Selection of features system and network parameters for hyperspectral images classification using convolutional neural networks

Victor I. Kozik¹, Evgeniy S. Nezhevenko¹

¹Institute of Automation and Electrometry of SB RAS, Novosibirsk, Russia

Abstract

A classification system for hyperspectral images using convolutional neural networks is described. A specific network was selected and analyzed. The network parameters, ensured the maximum classification accuracy: dimension of the input layer, number of the layers, size of the fragments into which the classified image is divided, number of learning epochs, are experimentally determined. High percentages of correct classification were obtained with a large-format hyperspectral image, and some of the classes into which the image is divided are very close to each other and, accordingly, are difficult to distinguish by hyperspectra.

Keywords

Hyperspectral images, convolutional neural networks, deep learning, principal components, probability of correct classification.

1. Introduction

Classification of the land areas is gaining in importance for a wide variety of applications, and one of the most effective systems of classification features is hyperspectral data. It is known that the greatest advances in the field of recognition in recent years have been obtained using deep learning and convolutional neural networks. This report examines exactly this problem. The most important question in this case — what features to use at the input of the neural network. Earlier it was shown that a high probability of correct classification is possible only using spatial-spectral features. The dimension of the input image of convolutional neural network is limited; therefore, a shortened system of features — the principal components — is formed from the spectral components. Spatial features are obtained by forming fragments from the resulting system of spectral features, and the methods of this formation significantly affect the quality of the classification. The analysis of these methods is the main subject of research in this report.

2. Description of the analyzed object

The object that will be investigated in this report has appeared in many publications (Figure 1) [1, 2, 3]. The reason for this is its unique properties: it is a satellite image of a sufficiently large size

SDM-2021: All-Russian conference, August 24-27, 2021, Novosibirsk, Russia

[🛆] kozik@iae.nsk.su (V. I. Kozik); nejevenko@iae.nsk.su (E. S. Nezhevenko)

^{© 0 2021} Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)



Figure 1:

Figure 2:

 $(1408 \times 614 \text{ pixels})$, pixel size is 20 m, and each pixel is characterized by 220 spectral components in the range of 0.4–2.5 μ m. A hyperspectral image of the site obtained within the framework of the AVIRIS program (Airborne Visible Infrared Imaging Spectrometer) at the Indian Pice test site (Indiana, USA) [4]. Figure 2 shows in pseudo colors the markup of this GSI into classes. There are 57 classes in total. However, the specificity of the spatial processing method we have chosen is such that in some areas classification cannot be formed due to the small size of these areas.

3. Convolutional network used for classification

Works in the direction of using neural networks for GSI classification have already been carried out [5, 6, 7], in our report another object is processed and a different method of element extraction is used. Currently, a huge number of networks have been published, designed to classify a wide variety of objects.

The use of networks pre-trained on millions of data is recommended. However, we have a special case. Our training and recognizable images are small terrain fragments that cover the marked (i.e. classified) areas of the GSI. Therefore, we will use a neural network that is not too complex and without such layers as Max Pooling, Dropout, etc. [8]. This network is shown in Figure 3. The network contains an input layer, convolutional layers, and a fully connected layer.



Figure 3:

We will not describe the functioning of the network, this is submitted in detail in the literature. Let's define the parameters of the network, and the most important of them is the character of the input signal. As such, a cube $M \times N \times F$ is selected, where $M \times N$ is the size of the fragment cut out from the 1408×614 image and shifted throughout this image, and F is the number of spectral features characterizing each pixel. As mentioned in the description of the object, the number of spectral components is 220, however, there are highly correlated components among them. As it's known from the theory of pattern recognition, using of correlated features reduces the correctness of recognition, therefore, for effective recognition, as a rule, decorrelation of features is carried out.

