Proceedings of the 37th Chinese Control Conference
July 25-27, 2018, Wuhan, China

Hyperspectral Remote Sensing Image Classification Based on Convolutional Neural Network

Xiangyang Dai[1,2], Wei Xue[1,2]

1. School of Automation, China University of Geosciences, Wuhan 430074

2. Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074

E-mail: 2485952510@qq.com, xw12982044@163.com

Abstract: Remote sensing hyperspectral imaging can obtain rich spectral information of terrestrial objects, which allows the indistinguishable matter in the traditional wideband remote sensing to be distinguished in hyperspectral remote sensing. Hyperspectral image has the characteristics of "combining image with spectrum". Making full use of spectral information and spatial information in hyperspectral image is the premise of obtaining accurate classification results. At present, most of hyperspectral data feature extraction algorithms mainly utilize local spatial information in the same channel and spectral information in the same spatial location of different channels. However, these methods require a large amount of prior knowledge, it is difficult to fully grasp the hyperspectral data of all spatial and spectral information, and the model generalization ability is poor. With the development of deep learning, convolutional neural network shows superior performance in all kinds of visual tasks, especially in the two-dimensional image classification, and could get a high classification accuracy. In this paper, an image classification method based on three-dimensional convolution neural network is proposed based on the structural properties of hyperspectral data. In the proposed method, first the stereo image blocks of hyperspectral data are intercepted, then multi-layer convolution and pooling operation of extracted blocks by convolutional neural network are implemented to obtain the essential information of hyperspectral data, finally the classification of hyperspectral data is completed. The experimental results show the proposed method could provide better feature expression and classification accuracy for hyperspectral image.

Key Words: Convolutional Neural Network, Hyperspectral Image, Deep Learning Classification

1 Introduction

After the development of the second half of the twentieth century, significant changes have taken place in theory and application of remote sensing technology. Among them, hyperspectral imagery remote sensing technology is undoubtedly a very important aspect in the major changes of remote sensing technology. Hyperspectral remote sensing technology uses hyperspectral sensor or imaging spectrometer to simultaneously image the target area in tens to hundreds of continuous subdivision bands, combine the image with the spectrum, which obtain the spatial and spectral information and get the pixel information in pixels Hyperspectral Imagery (HSI) [1,2]. Hyperspectral images provide abundant spectral information and occupy an important position in remote sensing earth observation system and are widely used in many fields such as modern military, precision agriculture and environmental monitoring [3,4]. With the further development of hyperspectral imagers, the amount of information contained in hyperspectral images will be even greater, and the application of hyperspectral images will be more extensive. In different applications, larger and larger amounts of data also pose more complex requirements for hyperspectral remote sensing to ground observation technology. Hyperspectral image classification technology is an important part of hyperspectral remote sensing technology for ground observation. Its specific task is to classify the target represented by each pixel in hyperspectral image. However, classification of hyperspectral data remains a challenging task because of the large number of spectral channels, limited training samples, and large spatial variability [5].

In the early hyperspectral image classification technologies, the hyperspectral image classification method only used the rich spectral information in hyperspectral images and did not dig deeper into the intrinsic information of the data. For example, distance classifier [6], K nearest neighbor classifier [7], maximum likelihood classifier [8] and Rogers regression [9]. Most of these methods are affected by Huygens phenomenon [10]. That is to say, when the training data is limited, if the data dimension is too high, the classification accuracy will greatly decrease. In recent years, with the continuous updating of feature extraction and classification methods, many methods such as spectral space classification and local Fisher discriminant [11] have been proposed and achieve good classification results. One representative method is Support Vector Machines (SVM) [12]. SVM is a kind of new statistical learning algorithm with the advantages of high precision, fast computing speed and strong generalization ability. The main idea of SVM is to transform the linearly inseparable problem of low-dimensional space into high-dimensional space for accurate classification. Although the kernel transformation method achieves a satisfactory classification accuracy to a certain extent, it is still difficult to choose the combination of the kernel function and the optimal parameters.

