A Two-Branch Neural Network Based on Superpixel Segmentation and Auxiliary Samples

Zhidong Dong, Caihong Mu, Haikun Yu, Yi Liu

A Two-Branch Neural Network Based on Superpixel Segmentation and Auxiliary Samples

Zhidong Dong[1], Caihong Mu[1][0000-0003-4373-3661], Haikun Yu[1] and Yi Liu[2]

[1] Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an 710071, China
[2] School of Electronic Engineering, Xidian University, Xi'an 710071, China
mucaihongxd@foxmail.com; yiliuxd@foxmail.com

Abstract. Existing hyperspectral image (HSI) classification methods generally use the information in the neighborhood of the samples but seldom utilize the regional homogeneity of the ground objects. We propose a two-branch neural network based on superpixel segmentation and auxiliary samples (TBN-SPAS) for HSI classification. TBN-SPAS uses superpixel segmentation to find samples within the superpixel, which have high spatial correlation with the sample to be classified. Then TBN-SPAS further selects samples from the samples within the superpixel as auxiliary samples, which have high spectral similarities with the sample to be classified. Finally, the neighborhood patch of the preprocessed HSI and the corresponding sorted auxiliary samples are input into a two-branch neural network for feature extraction and classification. TBN-SPAS achieves significantly better classification results compared with several state-of-the-art methods.

Keywords: Hyperspectral image classification, Superpixel segmentation, Two-branch neural network

1 Introduction

Hyperspectral images (HSIs) have rich spatial and spectral information and can be used for the detection or identification of ground objects, so they have been widely used in military target identification, geological resource detection, agricultural crop monitoring, archaeological relics restoration and other fields [1][2].

In recent years, deep learning techniques have been applied to HSI classification, such as stacked autoencoders [3], deep belief networks [4], and convolutional neural networks (CNNs) [5][6], etc., among which CNN-based HSI classification methods have been the most widely used. The CNN-based methods use the neighborhood block of the samples as input, and perform feature extraction through two-dimensional (2D) convolution and three-dimensional (3D) convolution. They can extract spatial features, spectral features or combined spatial-spectral features from the neighborhood patch, which greatly improves the classification accuracy of HSIs.

Hamida et al. [7] studied the classification effects of different 3D convolutional networks on hyperspectral images, and designed a network consisting of 4 layers of 3D

convolution and 1 fully connected layer (3DCNN) for feature extraction. The model achieved high classification accuracy. Zhong et al. [8] proposed the spectral-spatial residual network (SSRN). Referring to the residual structure, they designed a spatial feature extraction module and a spectral feature extraction module with multiple 3D convolutional layers, which also achieved high classification accuracy. Roy et al. [9] proposed the hybrid spectral convolutional neural network (HybridSN), which first extracted spatial spectral features to obtain feature maps by using multiple 3D convolutional layers, and then extracted spatial features from the obtained feature maps by using 2D convolutional layers, reducing the complexity of the model.

The above methods achieved good classification results, however, they made more use of the spectral features of HSIs and extract spatial-spectral joint features from the neighborhood patch of the input samples for classification. In reality, the distribution of ground objects is spatially continuous. Especially in HSI images with high spatial resolution, ground objects tend to be distributed in a large area with irregular shapes, the size of which generally exceeds the range of neighborhood patch. When the classifier uses the neighborhood patch as the input, it can only obtain the information within the neighborhood, and the distribution information of the objects beyond the neighborhood is not used. Therefore, some researchers proposed HSI classification methods that utilized a wider range of spatial information, such as introducing a wider range of spatial information through superpixel segmentation.

Superpixel segmentation is a commonly used unsupervised image segmentation method that divides adjacent pixels in an image into multiple disjoint regions, called superpixels [10]. Pixels belonging to the same superpixel in an image often have similar features such as texture, brightness, and color [11]. The purpose of superpixel segmentation is to achieve the following two effects: each superpixel contains only one class of objects; the set of superpixel boundaries is a superset of the object boundaries [12].