The most effective way to do this is by converting the array of spectral features to the principal components. The number of principal components will be determined by analyzing the "rocky talus" — the graph of eigenvalues decreasing. It is presented in [1]. It can be seen that already the 5th eigenvalue is 1/500 of the first value, this means that it accounts for 0.2% of the variance of the spectral components, therefore, most of the experiments will be carried out with the number of principal components equal to 5. Thus, based on the foregoing above, our classification system is a 3D convolutional neural network, the dimension of the input layer is 5, the dimension of the input signal is $M \times N \times 5$, total number of layers is 13. Subsampling is not used in our network, since the classified images are already relatively small size. The dimension of the output layer is equal to the number of classes.

The most important role is played the MxN parameter, the dimension of the fragment cut out from the input layer. Too small size of the fragment will not allow revealing its spatial features. A large size of fragments reduces their number in the class, since the areas belonging to the classes have an arbitrary shape, as a rule, curvilinear, and the fragments are rectangular, so that too few (and sometimes no one) fragments fall on some classes. Thus, it is necessary to find a compromise between the size of fragments and their number in an area belonging to one class, and this is the main theme of this report.

Let us explain how the training sample is formed in our case. Its elements are fragments of the GSI, divided into marked areas. Each area contains its own number of elements, depending on the size of the section and its configuration. When forming a sample, all GSI is covered with square fragments of a given size, and if all the pixels of a fragment belong to the same class, this fragment is considered as an element of the corresponding class.

4. GSI classification experiments

The sequence of classification stages.

- 1. The principal components of the GSI are calculated.
- 2. Directories of classes from 1 to 57 are formed.
- 3. From the file containing the GSI separation into classes (Figure 2), using a sliding window of size $M \times N$ and shift_M, shift_N, fragments are selected, all elements of which belong to the same class. Classes, the number of fragments of which exceeds the specified threshold, participate in training.
- 4. Network parameters are adjusted: number of layers, kernels size, number of feature maps, number of classes.
- 5. Parameters of the training procedure are adjusted: numbers of classes, number of training epochs, objects of each class are divided into training and validation samples (as a rule, in a ratio of 7 : 3).
- 6. Training procedure is started.

Trained network is visualized in Figure 4.

The number of weights adjusted as a result of training is 158184. There are 3 convolution layers in the network, 3 layers of normalization (batchnorn layers), which speed up the learning procedure; three activation layers (ReLu layers), which perform nonlinear transformation, and softmax and classoutput layers, provided the recognition procedure. Using the input layer (imageinput), training and recognizable images are introduced.

nalysis date: 01-Jun-2021 15:46:51				13 i layers	0 🛕 0 🃢 warnings error
	ANALYSIS RESULT				
imageinput		Name	Туре	Activations	Learnables
	1	imageinput 12x12x5 images with 'zerocenter' normalization	Image Input	12×12×5	-
• conv_1	2	conv_1 8 3x3x5 convolutions with stride [1 1] and paddin	Convolution	12×12×8	Weights 3×3×5×8 Bias 1×1×8
• batchnorm_1	3	batchnorm_1 Batch normalization with 8 channels	Batch Normalization	12×12×8	Offset 1×1×8 Scale 1×1×8
• relu_1	4	relu_1 ReLU	ReLU	12×12×8	-
conv_2	5	conv_2 16 3x3x8 convolutions with stride [1 1] and paddi	Convolution	12×12×16	Weights 3×3×8×16 Bias 1×1×16
• batchnorm_2	6	batchnorm_2 Batch normalization with 16 channels	Batch Normalization	12×12×16	Offset 1×1×16 Scale 1×1×16
• relu_2	7	relu_2 ReLU	ReLU	12×12×16	-
• conv_3	8	conv_3 32 3x3x16 convolutions with stride [1 1] and padd	Convolution	12×12×32	Weights 3×3×16×3 Bias 1×1×32
• batchnorm_3	9	batchnorm_3 Batch normalization with 32 channels	Batch Normalization	12×12×32	Offset 1×1×32 Scale 1×1×32
relu_3	10	relu_3 ReLU	ReLU	12×12×32	-
• fc	11	fc 33 fully connected layer	Fully Connected	1×1×33	Weights 33×4608 Bias 33×1
	12	softmax softmax	Softmax	1×1×33	-
• softmax	13	classoutput crossentropyex with '02' and 32 other classes	Classification Output	-	-