Compared with the traditional methods, deep learning technology such as Convolutional Neural Network (CNN) has good performance in image classification and pattern recognition. In recent years, with the development of Neural Network (NN), CNN is more and more widely applied to the classification of remote sensing data, such as Multilayer Perceptron (MLP) [14] and Radial Basis Function (RBF)

*This work is supported by the Fundamental Research Funds for the Central Universities under Grants CUGL170211, CUG2018JM15.

[15]. RATLE F et al [16] proposed a semi-supervised neural network structure for hyperspectral image database classification. In fact, in remote sensing image classification, SVM is superior to the traditional neural network in terms of classification accuracy and computational cost, but the neural network is still a powerful tool. HINTON G E et al [17] proposed a deep neural network for data reduction, classification and object recognition. In addition to image-related issues, it can also be used in areas such as speech recognition. CHEN Y et al [18] proposed an expand hyperspectral image classification method using Deep Neural Network (DNN).

CNN provides better classification effect in visual field than traditional SVM classifier [19] and DNN [20]. However, the related research does not directly apply the CNN method to the classification of hyperspectral images. In this paper, an appropriate network structure is established for effective classification of hyperspectral data. Experimental results show that the proposed network structure can promote the classification of hyperspectral images and provide higher classification accuracy.

2 Convolution Neural Network

Initially inspired by the neural mechanism of the visual system, CNN is a multi-layered perceptron designed for recognizing two-dimensional shapes with a high degree of invariance under translational conditions and a certain invariance under zooming and tilting [21]. The four basic features of CNN are local interconnection, weight sharing, down sampling, and using multiple convolutional layers, all of which ensure that CNN is not affected by the "gradient dispersion" of multi-layer feedforward neural networks under the deep network model.

2.1 Basic Structure of CNN

As shown in Fig. 1, a typical CNN consists of an input layer, a convolution layer, a down sampling layer, a fully connected layer and an output layer.

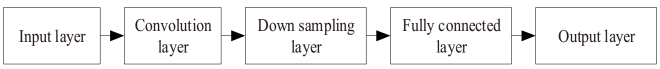

Fig. 1: Typical CNN structure

The input of a typical CNN is usually the original image X , and the $j\ th$ feature map of the $i\ th$ layer is $H_{i,j}(H_{0,1} = X)$.

(1) When the $i\ th$ layer is convolution layer:

$$H_{i,j} = f_i\left(\sum_{k=1}^{n_{i-1}} H_{i-1,j} \otimes W_{k,j}^i + b_j^i\right) \tag{1}$$

Where $\otimes$ represents convolution operation; $W_{k,j}^i$ denotes convolution kernel between the $k\ th$ feature map of the $i-1\ th$ layer and the $j\ th$ feature map of the $i\ th$ layer; b_j^i is the bias of the $j\ th$ feature map of the $i\ th$ layer ; n_{i-1} denotes the number of feature map of the $i-1\ th$ layer; $f_i(\cdot)$ represents the activation function of the $i\ th$ layer。

(2) If the $i\ th$ layer is down sampling layer:

$$H_{i,j} = g_i\left[\alpha_j^i down\left(H_{i-1,j}\right) + \beta_j^i\right] \tag{2}$$

Where α_j^i is the coefficient of the $j\ th$ feature map of the $i\ th$ layer; β_j^i is the bias of the $j\ th$ feature map of the $i\ th$ layer; $g_i(\cdot)$ represents the activation function of the $i\ th$ layer， generally choose a linear function as its activation function, that is, $g_i(x) = x$.

