algorithm for Estimation of Agriculture Acreage from Remote Sensing Images", *Proc. of World Conference on Futuristic Trends in Research and Innovation for Social Welfare*, pp. 5–9, 2016.

[11] Huang H., Deng J., Lan Y., Yang A., Deng X., Zhang L. "A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery", PLoS ONE, 13(4): e0196302, 2018.

[12] Inkyu Sa, Zetao Chen, Marija Popovic, Raghav Khanna, Frank Liebisch, Juan Nieto, Roland Siegwa, "weedNet: Dense Semantic Weed Classification Using Multispectral Images and MAV for Smart Farming", *IEEE Robotics and Automation Letters*, vol. 3(1), pp. 588-595, 2018.

[13] Potena C., Nardi D., Pretto A. "Fast and Accurate Crop and Weed Identification with Summarized Train Sets for Precision Agriculture" / *IAS 2016: Intelligent Autonomous Systems 14*, pp. 105-121, 2017.

[14] Philipe A. Dias, Amy Tabb, Henry Medeiros "Multispecies fruit flower detection using a refined semantic segmentation network", *IEEE Robotics and Automation Letters*, vol. 3, issue 4, pp. 3003-3010, 2018.

[15] Suchet Bargoti, James P. Underwood "Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards", *Journal of Field Robotics*, vol. 34(6), pp. 1039-1060, 2017.

[16] Chengming Zhang, Jiping Liu, Fan Yu, Shujing Wan; Yingjuan Han, Jing Wang, Gang Wang "Segmentation model based on convolutional neural networks for extracting vegetation from Gaofen-2 images", *J. of Applied Remote Sensing*, vol. 12(4), 2018.

[17] Lu Xu, Dongping Ming, Wen Zhou, Hanqing Bao, Yangyang Chen and Xiao Ling "Farmland Extraction from High Spatial Resolution Remote Sensing Images Based on Stratified Scale Pre-Estimation", *J. of Remote Sens*, vol. 11(2), pp. 10-19, 2019.

[18] Jing Zhang, [et al.] "Multi-Scale Context Aggregation for Semantic Segmentation of Remote Sensing Images", *Remote Sens*, vol.12, 701, 16 p., 2020.

[19] V. Ganchenko, A. Doudkin "Image Semantic Segmentation Based on Convolutional Neural Networks for Monitoring Agricultural Vegetation", *Communications in Computer and Information Science*, Springer, ch.5, vol. 1055, pp. 52-63, 2019.

[20] B. Sobkowiak [et al.]. *Zastosowanie technik analizy obrazu do wczesnego wykrywania patogenow ziemniaka*. Praca nie publicowana, Poznan: PIMR, 2006.

[21] B. Sobkowiak [et al.] *Zastosowanie technik analizy obrazu do wczesnego wykrywania zarazy ziemnechanej w warynkach polowych. Praca nie publicowana*, Poznan: PIMR, 2007

[22] Ke Sun, [et al.] "Deep High-Resolution Representation Learning for Human Pose Estimation", *Proc. of Conference on Computer Vision and Pattern Recognition,* pp. 5693-5703, 2019.

[23] S. Nikolenko, A. Kadurin, E. Archangelskaya *Deep learning*, Saint Petersburg: Piter, 480 pp, 2018. [In Russian]

[24] *Understanding Categorical Cross-Entropy Loss, Binary Cross-Entropy Loss, Softmax Loss, Logistic Loss, Focal Loss and all those confusing names* (04.06.2020) [Online] Available: https://gombru.github.io/2018/05/23/cross_entropy_loss/.

[25] Diederik P. Kingma, Jimmy Lei B "ADAM: A Method for Stochastic Optimization", *Proc. of ICLR 2015*, 15 pp. 2015.

[26] *Keras optimizers: Adam* (24.08.2020) [Online]. – Available: https://keras.io/api/optimizers/adam/.

[27] Marina Sokolova, Nathalie Japkowicz, Stan Szpakowicz "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation", *Advances in Artificial Intelligence*, *19th Australian Joint Conference on Artificial Intelligence,* Hobart, Australia, December 4-8, 2006, pp.1015-1021, 2006.

The work was partially supported by Belarusian Republican Foundation for Fundamental and Romanian Academy (Research (joint project No. F20RA-014) and the National Foundation of Natural Sciences of China (joint project BRFFR-NFNSC No F20-017).