Neural network texture segmentation of satellite images of woodlands using the U-net model

Anna E. Alyokhina[1], Dmitry S. Rusin[1], Egor V. Dmitriev[2] and Anastasia N. Safonova[1]

[1]Siberian Federal University, Krasnoyarsk, Russia
[2]Marchuk Institute of Numerical Mathematics of the Russian Academy of Sciences, Moscow, Russia

Abstract

With the advent of space equipment that allows obtaining panchromatic images of ultra-high spatial resolution ($<$ 1 m) there was a tendency to develop methods of thematic processing of aerospace images in the direction of joint use of textural and spectral features of the objects under study. In this paper, we consider the problem of classification of forest canopy structures based on textural analysis of multispectral and panchromatic images of Worldview-2. Traditionally, a statistical approach is used to solve this problem, based on the construction of distributions of the common occurrence of gray gradations and the calculation of statistical moments that have significant regression relationships with the structural parameters of stands. An alternative approach to solving the problem of extracting texture features is based on frequency analysis of images. To date, one of the most promising methods of this kind is based on wavelet scattering. In comparison with the traditionally applied approaches based on the Fourier transform, in addition to the characteristic signal frequencies, the wavelet analysis allows us to identify characteristic spatial scales, which is fundamentally important for the textural analysis of spatially inhomogeneous images. This paper uses a more general approach to solving the problem of texture segmentation using the convolutional neural network U-net. This architecture is a sequence of convolution-pooling layers. At the first stage, the sampling of the original image is lowered and the content is captured. At the second stage, the exact localization of the recognized classes is carried out, while the discretization is increased to the original one. The RMSProp optimizer was used to train the network. At the preprocessing stage, the contrast of fragments is increased using the global contrast normalization algorithm. Numerical experiments using expert information have shown that the proposed method allows segmenting the structural classes of the forest canopy with high accuracy.

Keywords

Neural network, segmentation, satellite images, U-net.

1. Introduction

Monitoring of forest areas, namely textural segmentation and forest mapping is an urgent task. One of the promising ways of global tracking of areas is the use of remote sensing data of the Earth (remote sensing). Due to the growth and diversity of information, there is a need to develop and modernize new methods of its processing. So, in recent years, due to the development of production capacities, one of the fastest growing areas is artificial intelligence. Thus, the idea of this experiment is to use remote sensing data and neural network methods to solve the problem of texture segmentation. In particular, there are works in the literature

SDM-2021: All-Russian conference, August 24–27, 2021, Novosibirsk, Russia
a.tolmacheva@solutionfactory.ru (A. E. Alyokhina); yegor@mail.ru (E. V. Dmitriev)

that are close to our experiment. In [1], the authors performed texture segmentation using the AGMSSeg-Net neural network, interactively selected by the user. Models based on convolutional neural networks, such as the second [2] and third versions of Deeplab [3], were also successfully used to create color labels on maps that allow solving the problem of textural segmentation of forest zones.

After the analysis of related works, we decided to use the Unet model to perform the task of textural segmentation of woodlands using Worldview-2 panchromatic images.

The main contribution to the work is as follows:

1. Preliminary image processing was performed using the global contrast normalization method.
2. The U-net model was trained on the original and pre-processed images.
3. Texture segmentation was performed by the final version of the trained model with the best result of metrics.
4. The errors were compared by the cross-validation method.

2. Materials and methods

2.1. The research area

The research area is located on the territory of the Moscow region, Bronnitsky forestry in the immediate vicinity of the territory of geographical landings of the forester P.I. Dementieva. The stands of the Bronnitsky forestry have an age of 40 years or more and, according to the variety of species, they cover all the main forest-forming breeds of Russia. The selected site contains natural and forest-cultural plantings with different species composition and visible textural differences of the forest canopy. The plot contains part of the territory of permanent larch forest-seed plantations, which have a pronounced regular structure. The plot also contains natural birch and pine (with an admixture of spruce) stands of various completeness.