Let's consider the results of experiments. Note, that the only criterion of the classification effectiveness is the classification accuracy, which is defined as the ratio of correctly objects classified number to the total number of objects (this term – accuracy – is used along with the term "probability of correct classification"). Let us note the feature of our method of a training sample forming, training and classification. With different sizes of fragments into which sections of classes are divided, and limiting the number of elements in a class, the number of classes will be different, which will not allow determining the actual dependence of the classification accuracy on the size of fragments, since two factors affect here: the size of the fragment and the number of classes. Therefore, we calculated the number of classes (18) for the maximum fragment size of 16x16 and then trained the network for all fragments size with this number of classes 18 is shown in Figure 5.

It can be seen from this graph that the optimal fragment size is near 14×14 .

A very important factor affecting the classification accuracy is the dimension of the input layer, which is equal to the number of principal components used at the input. This dependence is presented in Table 1.

It can be seen that, starting with 5 principal components, the classification accuracy increases insignificantly.

A very important parameter of the network is the number of learning epochs. Dependence of classification accuracy on this parameter for a fragment 14×14 and the number of principal components -5 is shown in Figure 6.

The classification accuracy monotonically increases with the number of epochs, taking a sharp jump from 20 to 30 epochs, although this function depends on the size of the fragments.



Figure 5:

Accuracy
0.253509
0.995038
0.997448
0.99844

rigure J.

Table 1



88 86 11 12 13 14 Number of layers

Figure 8:



Figure 9:

Class number	Class name	Number of classes	Accuracy
2	Buildings	2621	0.9987
4	Corn	2269	0.9927
7	Corn-EW	169	1
8	Corn-NS	368	1
9	Corn-CleanTill	2481	0.997
10	Corn-CleanTill-EW	4241	0.9914
12	Corn-CleanTill-NS-Irrigated	45	1
14	Corn-MinTill	896	0.9963
15	Corn-MinTill-EW	1099	0.997
16	Corn-MinTill-NS	93	1
17	Corn-NoTill	30	1
18	Corn-NoTill-EW	1304	1
21	Grass	91	0.963
26	Hay	443	0.9925
27	Hay-Alfalfa	191	1
30	Not cropped	70	1
33	Orchard	1996	1
35	Pond	326	1
38	Soybeans-NS	227	1
39	Soybeans-CleanTill	239	0.9722
40	Soybeans-CleanTill	2000	0.99
41	Soybeans-CleanTill-EW	1057	0.9905
42	Soybeans-CleanTill-NS	40	0.6667
44	Soybeans-CleanTill Weedy	1046	0.9777
45	Soybeans-Drilled	50	1
46	Soybeans-MinTill	65	1
47	Soybeans-MinTill-EW	689	1
48	Soybeans-MinTill-Drilled	721	1
49	Soybeans-MinTill-NS	185	1
50	Soybeans-NoTill	356	1
52	Soybeans-NoTill-NS	436	1
56	Trees	636	1
57	Wheat	8115	0.9988

Tab	le 2
-----	------

For the fragment 14×14 , it is already at 30 epochs actually comes out in saturation, and for a 5x5 size the classification accuracy continues to grow even at 50 epochs, as can be seen from Figure 7.

The classification accuracy also depends on the number of layers. In the previous work [1], only convolutional layers, input and output, were considered. Since here the entire network is shown, let's consider the dependence on the total number of layers. This dependence for a 14×14 fragment is shown in Figure 8. As follows from the graph, the optimal number of layers is 13.

So, we have chosen the following network parameters: fragment size -14×14 , number of principal components -5, number of network layers 13, number of learning epochs 50.

Classification results with the number of classes 33, (as seen in the learning function shown in Figure 9) - 99.72%, which, in our opinion, is a very good result.

The resulting table with the number of elements in each class and the probability of class recognition is shown in Table 2.