(3) If the $i\ th$ layer is fully connected layer:

If the layer is the first fully connected layer that connects the down sampling or convoluted layer, first all the top-level feature maps $H_{i-1,j}$ should be flatten, that is, they are converted into one-dimensional vectors I_{i-1} . Then,

$$I_i = f_i\left(W^i I_{i-1} + b^i\right) \tag{3}$$

Where W^i is the connection weight matrix of the $i\ th$ layer; b^i is the bias of the $i\ th$ layer; $f_i(\cdot)$ represents the activation function of the $i\ th$ layer.

(4) If the $i\ th$ layer is output layer:

The layer is the same as fully connected layer, the softmax classifier can also be chosen as the output layer, then the appropriate loss function is selected as the CNN objective function, denoted as $L(W,b)$. Where, W is the weight of all connections in the network, b represents all offsets in the network.

2.2 CNN Classifier

Hyperspectral image data of three-dimensional structure contains both spatial information and rich spectral information. Neither the spectral feature nor the spatial feature-based classification method fully exploits the advantages of hyperspectral images. Compared with these two methods, the combination of spatial information and spectral information for hyperspectral image classification is more consistent with the characteristics of hyperspectral and beneficial to improve the accuracy of spectral image classification. In the processing of three-dimensional data, 2D-CNN can not meet the requirements. Ji proposed 3D-CNN [22] and applied it to human motion recognition. Molchanov also achieved good results by using the 3D-CNN recognition gesture [23]. For such three-dimensional data, 3D-CNN is more suitable for the representation than 2D-CNN, hyperspectral image data can be considered as three-dimensional data when joint spectral-space information is classified, so it is feasible to classify it using 3D-CNN. The basic structure of 3D-CNN is similar to that of 2D-CNN, except that both the convolution kernel and the feature map are three-dimensional. In this paper, a 3D-CNN method is proposed for hyperspectral image-space joint classification, the principle of proposed method is shown in Fig.2. The method consists of five layers: input layer, convolution layer, maximum pooling layer, full connection layer and output layer.

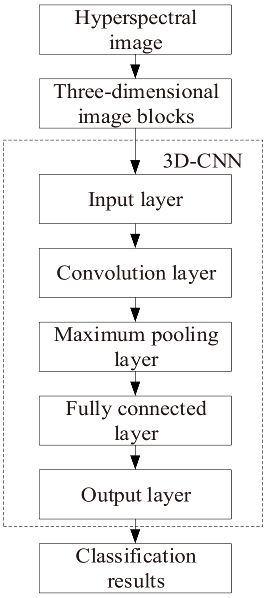

Fig. 2: 3D-CNN classification structure diagram

The models of the 3D-CNN layers are as follows:

(1) Convolutional Layer. It is used for feature extraction. The input image and the corresponding position elements of the three-dimensional convolution kernel are multiplied and then summed to obtain a three-dimensional feature map. Different convolution kernel will get different output data, such as color depth, contour. This is equivalent to extracting different desired convolutional kernels with specific information about the image if different features of the image is extracted.

(2) The maximum pooling layer. It is used to compress the input feature map. On the one hand, the pooling layer makes the feature map smaller and simplifies the network computing complexity; on the one hand, the pooling layer performs feature compression and main features extraction. There are two kinds of pooling operations, one is the average pool and the other is the maximum pool. The maximum pooling used in this paper is to find the maximum value in each region as the new eigenvalue.

(3) Full connection layer. It is used to connect all the features and output value to the classifier.

2.3 CNN Training Algorithm

The training algorithm mainly studies how to train each coefficient in the network structure. Before starting to train the CNN structure, each coefficient should be given an initial random value in the range [-0.05,0.05]. The training process mainly consists of forward propagation and backward propagation.

Forward propagation mainly calculates the real classification effect of the network structure for the current coefficients.

$$x^{i+1} = f_i(u^i) \tag{4}$$

$$u^i = W^i x^i + b^i \tag{5}$$

Where x^i is the input of the i th layer; W^i is the weight vector of the i th layer, which acts on the input data; b^i is the additional bias vector of the i th layer; $f_i(\cdot)$ represents the activation function of the i th layer.