Multispectral and panchromatic images of WorldView-2 with a spatial resolution of 1.85 and 0.46 m, respectively, were used as satellite information. The photo was taken on June 28, 2011, before the construction of the Novoryazansky Highway and the Central Ring Road began. For texture processing, a panchromatic image was used, which, after correction, has a spatial resolution of 0.5 m (Fig. 1).

2.2. Pre-processing of images

This subsection presents an algorithm for preprocessing a satellite image, which consists of:

1. Converting fragments from the format .tiff to .png format for further work with the neural network. In this study, fragments of images with a size of 27×27 pixels were prepared for training.
2. Increasing the contrast of fragments using the global contrast normalization algorithm [4]:

$$X'_{i,j,k} = s \frac{X_{i,j,k} - \overline{X}}{\max\left\{\epsilon, \sqrt{\lambda + \frac{1}{3rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{3}(X_{i,j,k} - \overline{X})^2}\right\}}, \quad (1)$$

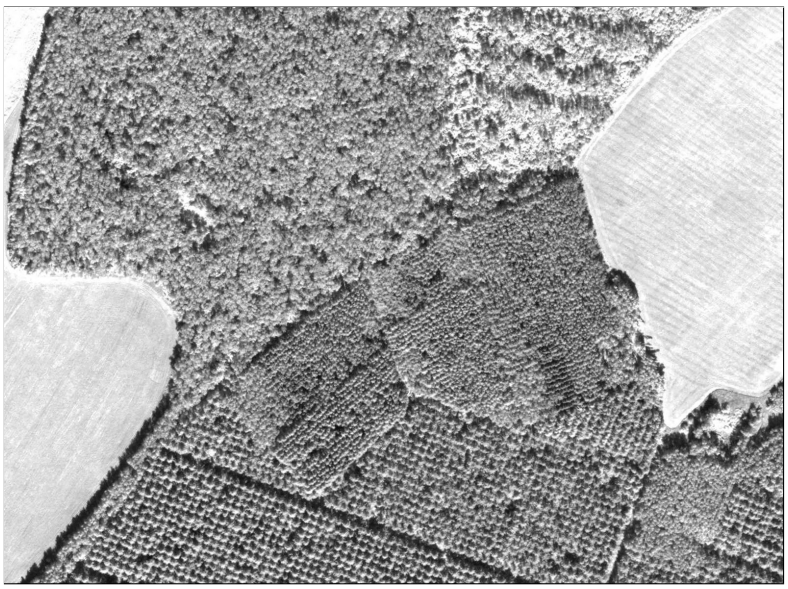


Figure 1: A fragment of a satellite image of the test site of the Bronnitsky forestry.

where $X_{i,j,k}$ is the tensor of the original image, $X'_{i,j,k}$ is the tensor of the normalized image, $\overline{X} = \frac{1}{3rc}\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{3} X_{i,j,k}$ is the average pixel value of the original image, ϵ and λ are some constants, in our solution $\lambda = 10$, $\epsilon = 0.000000001$, respectively.

Figure 2 below shows the results of image processing by the global contrast normalization algorithm.

A test sample of the studied textures was prepared from 3500 transformed fragments, of which 80% were allocated for training, 20% for validation, and one image was used for independent verification of the trained U-net model. For each class, 400 training segments, 100 test segments and 56 segments per test zone A were allocated.

There was one color annotation label per fragment that belongs to a certain class.

Most classes differ in the density of green spaces, as well as the variety of types of rocks. Two classes are an ordinary grass field.

2.3. The U-net model

In this work, the Xception model [5] was used, due to the fact that with the help of this convolutional network architecture, it is possible to obtain a better result compared to Inception V3 [6], as presented in [7]. The Xception architecture represents a fully connected convolutional network that is able to work with a small number of training examples for segmentation tasks. The generalized U-net architecture [8] is shown in Figure 3.

