Determination of the Requirements for the Neural Network for Recognition Algorithm UHRSI

View the article online for updates and enhancements.

Determination of the Requirements for the Neural Network for Recognition Algorithm UHRSI

A A Ostrovskaya[1], N E Semenov[1] and A O Rubtsov[2]

[1] Peoples' Friendship University of Russia, Moscow 117198, Russia
[2] Russian Space Systems Corporation, Moscow 111250, Russia

E-mail: ostrovskaya_aa@rudn.university

Abstract. The Earth remote sensing is becoming a new and quickly developing multiscience area of a practical importance and different business applications. The IT industry now in accordance with the academic sector is developing new competences in the area which includes the development of new algorithms and calibration of existing and already working program complexes and libraries. Investigation about using neural networks for detection geo-objects on the satellite images are held by different research organizations. In this paper we present the new way of determination of the initial data, the architecture type of the neural network, the accuracy characteristics of the recognized objects for the algorithm of recognition of soil use areas of open type based on the data of ultrahigh resolution space imagery (UHRSI) which was made according to the practice on the newest satellite images. The results could be used in the various application of remote data analysis including satellite application in different branches of military and economic needs.

1. Introduction

The Earth remote sensing is becoming a new and quickly developing interdisciplinary area of a practical importance. In machine learning applications for remote sensing, aerial image interpretation is usually formulated as a pixel labeling task. Given an aerial image the goal is to produce either a complete semantic segmentation of the image into classes such as building, road, tree, grass, and water or a binary classification of the image for a single object class. So the new algorithms should be devised and studied. Below we present a way to set the preliminary conditions for the algorithm developed in the project of RUDN University and Russian Space Systems Corporation as an approach for a new competence development. Below we present our findings in determination of the requirements for the original data, the architecture of the neural network, the accuracy characteristics of the recognized objects and the development of an algorithm for recognition of quarries in accordance with ultra-high and high-resolution space imagery data.

2. Findings

1. Requirements to the initial data for the algorithm of detecting of soil use areas of open type.
- Images from the sensor of Geoton and Wide-resolution multispectral imaging equipment of high resolution of Resource-P Satellite, panchromatic and multizonal systems of Canopus-V, Planet;

- Presence of 4 channels (red, blue, green and near infrared);
- Processing level 4A is the integrated panchromatic image (processing level 2A) and multi-spectral (processing level 2A1) images of the same territory (Pansharpening);
- Cloudiness less than 70% of the image coverage;
- The angle of deviation from the nadir of the original image should not exceed 30 degrees;
- The point sizes of each satellite image channel must be pairwise identical;
- The point sizes of the channels must have at least 1024 points in each dimension;
- The information component of each satellite channel point must be 10 bits

2. Determination of the requirements to the neural network architecture

- The neural network must have at least 10 layers;

- The neural network must have multiple inputs and one output;

- The point sizes of the result should correspond to the point sizes of the input image;

- The architecture of a neural network should provide the possibility of using methods of regularization and retraining prevention, namely, the methods of "normalizing batches" [1] and "exceptions" [2];

- The architecture should include links of the "passthrough characteristics" type [3];

- The architecture of the neural network should be adapted to the learning transfer strategy for a part of the layers

3. Determination of requirements to the accuracy characteristics of recognized objects

- The soil use areas of open type with a total linear dimension of less than 100 meters, should achieve the acquisition accuracy of 0.5 and segmentation accuracy (measure of Sørensen) of each object of no less than 0.2;

- The soil use areas of open type with a total linear dimension exceeding 100 meters, should achieve the acquisition accuracy of 0.8 and segmentation accuracy (measure of Sørensen) of each object of no less than 0.75;

- The soil use areas of open type with a total linear dimension exceeding 750 meters, should achieve the acquisition accuracy of 0.94 and segmentation accuracy (measure of Sørensen) of each object of no less than 0.8;

- The soil use areas of open type with a total linear dimension exceeding 2000 meters, should achieve the acquisition accuracy of 0.97 and segmentation accuracy (measure of Sørensen) of each object of no less than 0.95. Accuracy = TP + TN/Total (1) where TP is the number of true object definitions, TN is the number of false object definitions, and Total is the total number of unique objects. Another value 2TP/(2TP+FP+FN) (2) is the measure of Sørensen, where TP is the number of true definitions of points in the object, FP is the number of false positive definitions of points in the object, FN is the number of false negative definitions of points in the object.

3. Algorithm

Development of an algorithm for recognition of soil use areas of open type based on ultra-high resolution space imagery.

