The intelligent analysis system and remote sensing images segmentation engineering by using methods of advanced machine learning and neural network modeling

K E Tokarev¹², Yu A Orlova², A F Rogachev¹², A N Chernyavsky¹ and Yu M Tokareva³

¹Volgograd State Agricultural University, 26, Universitetskiy Avenue, Volgograd, 400002, Russia
²Volgograd State Technical University, 28, Lenina Avenue, Volgograd, 400005, Russia
³Volgograd State Medical University, 1, Pl. Pavshikh Bortsov Square, Volgograd, 400131, Russia

E-mail: tke.vgsha@mail.ru

Abstract. The article describes the main stages of the development of an intelligent system for analysis and segmentation of remote sensing data using deep machine learning and neural network modeling. In particular, the environment for creating an experimental software module was chosen, an experiment on training a neural network was implemented and the results of the training sample were evaluated.

1. Introduction
The rapid growth of modern technical air monitoring of the earth's surface using unmanned aerial vehicles, as well as the growth performance of hardware processing that would result from aerial photographs, allows to apply modern methods of mathematical processing, greatly improve the degree of automation.

The problem of image segmentation using approaches and methods of deep learning with the development of unmanned aerial vehicles is becoming increasingly relevant. Today there is a significant increase in the need for aerial monitoring of agricultural infrastructure.

Classification of objects in the images can be carried out using various methods, but given the characteristics of the problem, computer vision methods are the most suitable and promising. The peculiarities of the task include the requirement for the speed of object recognition. This is because the seeds will fall in front of the camera and when motion is detected in the frame, it will be transferred to the processing. When processing, you need to quickly give the result of belonging to a particular culture. Another feature is the accuracy of recognition. It is necessary to determine weed seeds as accurately as possible to minimize the number of sorting iterations to achieve the desired seed purity. All these conditions are fully satisfied by classifiers based on machine and deep learning.
2. Materials and methods
The neural network used to construct the segmentation function is based on a convolutional neural U-network that was created specifically for image segmentation. Algorithm of the network in the form of a graphic scheme is similar to the Latin letter U, from which this neural network got its name.

![Algorithm of U-networks](image)

Figure 1. Algorithm of U-networks.

The algorithm of the U-network (figure 1), is described as follows: each vertical line denotes a 3-dimensional matrix, which at the bottom left shows its dimensions in the first and second dimensions (their physical meaning - height and width), and the top dimensions in the third dimension (physical meaning – the number of channels). For example, in this notation, the input image corresponds to a 572x572 matrix with 1 channel (grayscale image). The output is a 2-channel mask.

The algorithm consists in a process of transformations. The network uses the following types of transformations:

- Conv 3x3: convolution conversion of an image with a 3x3 cm window by a 3x3 dimension filter with N channels, resulting in an image with MxN channels.
- ReLU: denotes the activation transformation (replacing negative values with 0), more complex functions such as hyperbolic tangent can also be used.
- Max pool 2x2: denotes a subdiscretization conversion when the maximum value is selected instead in the 2x2 window.
- Upconv 2x2: denotes the conversion of convolution of the image with a window 2x2 on the filter dimension 3x3xN. to perform this algorithm from the window 2x2 window is formed 3x3 by filling the Central elements with zeros, and angular-the values of the original window.
- Soru and crop: denotes the transformation of the merger of two matrices by supplementing the processed channels with the original channels.
- Conv 1x1: denotes convolution of the image with the window 1x1 on the filter of dimension 1x1. Used to reduce the dimension of the number of channels.

Transformations 1), 4) and 6) require adjustment of convolution matrix coefficients, which is implemented by an iterative process on a training sample, in which the initial values are selected as
Thirty-one the method of optimization is the Adam method, which is a modification of the gradient descent method with the accumulation of momentum of motion, and for the recalculation of the coefficients of the matrices, the method of back propagation of the error is chosen, which consists in a special way of calculating the gradient of the coefficient vector, taking into account that they are matrices.

Segmentation of one image by given K classes is understood as a function:

$$F: [0..1, 0..1, 0..1]^{W \times H} \rightarrow [0..K]^{W \times H}$$

The input value is an image with a width \( W \) of pixels and a height \( H \), represented as a matrix of pixels, where each pixel is encoded with the color R, G, B corresponding to the red, green and blue components.