Back propagation iteratively update the coefficients through the network output and the expected comparison, and the output results are close to the expected value. Where, the learning rate α controls the intensity of back propagation:

$$W^i := W^i - \alpha \frac{\partial L(W,b)}{\partial W^i} \tag{6}$$

$$b^i := b^i - \alpha \frac{\partial L(W,b)}{\partial b^i} \tag{7}$$

As the number of training increases, the return value of the cost function gradually decreases, that is, the actual output becomes closer to the expectation. In order to effectively avoid the problem of falling into the local extreme, the parameters are updated by random mini-batch.

3 Experiments and Analysis

In this section, we will study the representative capabilities of 3D-CNN features, classification experiments are performed on data sets using traditional SVM methods and our proposed methods. The detailed experimental setup and numerous experiments with reasonable analysis are also presented. Finally, we compare our classification accuracy with the latest remote sensing image classification methods.

3.1 Selection of Dataset

The hyperspectral dataset is the Pavia University dataset acquired by Reflected Optical System Imaging Spectrometer (ROSIS). The image covers 610 × 340 pixels from engineering institute of Pavia University, which was collected by the HySens project of German space agency. The ROSIS-03 sensor contains 115 spectral channels from 430 to 860 nm. In the dataset, with the exception of 12 noise channels, the remaining 103 spectrum channels are processed. The spatial resolution is 1.3 meters per pixel. The training samples available for this dataset cover nine categories of interest. Table 1 shows the information on the different categories and the corresponding training and test samples.

Table 1: The number of training and test samples of the Pavia University dataset

Number	Class	Training	Test
1	asphalt	548	6631
2	meadows	540	18649
3	Gravel	392	2099
4	Trees	524	3064
5	Painted metal sheets	265	1345
6	Bare soil	532	5029
7	Bitumen	375	1330
8	Self-blocking bricks	514	3682
9	Shadows	231	947
Total		3921	42776

3.2 Comparison with SVM Method

The proposed method is compared with the traditional SVM classifier, here the kernel of the RBM classifier is calculated according to the algorithm in the library [24]. Table 2 lists the classification accuracy of the two methods for each category and Fig. 3 shows the classification results of the two methods.

All experiments were performed on a 64-bit Intel i5-4200U machine with a 1.6 GHz clock and 4 G RAM. The graphics processing unit (GPU) used is an NVIDIA GeForce GT 740M with 3.5 GB of memory under the CUDA 8.0 release.

Table 2: Performance Comparison Between SVM and 3D-CNN on Pavia University Dataset

Number	Class	SVM (%)	3D-CNN (%)
1	asphalt	98.26	99.50
2	meadows	99.00	99.78
3	Gravel	78.13	94.71
4	Trees	97.78	99.73
5	Painted metal sheets	99.85	99.98
6	Bare soil	86.63	99.62
7	Bitumen	90.15	97.74
8	Self-blocking bricks	86.63	98.45
9	Shadows	99.89	100.00
Overall Accuracy		95.03	99.30
Average Accuracy		92.92	98.83
Kappa Coefficient		93.41	99.08

It can be seen from Table 2 that the classification accuracy of the SVM method is not very high for some small sample data, but the improved 3D-CNN can obtain higher classification accuracy for small samples, and for most categories the 3D-CNN can get a higher classification accuracy. It can also be seen from Figure 3 that the edge of the classification graph obtained by 3D-CNN is more smooth than that of SVM and the noise point is also smaller. The experimental results show that 3D-CNN provides better explaining ability for the information extraction of spectral hyperspectral image.

To illustrate the efficiency of the proposed method, the time consumption by the three major stages of 3D-CNN on the Pavia University dataset is shown in Table 3, namely image preprocessing, training, and classification.