Jiang et al. [13] proposed a superpixel-based principal component analysis (PCA) method (SuperPCA) for HSI classification. This method first performed superpixel segmentation on the first principal component of the HSI at different scales, and applied PCA method to samples inside the obtained superpixels to reduce the data dimension. Then the classifier was trained on the data of each scale, and the final classification result was obtained through decision fusion. This method achieved high classification accuracy even with limited samples. However, the utilization of neither the superpixel segmentation results nor the regional homogeneity of the ground objects was sufficient. Samples belonging to the same superpixel have higher distribution consistency in space and are more likely to belong to the same class. This fact can provide more auxiliary information for HSI classification. Therefore, we propose a two-branch neural network based on superpixel segmentation and auxiliary samples (TBN-SPAS) for HSI classification. TBN-SPAS makes full use of superpixel segmentation and spectral similarities to obtain the auxiliary samples that have high spatial correlation and spectral similarities with the sample to be classified, and then uses a two-branch neural network for feature extraction and classification. Experimental results on two HSI datasets demonstrate that TBN-SPAS achieves better classification results compared with several state-of-the-art methods.

2 Proposed method

2.1 Selection of auxiliary samples

We first reduce the dimension of the HSI by PCA, and obtain the first principal component (FPC) of the HSI. Suppose $\boldsymbol{x_i}$ is the sample to be classified. We perform the entropy rate superpixel segmentation (ERS) [12] on the FPC, and find samples around $\boldsymbol{x_i}$ within the obtained superpixels, which can ensure that these samples have high spatial correlation with $\boldsymbol{x_i}$. Since the spatial distribution of ground objects in the HSI may be large, the obtained superpixels may contain many samples. To ensure the computational efficiency and avoid introducing too much redundant information, we further select some samples within the superpixel as the final auxiliary samples by measuring the spectral similarities between these samples and $\boldsymbol{x_i}$.

We use the cosine distance to measure the spectral similarities of the samples. For other samples in the superpixel where $\boldsymbol{x_i}$ is located, we sort them in ascending order according to the cosine distances between their spectral vectors and the spectral vector of $\boldsymbol{x_i}$, and select a certain number (denoted as m) of samples at the top of the ranking list as auxiliary samples, which will be introduced into the process of classification.

2.2 The structure of TBN-SPAS

We design a two-branch neural network called TBN-SPAS. The main branch takes the preprocessed neighborhood patch of $\boldsymbol{x_i}$ as input. The input of the auxiliary branch is the 2D data composed of m sorted auxiliary samples. The sorted auxiliary samples have sequential features that can be learned by the network.

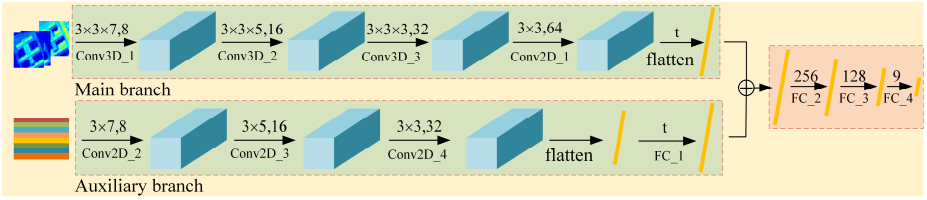


Fig. 1. The structure of the two-branch neural network based on superpixel segmentation and auxiliary samples (TBN-SPAS)

The structure of TBN-SPAS is shown in Fig. 1. The main branch includes three cascaded 3D convolutional layers and one 2D convolutional layer, namely Conv3D_1, Conv3D_2, Conv3D_3 and Conv2D_1. The auxiliary branch includes three 2D convolutional layers and one fully connected layer, namely Conv2D_2, Conv2D_3, Conv2D_4 and FC_1. Three fully connected layers (FC_2, FC_3, and FC_4) are used to perform feature fusion on the outputs of the main branch and the auxiliary branch. The network layers Conv3D_1, Conv3D_2, Conv3D_3, Conv2D_1, Conv2D_2, Conv2D_3, Conv2D_4 and FC_1 are all followed by a BatchNorm layer and a rectified linear unit (ReLU) activation layer. The network layers FC_2 and FC_3 are followed

by a ReLU activation layer and a Dropout layer with a dropout rate of 0.4. The network layer FC_4 is followed by a Softmax function to map the network output to a vector of the probability that the sample $\boldsymbol{x}_i$ belongs to each class. In the main branch, t represents the length of the flattened output of Conv2D_1. The number of output neurons of FC_1 in the auxiliary branch is set to t. The hyper-parameters in the network are shown in Fig. 1.

3 Implementation process of TBN-MERS

The overall process of the TBN-SPAS method is as follows.