It can be suggested, that high values of classification accuracy are obtained due to overfitting of the neural network. This is an undesirable phenomenon that occurs under solving learning problems by precedents, when the probability of the trained algorithm average error on the test sample is significantly higher than the error on the training sample. From Figure 9 (bottom part), characterizes the behavior of the error in the learning process, follows that the error on the test sample is very insignificantly (by a part of a percent) higher than the error on the training sample, which means that there is no overfitting in this case. Table 2 shows the classification results, indicating the probabilities for each class.

From the class names it is clear that we did not integrate closely related classes into one (for example, crops of corn, crops of soybeans), as it was done in other publications [9, 10]. It is clear that it is much easier to distinguish corn crops from buildings than to distinguish between different planting options for the corn or soybeans for different types of plowing. Note that with fragments of 14×14 , almost indistinguishable objects – crops of corn, soybeans – are classified with a very high (often 100%) probability.

It should be said that the results obtained in this work significantly exceed the results of [11], with one caveat: the latter does not contain the problem of covering an area belonging to a class by rectangular windows; therefore the regions with a complex configuration can be classified there.

5. Conclusion

Thus, in this report we have analyzed the influence of the neural networks parameters on the accuracy of hyperspectral images classification. The network parameters and methods of forming a training sample are selected, which provide a very high classification accuracy (integral accuracy is 99.72%), and such a high accuracy is ensured on close classes (11 types of corn plowing, 14 types of soybean plowing). Analyzing such high classification accuracy, the following should be said. This is largely due to the way as the training and validation samples are formed, characterized by their very close mixing. At the same time, it is obvious that this method shows a certain limit of classification accuracy, from which it's possible to deviate, for example, by increasing the fragments coverage step or forming the training and validation samples spatially separated.

References

- [1] Kozik V.I., Nezhevenko E.S. Classification of hyperspectral images using convolutional neural networks // Avtometriya. 2021. No. 1. P. 13–21.
- [2] Borzov S.M., Potaturkin O.I. Spectral-spatial methods of classification of hyperspectral images, a review // Avtometriya. 2018. Vol. 54. No. 6. P. 64–86.

- [3] Nezhevenko E.S., Feoktistov A.S. Investigation of the efficiency of neural network classification of hyperspectral images using the Hilbert – Huang transform // Collection of Articles Based on the Materials of the International Scientific Congress "Interexpo Geo-Siberia". Novosibirsk, April 18–22, 2016. Vol. 1. P. 60–64.
- [4] Nezhevenko E.S, Feoktistov A.S, Dashevsky O.Yu. Neural network classification of hyperspectral images based on the Hilbert – Huang transform // Avtometriya. 2017. Vol. 53. No. 2. P. 79–85.
- [5] Baumgardner M. F., Biehl L. L., Landgrebe D. A. 220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3. Purdue University Research Repository. 2015. doi:10.4231/R7RX991C.
- [6] Audebert N., Saux B., Lefèvre S. Deep learning for classification of hyperspectral data: A comparative review // Geoscience and Remote Sensing Magazine. IEEE, 2019. Vol. 7. No. 2. P. 159–173.
- [7] Li Y., Zhang H., Shen Q. Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network // Remote Sensing. 2017. Vol. 9. No. 67. P. 1–21, DOI:10.3390/rs9010067.
- [8] Krizhevsky A. Learning multiple layers of features from tiny images. Master's Thesis, Department of Computer Science, University of Toronto, 2009.
- [9] Borzov S.M, Potaturkin O.I Research of the efficiency of spectral-spatial classification of hyperspectral observation data // Avtometriya. 2017. Vol. 53. No. 1. P. 32–42.
- [10] Borzov S.M., Potaturkin O.I. Classification of hyperspectral images with different methods of forming training samples // Avtometriya. 2018. Vol. 54. No. 1. P. 89–97.
- [11] Nezhevenko E.S. Neural network classification of difficult to distinguish types of vegetation byhyperspectral features // Avtometriya. 2019. No. 3. P. 62–70.