Table 3: Time Consumption of 3D-CNN on the Pavia University dataset

Number	Stage	Time (h)
1	Image preprocessing	0.2
2	Training	17
3	Classification or test	0.6
Total		17.8

In summary, 3D-CNN can achieve higher and more stable classification accuracy than SVM. Although it takes more time than SVM, when the training set is small and equipped with high-grade experimental equipment, 3D-CNN may also result in rapid convergence speed.

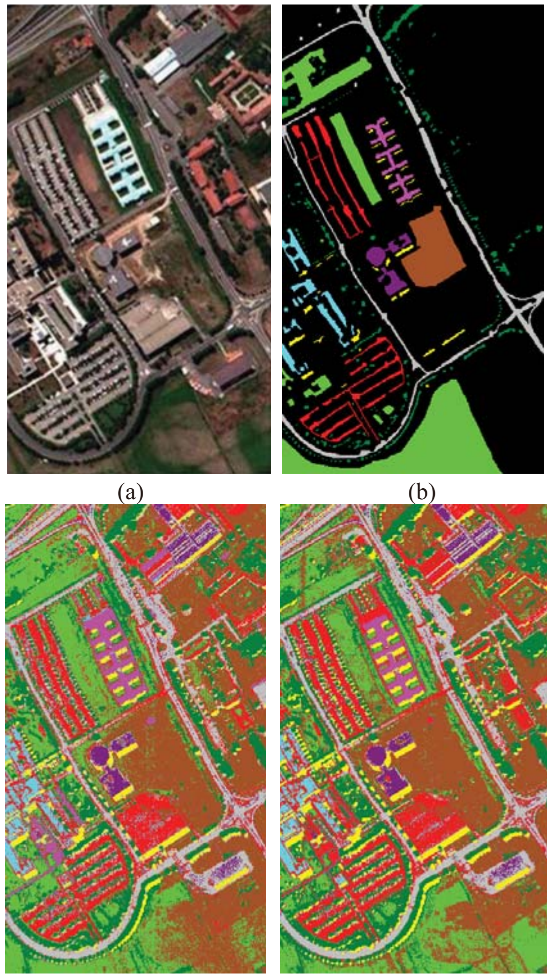
(a) (b) (c) (d)

Fig. 3: Classification results of the Pavia University dataset. (a) Original data; (b) Ground reference; (c) SVM-RBF; (d) 3D-CNN.

3.3 Comparison with State-of-the-Art Methods

To further illustrate the performance of 3D-CNN, it is compared with the most advanced methods. These studies were selected because all of these studies were evaluated on Pavia University datasets. Table 4 shows the overall accuracy. As can be seen from Table 4, the proposed method has higher classification accuracy than others. This once again proves the superiority of the proposed method.

Table 4: Performance Comparison of State-of-the-Art Methods

Number	Methods	Overall Accuracy (%)
1	UFL-SC [5]	90.26
2	SAE-LR [18]	98.12
3	CNN [25]	80.51
4	RNN-GRN-PRetanh [25]	88.85
5	SMLR [13]	93.00
6	K- SMLR [13]	97.41
7	3D-CNN	99.30

4 Conclusion

In order to solve hyperspectral image classification problem, a 3D-CNN-based hyperspectral spectral hyperspace joint classification method is proposed in the paper. The proposed method improves the classification accuracy of small sample by increasing the weight of the network before outputting the loss functions of different classes, which could increase the attention of the network to the small sample data. The experimental results show that the proposed method can obtain higher classification accuracy for different types of samples and achieve better performance than the traditional SVM classifier and some state-of-the-art methods.

Future work will focus on the optimization of the required computational time for the 3D-CNN and further improve the classification accuracy by using an idealized regularized compound kernel.

References

[1] Bioucas-Dias J M, Plaza A, Camps-Valls G, Scheunders P. Hyperspectral remote sensing data analysis and future challenges. IEEE Geoscience and Remote Sensing Magazine,2013, 1(2): 6-36.

[2] Benediktsson J A, Chanussot J, Moon W M. Very high-resolution remote sensing: challenges and opportunities [point of view]. Proceedings of the IEEE, 2012, 100(6):1907-1910.