An HSI is denoted as $\boldsymbol{X} \in \mathbb{R}^{w \times h \times b}$ and the corresponding label map is denoted as $\boldsymbol{Y} \in \mathbb{R}^{w \times h}$, where w、h are the width and height of the HSI, respectively, and $\boldsymbol{b}$ is the number of spectral bands of the HSI. The value of each position y_{ij} in the label map $\boldsymbol{Y}$ is among the set $\{0,1,2,...,c\}$, in which $\boldsymbol{c}$ represents the total number of object classes.

PCA is applied to $\boldsymbol{X}$ to extract the FPC $\boldsymbol{X}_{pca1} \in \mathbb{R}^{w \times h}$, and the values of $\boldsymbol{X}_{pca1}$ are scaled to the interval $[0,255]$. Then ERS is applied to $\boldsymbol{X}_{pca1}$ to obtain the corresponding 2D segmentation map $\boldsymbol{S}_{pca1}$.

We standardize the data of each band of $\boldsymbol{X}$, and denote the preprocessed HSI as $\boldsymbol{X}'$. Each sample $\boldsymbol{x}_i$ consists of two parts of data: (1) the neighborhood patch $\boldsymbol{p}_i^{\mathrm{A}} \in \mathbb{R}^{p \times p \times b}$ taken from $\boldsymbol{X}'$, where p is the length or width of the neighborhood patch, and b is the number of bands of the hyperspectral image; (2) 2D data $\boldsymbol{p}_i^{\mathrm{B}} \in \mathbb{R}^{m \times b}$ composed of auxiliary samples extracted from superpixels, where m is the number of auxiliary samples.

We randomly select a certain number of samples from each class to form the training set, and draw an equal number of samples from the remaining samples of each class to form the validation set. The remaining samples are used as the test set.

We feed the two parts $\boldsymbol{p}_i^{\mathrm{A}}$ and $\boldsymbol{p}_i^{B}$ of each sample $\boldsymbol{x}_i$ into the two branches of TBN-SPAS, and train the network. The model is tested on the verification set every epoch. When the classification accuracy of the model on the validation set no longer rises, the training is completed. The best model on the validation set is used to predict the classes of the samples in test set.

The network is trained with the cross-entropy loss function:

$$L = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{c} y_{ij}\log(p_{ij}) \tag{1}$$

where N is the number of training samples, c is the number of classes, and p_{ij} is the probability predicted by the model that the sample x_i belongs to the class j. If sample $\boldsymbol{x}_i$ belongs to the class j, the value of y_{ij} is 1; otherwise, it is 0.

4 Experiment and Analysis

4.1 Experimental settings

To explore the role of auxiliary branch in TBN-SPAS and verify the performance of the proposed TBN-SPAS, we design the following experiments: (1) verification of the role of auxiliary branch; (2) comparison with other existing methods. The comparison methods used include: 3DCNN [7], SSRN [8], HybridSN [9] and SuperPCA [13]. Indian Pines (IP) and Pavia University (PU) datasets [9] are used for the experiments. Overall classification accuracy (OA), Average classification accuracy (AA) and Kappa coefficient (Kappa) [9] are used to measure the performance of these methods.

In TBN-SPAS, the number of target superpixels in ERS is set to 800, α (a parameter in ERS) is set to 0.5, and the number of auxiliary samples m is set to 8. We randomly select 50 samples from each class (10 samples for the class with less than 50 samples) to form the training set. We randomly select the same number of samples for each dataset as the validation set, and the remaining samples as the test set. The batch size in the experiment is set to 32. We use the Stochastic Gradient Descent (SGD) optimizer, and set the learning rate to 0.0005. In 3DCNN, SSRN, HybridSN, and TBN-SPAS, the patch size of input samples is set to 25×25. The experimental parameters and SuperPCA are consistent with those described in the original paper. All experimental data in this paper are the average of five runs of each method on the RTX TITAN.

4.2 The role of auxiliary branch

We design comparative experiments to verify the effect of auxiliary branch. The comparative experiments include: (1) a single-branch network without auxiliary branches (marked as SBN-XAB), where only the main branch and the classification part in TBN-SPAS are used to form a model and the auxiliary branch is deleted; (2) a two-branch model with the main and the auxiliary branches, that is, TBN-SPAS. The other hyperparameters in the comparative experiments are the same. Table 1 shows the classification results of the comparative experiments.