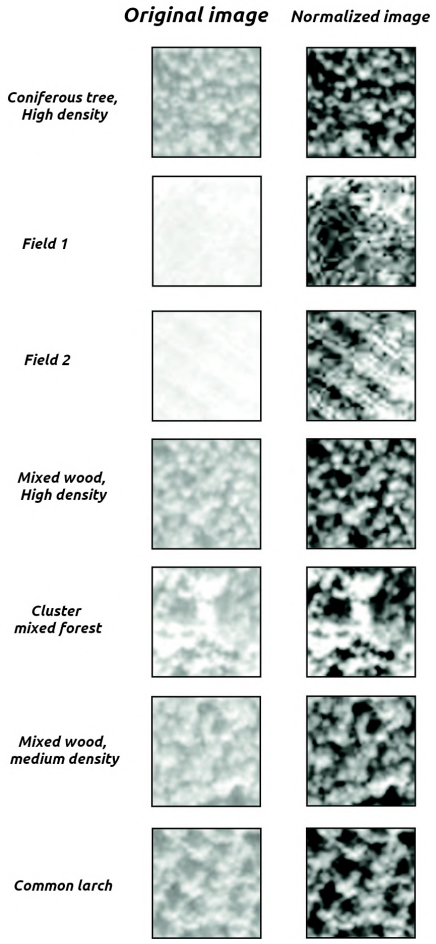


Figure 2: The result of normalization of global contrast on the example of seven texture zones, where the first image is a coniferous tree, the second is field 1, the third is field 2, the fourth is a mixed forest, the fifth is a cluster mixed forest (average density), the sixth is ordinary larch.

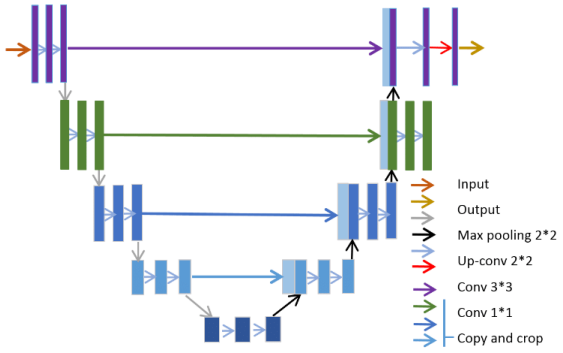


Figure 3: Generalized U-net architecture (figure from https://arxiv.org/abs/1505.04597).

2.4. Metrics

To calculate the effectiveness of the trained model, we used the mAP and IoU metrics [9]. IoU is just a score indicator. Any algorithm that provides predicted bounding rectangles as output can be evaluated using IoU.

1. Reliably marked areas manually by an expert.
2. Certain results of the trained network (2):

$$IoU = \frac{Areaofoverlap}{Areaof\Cup} \tag{2}$$

where $Areaofoverlap$ is the label of ground truth, and $Areaof\Cup$ are the labels of prediction and truth.

Additionally, $F1 - score$ (3) and mAP (4) are calculated to evaluate the performance of the model. The $F1 - score$ is calculated based on Accuracy (Precision) (5) and memorization (Recall) (6). mAP is the average value for all classes or finding the area under the Precision-Recall curve above [10] mAP is calculated in the range from 0 to 1 using the following formula (4):

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall}, \tag{3}$$

$$mAP = \sum_{i=1}^{N} AP_i = \frac{1}{N} \sum_{Recall} Precision(Recall), \tag{4}$$

$$Precision = \frac{TP}{TP + FP}, \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

where TP is a true positive result, FP is a false positive result, and FN is a false negative result.

The sparse categorical cross entropy ($SCCE$) (7) was used to calculate the model loss parameter

$$SCCE = -\sum_{i=1}^{n} (x_i * \log(\sigma(y_i))) \tag{7}$$

where $\sigma(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{n} e^{y_i}}$ or is the normalized exponent.