The general developed scheme of the algorithm for recognition of soil use areas of open type according to the data of ultra-high resolution space imagery is shown in Figure 1 and 2.

On entry the algorithm is awaited by 4-channel satellite image, made by canopus-B device, resurs-P, 4A processing level planet - a complex image of panchromatic (2A processing level) and multispectral (2A1 processing level) images of the same territory (Pansharpening).
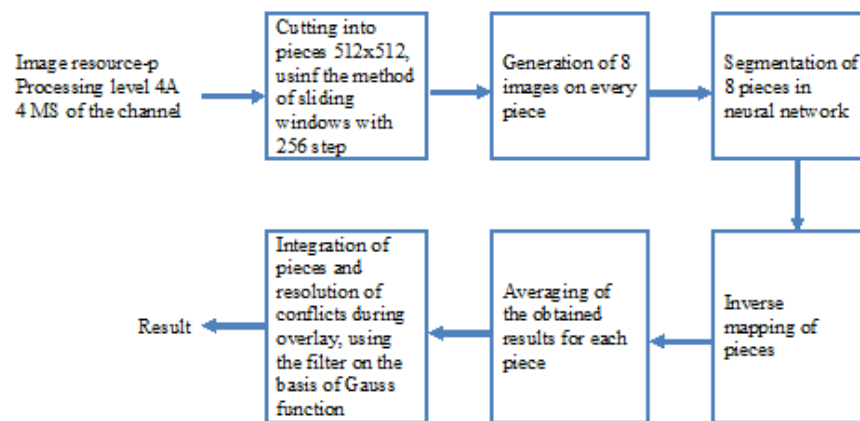
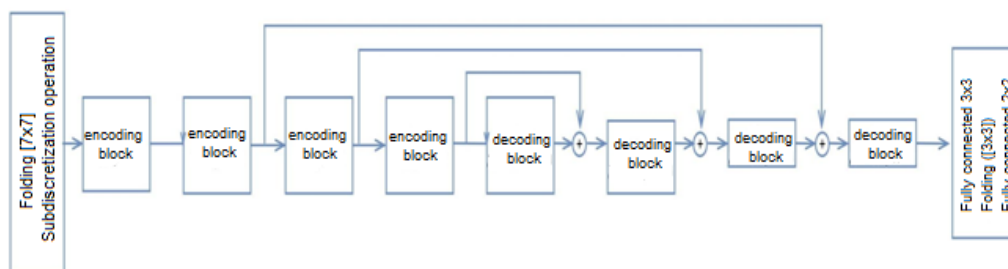**Figure 1.** Diagram of the algorithm for recognition of buildings and structures

Figure 2. Architecture of the neural network

Figure 3. Splitting the original image into pieces of fixed size, using the sliding window method

Sliding window method (example on figure 3):
1. Make the size of the sliding window - $H_{SW} \times W_{SW}$ pixels;
2. Make the increments of the sliding window - S_H pixels (vertically) and S_W pixels (horizontally);
3. Complete the edges of the initial image I with the height H and width W to the size of the sliding window;
4. Form the pieces of the image (crop) of the size $H_{SW} \times W_{SW}$ from the complemented image in increments of $S_H \times S_W$

The following values were chosen for this algorithm:

- $H_{SW} = W_{SW} = 512$ pixels, to ensure the hit of rather big objects in one crop, which will allow to achieve the specified accuracy, and the practical possibility to create the software implementation on one hand, on the other hand, since the size of the window directly determines the size of the layer of the neural network, which increase leads to the increase of the requirements for the size of memory and the speed of the hardware and exponential increase of the learning time.

- $S_H = \frac{H_{SW}}{2} = 256$ pixels, $S_W = \frac{W_{SW}}{2} = 256$ pixels, to ensure the overlay of the crops so that the edge of one crop coincides with the center of the adjacent. It will enable to avoid artifacts and conflicts on the edges of the crop when restoring the segmentation of the entire image from the segmentations of certain crop areas.