Table 1. Results of SBN-XAB and TBN-SPAS

Dataset	Metrics	SBN-XAB	TBN-SPAS
IP	OA (%)	93.77	**96.56**
	AA (%)	97.01	**98.34**
	Kappa (×100)	92.88	**96.05**
PU	OA (%)	95.54	**98.96**
	AA (%)	95.21	**99.21**
	Kappa (×100)	94.09	**98.63**

It can be seen in Table 1 that on the IP dataset, compared with the classification results of SBN-XAB, TBN-SPAS achieves an OA with 96.56% that is increased by

2.79%, an AA with 98.34% that is increased by 1.33%, and a Kappa with 0.9605 that is increased by 0.0317. On the PU dataset, compared with the classification results of SBN-XAB, TBN-SPAS achieves the values of OA, AA and Kappa with 98.96%, 99.21% and 0.9863, which are increased by 3.42%, 4.00%, and 0.0454, respectively. Such significant improvement demonstrates that the information of auxiliary samples is very effective in improving the HSI classification effect.

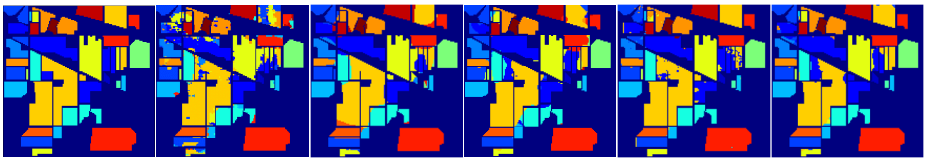
4.3 Comparison with existing methods

We compare TBN-SPAS with four existing methods: 3DCNN [7], SSRN [8], HybridSN [9] and SuperPCA [13].

The experimental results and analysis of the five different methods on the IP and PU dataset are as follows. In the following tables, the best values are marked in bold. The numbers in parentheses are the standard deviations.

Table 2. Experimental results of different methods on IP dataset

Metrics	3DCNN	SSRN	HybridSN	SuperPCA	TBN-SPAS
OA (%)	82.16 (1.47)	88.38 (1.58)	93.77 (1.22)	95.06 (1.24)	**96.56** (0.58)
AA (%)	84.47 (1.58)	75.06 (1.07)	97.01 (0.55)	96.70 (1.00)	**98.34** (0.26)
Kappa (×100)	79.72 (1.65)	86.73 (1.79)	92.88 (1.38)	94.32 (1.42)	**96.05** (0.67)



(a) (b) (c) (d) (e) (f)

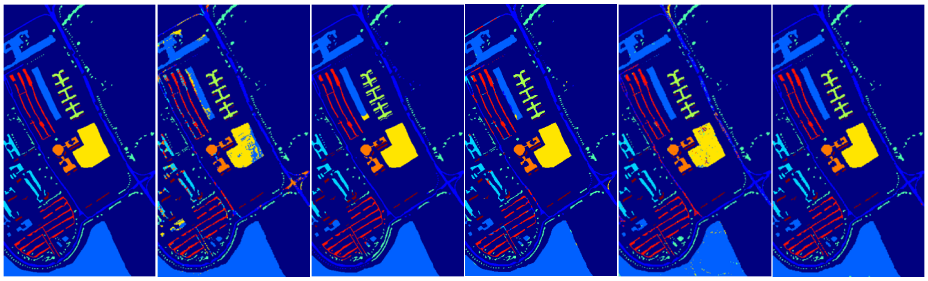
Fig. 2. The ground-truth label map of IP and the classification result maps of different methods (a) Ground-truth (b) 3DCNN (c) SSRN (d) HybridSN (e) SuperPCA (f) TBN-SPAS

(1) Results and analyses of IP: Table 2 shows the three metrics of the five methods on IP. From Table 2, it can be seen that TBN-SPAS achieves an OA of 96.56%, which is increased by 14.40%, 8.18%, 2.79%, 1.50% over 3DCNN, SSRN, HybridSN and SuperPCA, respectively. TBN-SPAS achieves an AA of 98.34%, which is increased by 13.87%, 23.28%, 1.33%, 1.64% over 3DCNN, SSRN, HybridSN and SuperPCA, respectively. A higher AA indicates that TBN-SPAS is good at classifying difficult classes on IP datasets, and also indicates that TBN-SPAS has better generalization. TBN-SPAS also achieves the best results in terms of Kappa. The standard deviation of each metric is low, indicating that TBN-SPAS has better robustness on the IP dataset. Fig. 2 shows the ground-truth label map of IP and the classification results of different methods, demonstrating the effectiveness of TBN-SPAS.