The output value is a mask matrix of size \( W \times H \), where each value is encoded with the object class code (0 means that the object does not belong to any class, a value of \( k \) from 1 to \( K \) means that the object belongs to the \( k \)-th class).

To further describe the algorithm, another equivalent definition of the segmentation function is used – it is a function according to:

$$F: [0..1, 0..1, 0..1]^{W \times H} \rightarrow [0..K]^{W \times H(K+1)}$$

The input value is an image with a width \( W \) of pixels and a height \( H \), represented as a matrix of pixels, where each pixel is encoded with the color R, G, B corresponding to the red, green and blue components.

The output value is a \((K+1)\)-channel mask of size \( W \times H \) each, where in the \( k \)-th channel \((k = 0..K)\) in each cell, a real number from 0 to 1 is encoded, which is the value of the probability that the corresponding pixel of the original image belongs to the \( k \)-th class, while the zero channel stores information about the probability that the pixel does not belong to any class. Thus, the sum of the values of all channels for each pixel is 1.

In this case, to move from the second output value to the first, it is enough to round off the mask values and form a single mask with class codes for a set of masks for each class.

Under the reference mask \( me \) of the original img image we will understand the mask with true values, formed manually. The values of the reference mask cells can only be 0 or 1. For the reference mask \( me \) and calculated using the segmentation function \( f \) mask \( m \) \((m = f(\text{img}))\), you can determine the loss function \( H \) \((me, m)\), characterizing the correlation – the degree of deviation of \( m \) from \( me \). As such, the function is used to calculate the binary cross-entropy according to:

$$H(me, m) = -\sum_{i=0}^{K} \sum_{i=1}^{W} \sum_{j=1}^{H} me_{ijk} \ln(m_{ijk})$$

when two masks are completely identical, \( H(me, m) = 0 \), and the less their similarity (correlation), the greater the values the loss function will take.

Taking into account the above definitions, the delivery of the work task is described as follows:

The input data for the problem is an input sample of images and their corresponding reference masks, divided into a training sample according to:

$$T=(\text{img}, me), i=1..N_T$$

and test sample according to:

$$V=(\text{img}, me), i=1..N_V$$

The basic approach to designing a software product was to decompose the functionality into the software modules that implement them, taking into account the fact that most of them can be used existing open source software libraries.

The software was developed using Python, a high-level General-purpose programming language focused on improving developer productivity and code readability. The learning and segmentation
algorithms implemented with Python have a minimalist structure, using a large number of functions that are contained in the Python standard library.

As a library for the implementation of machine learning algorithms, the PyTorch library was chosen, which was initially developed by the artificial intelligence algorithms research team at Facebook, and allows using the computing power of the video card for calculations.

In the PyTorch library there are ready-made functions for specifying a model of a neural network of the U-network type by combining possible transformations, as well as its calculation and the function for optimizing the coefficients of the neural network in training using CUDA parallel computing technology (figure 2). Python imaging library is used as a library to implement work with images.

```python
def __init__(self, config=config_dict):
    super(UNet, self).__init__()

    self.dropout = config.get('dropout', 0.8)
    self.pool_ksize = config.get('pool_ksize', 2)
    self.n_channels = config.get('n_channels', 3)
    self.n_classes = config.get('n_classes', 2)

    self.upsample = nn.Upsample(scale_factor=self.pool_ksize)

    self.down0 = self._module(self.n_channels, 32)
    self.down1 = self._module(32, 64)
    self.down2 = self._module(64, 128)
    self.down3 = self._module(128, 256)
    self.down4 = self._module(256, 512)

    self.up1 = self._module(512 + 256, 256)
    self.up2 = self._module(256 + 128, 128)
    self.up3 = self._module(128 + 64, 64)
    self.up4 = self._module(64 + 32, 32)

    self.pool = nn.MaxPool2d(self.pool_ksize)

    self.final_conv = nn.Conv2d(32, self.n_classes, 1)
```

**Figure 2.** Part of the main algorithm of the U-network model.

Blue on the graph shows the loss function for the training sample and orange for the test sample (figure 3).