3. Results

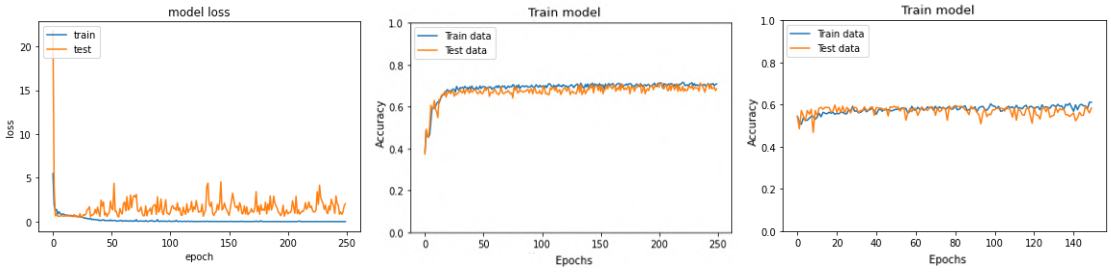
The *RMSprop* optimizer was used to train the network. The number of epochs was 250. The time spent for one epoch is about 5–6 minutes. The training was carried out on the Google

Colab platform [11]. The learning process is shown in Figure 4. The quality of using normalized images was compared.

As you can see, increasing the contrast slightly improves the quality of the model, a more detailed comparison of the results is presented below.

At the exit from the network, a mask is formed that corresponds to a certain forest structure. An example of the final processing of a test image is shown in Figure 5.

The main metrics of this work are presented in Table 1. mAP and $F-score$ take an average value for small images of the test area.



a *b* *c*

Figure 4: The value of the error function (*a*) and accuracy value (*b* and *c*) at each epoch for the test and training processed data.

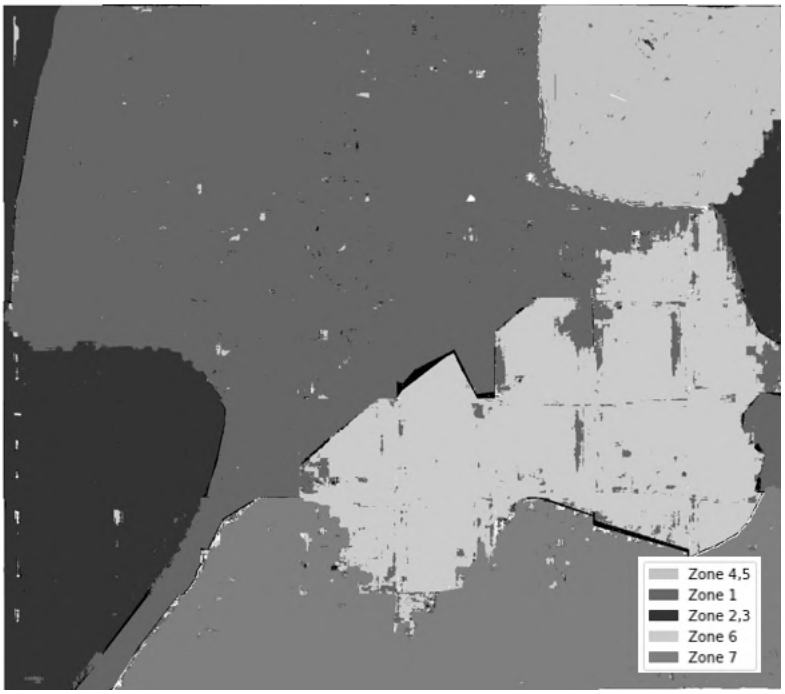


Figure 5: An example of an output mask based on a photo of the study area, where zone 1 is a coniferous tree, zone 2 is field 1, zone 3 is field 2, zone 4 is a mixed forest, zone 5 is a cluster mixed forest (average density), and zone 6 is ordinary larch.

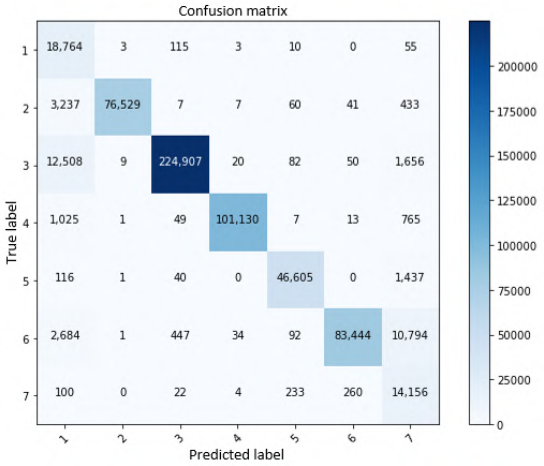


Figure 6: The confusion matrix of the trained model in the test image, where class 1 is a coniferous tree, class 2 is field 1, class 3 is field 2, class 4 is a mixed forest, class 5 is a cluster mixed forest (average density), and class 6 is an ordinary larch.