Generating of 8 reflections for each crop - all possible variations of the reflection of the initial piece, using the rotation operations on $\pi/2$ and the mirror reflection (can be seen on figure 4).
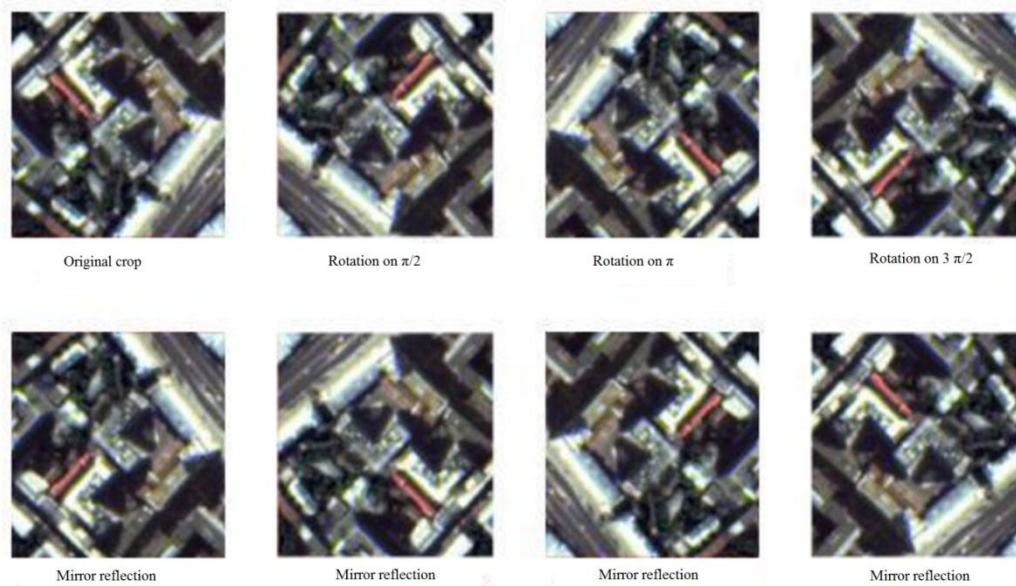

Figure 4. Turns and reflections of the crop

Let C be the crop, $\varphi(x)$ is a rotation by $\pi/2$, $\psi(x)$ is a mirror reflection, then the set 8 of reflections O can be represented as:

$$O = \left\{ C, \varphi(C), \varphi(\varphi(C)), \varphi\left(\varphi(\varphi(C))\right), \psi(C), \psi(\varphi(C)), \psi\left(\varphi(\varphi(C))\right), \psi\left(\varphi\left(\varphi(\varphi(C))\right)\right) \right\} \quad (3)$$

Obtain a segment map for each of the 8 reflections. At this stage, the original crop is passed through a trained deep neural web. Denote the function of obtaining the prediction P on the crop image C, as h. Then:

$$P = h(C). \quad (4)$$

The network architecture is a sequence of encoding blocks that reduce the spatial resolution of the original crop, and decoding blocks that increase the spatial resolution, combining input data with feature maps, obtained by the transmission method from the encoding blocks of the corresponding resolution, which ensures the ensemble of the results of all layers and resolutions.

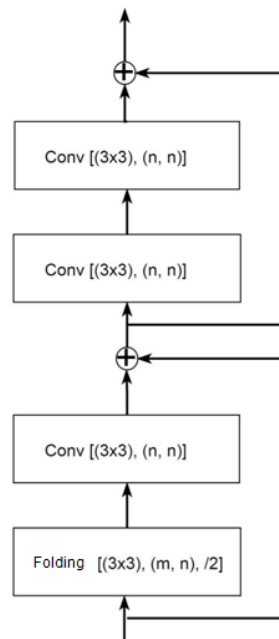The encoding block is a set of 3 operations on feature cards (figure 5).

Figure 5. Block diagram of the block of encoding

These operations are:

- the convolution layer is the main block of the convolutional neural network (figure 6). The convolution layer includes its own filter for each channel, which convolution kernel processes the previous layer in fragments (summing the results of the matrix product for each fragment). It is denoted as Convolution ([k x k], m, n), where k x k is the size of the convolution kernel, m and n are the number of input and output filters for the layer, respectively;
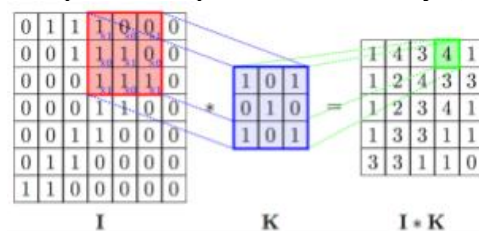

Figure 6. Example of convolutional layer operation

- ReLU activation function (figure 7). The scalar result of each convolution falls on the activation function, which is a function $\begin{array}{l} += max(0,x) \\ ReLU = x \end{array}$ this function enables to avoid the problems of a damping and exploding gradient, and is computationally simple.
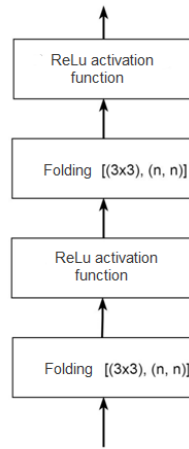
Figure 7. Diagram of the decoding unit

The decoding block is a set of 2 sequences of two operations:
- convolution layer;
- ReLU activation function.