**Figure 3.** Calculation of the dependence of the loss function on the number of cycles.
As a result of the experiments, a graph of the loss function (the numerical value of the probability of non-recognition) was constructed depending on the number of training cycles conducted, within each of which training was conducted for all sets of training samples, and after each training for each training cycle, the calculation of the loss function for the test set was performed.

3. Conclusion
Studies have shown that the neural U-network is applicable as a basic mathematical apparatus for solving problems of image segmentation. One of its key features is a significantly small number of weight coefficients to be optimized in the learning process compared to the classical neural network built on the basis of perceptrons. In the case of a network, one small filter (3x3, 2x2, 1x1) is used for the entire image and does not require personal weighting for each pixel. Thus, training a neural network consists in generalizing the allocated informative elements, rather than images of the trained sample, as does a neural network based on perceptrons by tuning billions of weight coefficients.

Another important advantage of the neural network is the ability to parallelize calculations to use the graphics card architecture as a General-purpose coprocessor and provides algorithms in real-time systems. The chosen method of forming a training sample ensures the stability of the segmentation algorithm to the rotation and shift of the recognized image, as well as meteorological conditions of shooting.

The disadvantages of the chosen method is the absence at the moment, recommendations on the optimal formation properties of neural network, including number of layers, the dimensions of the filter to perform convolution for each of the layers, the choice of functions with decreasing dimensions (maximum, average value and etc.). All these parameters are selected in this paper in an empirical way, can have a significant impact on the result.

4. Acknowledgments
The reported study was funded by RFBR and Volgograd region according to the research project № 19-416-343006.

References
[1] Van Der Malsburg C 1986 Frank Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms Brain Theory pp 245-8
[2] Jang J-S R 1993 ANFIS: Adaptive-Network-Based Fuzzy Inference System IEEE Trans. Systems & Cybernetics 23 665–85
[3] Tahmasebi P 2012 A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation Computers & Geosciences: journal 42 18-27
[4] Rogachev A F, Medvedeva L N, Timoshenko M A, Tokarev K E, Shiro M S 2019 Using Computer Modeling During Evaluation of Socio-Labor Potential of the Region. Advances in Intelligent Systems and Computing 726 989–1004
[5] Rogachev A F, Shokhnekh A V, Melikhova E V 2018 Monitoring and Economic & Mathematical Modeling of Manufacture and Consumption of Agricultural Products as a Tool of Food Security Management Revista Espacios 39 (1)
[6] Nikitin N A, Rozaliy V L, Orlova Y A and Alekseev A A 2018 Automated sound generation based on image colour spectrum with using the recurrent neural network CEUR Workshop Proceedings 2212 399–408
[7] Kruzhilin I P, Ovchinnikov A S, Kuznetsova N V, Kozinskaya O V, Fomin S D, Bocharnikov V S, Vorontsova E S 2018 Water pressure monitoring in irrigation piping as quality management tools of sprinkler irrigation ARPN Journal of Engineering and Applied Sciences 13(13) 4181–4
[8] Ovchinnikov A S, Bocharnikov V S, Skorobogatchenko D A, Borisenko I B, Chernyavsky A N 2018 The optimum geometrical form modeling of the “Striegel” type harrow ARPN Journal
of Engineering and Applied Sciences 13(23) 9138–44
[9] Kruzhilin I P, Dubenok N N, Ganiev M A, Ovchinnikov A S, Melikhov V V, Abdou N M, Rodin K A Mode of rice drip irrigation 2017 ARPN Journal of Engineering and Applied Sciences 12(24) 7118–23
[10] Timoshenko M A, Rogachev A F, Medvedeva L N, Tokarev K E 2019 Analysis and support of decision making with the use of latest information and communication technologies for development of social and economic area with regard to capabilities of different categories of population The Leading Practice of Decision Making in Modern Business Systems: Innovative Technologies and Perspectives of Optimization p 214
[11] Yurchenko I F 2018 Information support for decision making on dispatching control of water distribution in irrigation Journal of Physics: Conference Series 1015(4) 042063