Table 1
Results of network metrics

Name	The value of the processed data	The value of the raw data
loss	2.3654e−04	0.18
mAP	the average for each segment is 0.71	the average for each segment is 0.63
$F-score$	average 0.73	average 0.59

Table 2
Results of network metrics

No.	Data without pre-processing		Data from pre-processing	
	$F-score$	loss	$F-score$	loss
1	74.68	38.34	77.78	24.68
2	74.03	32.03	82.97	20.79
3	70.13	33.89	74.53	40.57
4	69.28	59.12	82.82	21.68
5	75.16	29.88	83.47	20.37
Average	72.66	38.65	80.31	25.61

After training, the model was tested on test images. The results of the predicted masks were compared with the test masks using several quality metrics. The results are presented in Table 1.

Cross-validation was also carried out for evaluation on independent data. The parameter k was equal to 5. The results of the sliding control are presented in Table 2.

As can be seen from Figure 4, *c*, we have a predominance of the third class, and also sometimes the model considered the 3rd class as the 1st class, perhaps due to the close intersection of these classes, this error occurred, you can also see the erroneous definition of the ForestMixedNormal class as the LarchRegularNormal class. This was due to the similar data structure of the classes.

4. Conclusion

Based on the results, we can conclude that the U-net model copes with the processing of satellite images of forest areas for segmentation tasks. The main structure of each type of forest is clearly highlighted in the image for a better result, a larger set of image data is still needed, on which several classes will intersect.

In the future, it is planned to process high-resolution images (36 pixels) this will allow using several classes on one image, and it is also planned to use classical architectures of convolutional models of neural networks with a change in architecture to increase efficiency and compare with the U-net model in new areas.

Acknowledgments

The research was carried out with the financial support of the RFBR (projects No. 19-01-00215 and No. 20-07-00370).

References

[1] Li K., Hu X., Jiang H., Shu Z., Mi Z. Attention-guided multi-scale segmentation neural network for interactive extraction of region objects from high-resolution satellite imagery // Remote Sensing. 2020. Vol. 12. P. 789. DOI:10.3390/rs12050789.

[2] Bengana N., Heikkilä J. Improving land cover segmentation across satellites using domain adaptation // Remote Sensing. 2020. DOI:1912.05000.

[3] Barmpoutis P., Stathaki T., Dimitropoulos K., Nikos G. Early fire detection based on aerial 360-degree sensors, deep convolution neural networks and exploitation of fire dynamic textures // Remote Sensing. 2020. Vol. 12. P. 3177. DOI:10.3390/rs12193177.

[4] Bengio Y. Deep learning. 2016. URL: https://www.deeplearningbook.org.

[5] François C. Xception: Deep learning with depthwise separable convolutions // arXiv preprint. 2017. arXiv:1610.02357v3 [cs.CV].

[6] Szegedy C. et al. Rethinking the inception architecture for computer vision // arXiv preprint. 2016.

[7] Canziani A., Paszke A., Culurciello E. An analysis of deep neural network models for practical applications. URL: https://arxiv.org/abs/1605.07678.

[8] Hui J. mAP (mean Average Precision) for object detection. 2018. URL: https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173.

[9] Scikit-learn. Precision-Recall. URL: https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html.

[10] Canziani A., Paszke A., Culurciello E. An analysis of deep neural network models for practical applications. URL: https://arxiv.org/abs/1605.07678.

[11] Hinton G. Neural networks for machine learning. Online course. URL: https://www.coursera.org/leture/neural-networks-deep-learning/geoffrey-hinton-interview-dcm5r.

[12] Colab G. Research notebooks. URL: https://colab.research.google.com.