As a result, at the output of the neural network segmentation, a set of 8 maps of probability that each point of the original crop belongs to the class "open-type subsurface resources" is obtained.

Reverse mapping operations (rotation by $-\pi/2$ and mirror mapping) are applied to the obtained set of probability maps in order to obtain the preimages of the used images with respect to the output probability maps. Since $x = \psi(\psi(x))$ and $x = \varphi^{-1}(x) = \varphi\left(\varphi(\varphi(x))\right)$, the target set O_P will have the form:

$$O_P = \{P_1, \varphi^{-1}\left(\varphi^{-1}(\varphi^{-1}(P_2))\right), \varphi^{-1}(\varphi^{-1}(P_3)), \varphi^{-1}(P_4), \psi(P_5),$$

$$\psi\left(\varphi^{-1}\left(\varphi^{-1}(\varphi^{-1}(P_6))\right)\right), \psi\left(\varphi^{-1}(\varphi^{-1}(P_7))\right), \psi(\varphi^{-1}(P_8))\} \qquad (5)$$

The amendment of the boundaries of the obtained segments by averaging the predictions. Then the final prediction for each point of the crop image will be calculated, using the following formula

$$P_{result}(i,j) = \frac{\sum_{k=1}^{8} P_k(i,j)}{8}$$

$$(6)$$

This approach improves the segmentation results, obtained at the previous stage.

Let us combine the obtained intersecting prediction maps with the help of the weighted sum, using a two-dimensional Gaussian distribution with zero at the center of the crop and a mean-square deviation $\sigma = \frac{H_{SW}}{2*3} \approx 85$, calculated at points, corresponding to the centers of the crop pixels to get the segmentation of the original image. This will eliminate conflicts and artifacts on the crop boundaries, since the crop, which is closer to its center, will make the greatest contribution to each pixel value and the contribution of the terminal crop points is ~ 0.2.

4. Conclusion

In this paper we presented the way of determination of the initial data, the architecture of the neural network, the accuracy characteristics of the recognized objects for the algorithm of recognition of soil use areas of open type based on the data of ultra-high resolution space imagery (UHRSI) which was made according our research. The results could be used in the different application of remote data analysis. This determination is the necessary part in the development of new competences of remote

data sensing [4-8]. This work lays a foundation for the new competence hub being created by RUDN and RSS, which will interact with leading Russian collectives working in remote monitoring.

Acknowledgments

References

[1] Chaurasia A and Culurciello E 2017 Linknet Exploiting encoder representations for efficient semantic segmentation *2017 IEEE Visual Communications and Image Processing (VCIP)* pp 1-4

[2] Ioffe S and Szegedy C 2015 *Batch* Normalization Accelerating Deep Network Training by Reducing Internal Covariate Shift *International Conference on Machine Learning* pp 448-456

[3] Srivastava N et al. 2014 Dropout: a simple way to prevent neural networks from overfitting *The Journal of Machine Learning Research* **15**(1) 1929-1958.

[4] Kashirin A, Semenov A, Ostrovskaya A and Kokuytseva T 2016 The Modern Approach to Competencies Management Based on IT Solutions. *JIBC-AD - Journal of Internet Banking and Commerce* **21** 1

[5] Chursin A, Kashirin A et al. 2018. The approach to detection and application of the company's technological competences to form a business-model *IOP Conference Series Materials Science and Engineering* **313** 012003 (doi:10.1088/1757-899X/312/1/012003)

[6] Kashirin A, Semenov A, Ostrovskaya A and Kokuytseva T 2016 The Modern Approach to Competencies Management Based on IT Solutions *JIBC-AD - Journal of Internet Banking and Commerce* **01** 1-12

[7] Kashirin A, Semenov A, Ostrovskaya A, Kokuytseva T and Strenaluk V 2016 The Modern Approach to Competence Management and Unique Technological Competences *Quality – access to success* **17**(154) 105–109

[8] Chursin A, Semenov A and Danilchankau A 2016 Analysis of Innovation Development in the Economy with Exhaustible Resource Sector by First Order Dynamical Systems Application *Nonlinear phenomena in complex systems* **19**(3) 254-270