(2) Results and analyses of PU: Table 3 shows the OA, AA, and Kappa of the five different methods on PU. From Table 3, it can be seen that TBN-SPAS achieves an OA of 98.96%, an AA of 99.21% and a Kappa of 98.63%. Compared with other comparison methods, TBN-SPAS has achieved a great improvement of more than 3% in OA and AA, which shows that the introduction of auxiliary samples and the use of two branch network are effective. Fig. 3 shows the ground-truth label map of PU and the classification results of different methods, which also demonstrates the effectiveness of TBN-SPAS.

Table 3. Experimental results of different methods on PU dataset

Metrics	3DCNN	SSRN	HybridSN	SuperPCA	TBN-SPAS
OA (%)	88.38 (0.99)	95.08 (1.27)	95.54 (0.71)	93.24 (0.67)	**98.96** (0.73)
AA (%)	85.32 (0.94)	91.51 (1.00)	95.21 (0.70)	94.42 (0.37)	**99.21** (0.18)
Kappa (×100)	84.69 (1.25)	93.47 (1.66)	94.09 (0.94)	91.10 (0.85)	**98.63** (0.96)



(a) (b) (c) (d) (e) (f)

Fig. 3. The ground-truth label map of PU and the classification result maps of different methods (a) Ground-truth (b) 3DCNN (c) SSRN (d) HybridSN (e) SuperPCA (f) TBN-SPAS

5 Conclusions

We design a two-branch neural network based on superpixel segmentation and auxiliary samples (TBN-SPAS) for HSI classification. TBN-SPAS obtains the auxiliary samples that have high spatial correlation and spectral similarities with the sample to be classified by fully utilizing superpixel segmentation and spectral similarities, which are sorted and then input together with the neighborhood patch of the preprocessed HSI into a two-branch neural network for further feature extraction and classification. Experimental results demonstrate that TBN-SPAS achieves significantly better classification results compared with several state-of-the-art methods, indicating that the idea of selecting auxiliary samples through superpixel segmentation and spectral similarities is effective. In the future, we will study more effective way of making use of superpixel segmentation.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 62077038 and 61672405).

References

1. Li S, Song W, Fang L, et al.: Deep learning for hyperspectral image classification: An overview. IEEE Transactions on Geoscience and Remote Sensing. 57(9), 6690-6709 (2019).
2. Camps-Valls G, Tuia D, Bruzzone L, et al.: Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. IEEE signal processing magazine. 31(1), 45-54 (2013).
3. Chen Y, Lin Z, Zhao X, et al.: Deep learning-based classification of hyperspectral data. IEEE Journal of Selected topics in applied earth observations and remote sensing. 7(6), 2094-2107 (2014).
4. Chen Y, Zhao X, Jia X. Spectral–spatial classification of hyperspectral data based on deep belief network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 8(6), 2381-2392 (2015).
5. Hu W, Huang Y, Wei L, et al.: Deep convolutional neural networks for hyperspectral image classification. Journal of Sensors. 2015, (2015).
6. Cao X, Zhou F, Xu L, et al.: Hyperspectral image classification with Markov random fields and a convolutional neural network. IEEE Transactions on image processing. 27(5), 2354-2367 (2018).
7. Hamida A B, Benoit A, Lambert P, et al.: 3-D deep learning approach for remote sensing image classification. IEEE Transactions on geoscience and remote sensing. 56(8), 4420-4434 (2018).
8. Zhong Z, Li J, Luo Z, et al.: Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework. IEEE Transactions on Geoscience and Remote Sensing. 56(2), 847-858 (2017).
9. Roy S K, Krishna G, Dubey S R, et al.: HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters. 17(2), 277-281 (2019).
10. Ren X, Malik J.: Learning a classification model for segmentation. In: Computer Vision IEEE International Conference, pp. 10-10, IEEE Computer Society, (2003).
11. Achanta R, Shaji A, Smith K, et al.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE transactions on pattern analysis and machine intelligence. 34(11), 2274-2282 (2012).
12. Liu M, Tuzel O, Ramalingam S, et al.: Entropy rate superpixel segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2097-2104, IEEE, (2011).
13. Jiang J, Ma J, Chen C, et al.: SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing. 56(8), 4581-4593 (2018)