[3] Landgrebe D. Hyperspectral image data analysis. IEEE Signal Processing Magazine, 2002, 19(1): 17-28.

[4] Nasrabadi N M. Hyperspectral target detection: an overview of current and future challenges. IEEE Signal Processing Magazine, 2014, 31(1): 34-44.

[5] Hu, F.; Xia, G.; Wang, Z.; Huang, X.; Zhang, L.; Sun, H. Unsupervised Feature Learning via Spectral Clustering of Multidimensional Patches for Remotely Sensed Scene Classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2015, 8, 2015–2030.

[6] Kumar S, Ghosh J, Grawford M. Best-bases feature extraction algorithms for classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing,2001, 39(7): 1368-1379.

[7] Samaniego L, Bardossy A, Schulz K. Supervised classi?cation of remotely sensed imagery using a modified k-NN technique. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(7): 2112-2125.

[8] Ediriwickrema J, Khorram S. Hierarchical maximum likelihood classification for improved accuracies. IEEE Transactions on Geoscience and Remote Sensing, 1997, 35(4):810-887.

[9] Foody G M, Mathur A. A relative evaluation of multiclass image classification by support vector machines. IEEE Transactions on Geoscience and Remote Sensing, 2004,42(6): 1335-1343.

[10] Hughes G. On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory, 1968,14(1): 55-63.

[11] Du Peijun, Xia Junshi, Xue Chaohui, Tan Kun, Su Hongjun, Bao Rui. Progress in Classification of Hyperspectral Remote Sensing Images [J]. Journal of Remote Sensing, 2016,20 (02): 236-256.

[12] Yin Ming, Tan Xiong, Zhang Pengqiang, et al.A hyperspectral image information vector machine classification method based on texture and spectral features [J] .Chinese Journal of Surveying and Mapping, 2015,32 (4): 368-372.

[13] Cao F, Yang Z, Ren J, et al. Extreme Sparse Multinomial Logistic Regression: A Fast and Robust Framework for Hyperspectral Image Classification[J]. 2017.

[14] ATKINSON P M, TATNALL A R L. Introduction neural networks in remote sensing [J]. International of Remote Sensing, 1997,18(18):699-709.

[15] BRUZZONE L, PRIETO D F. A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images[J]. IEEE Transaction on Geoscience & Remote Sensing,1999,37(2):1179-1184.

[16] RATLE F, CAMPS G, WESTON J. Semi-supervised neural networks for efficient hyperspectral image classification [J]. IEEE Transaction on Geoscience & Remote Sensing,2010,48(5):2271-2282.

[17] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006,313(5786):504-507.

[18] CHEN Y, LIN Z, ZHAO X, et al. Deep learning-based classification of hyperspectral data [J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2014,7(6):2094-2107.

[19] SUTSKEVER I, HINTON G E. Deep narrow sigmoid belief networks are universal approximators [J]. Neural Computation, 2008,20(20):2629-2636.

[20] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional networks[C]//International Conference on Neural Information Processing Systems. Lake Tahoe, USA, 2012:1097-1105.

[21] Krizhevsky A， Sutskever I， Hinton G E． Image net classification with deep convolutional neural networks:proceedings of advances in neural information processing systems. Cambridge ， MA: MITPress ， 2012:1106-1114.

[22] Ji S, Yang M, Yu K. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 35(1):221-231.

[23] Molchanov P, Gupta S, Kim K, et al. Hand gesture recognition with 3D convolutional neural networks[C]// Computer Vision and Pattern Recognition Workshops. IEEE, 2015:1-7.

[24] Chang C C, Lin C J. LIBSVM: A library for support vector machines[M]. ACM, 2011.

[25] Mou L, Ghamisi P, Zhu X X. Deep Recurrent Neural Networks for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience & Remote Sensing, 2017, 55(7):3639-3655.