A NOVEL DEEP CONVOLUTIONAL NEURAL NETWORK FOR SPECTRAL–SPATIAL CLASSIFICATION OF HYPERSPECTRAL DATA

Na Li [1, *], Chengguo Wang [1], Huijie Zhao [1, *], Xuemei Gong [1], Daming Wang [2]
[1]School of Instrumentation Science and Opto-electronics Engineering, Beihang University, Beijing, China - lina_17@buaa.edu.cn, wcgsdu@163.com, hjzhao@buaa.edu.cn, gxmgxm201303@163.com
[2] China Geological Survey - wangdaming@mail.cgs.gov.cn

Commission III, WG III/4

KEY WORDS: Hyperspectral Data, Classification, Three-dimensional Convolution, Deep CNN, Feature Extraction

ABSTRACT:

Spatial and spectral information are obtained simultaneously by hyperspectral remote sensing. Joint extraction of these information of hyperspectral image is one of most import methods for hyperspectral image classification. In this paper, a novel deep convolutional neural network (CNN) is proposed, which extracts spectral-spatial information of hyperspectral images correctly. The proposed model not only learns sufficient knowledge from the limited number of samples, but also has powerful generalization ability. The proposed framework based on three-dimensional convolution can extract spectral-spatial features of labeled samples effectively. Though CNN has shown its robustness to distortion, it cannot extract features of different scales through the traditional pooling layer that only have one size of pooling window. Hence, spatial pyramid pooling (SPP) is introduced into three-dimensional local convolutional filters for hyperspectral classification. Experimental results with a widely used hyperspectral remote sensing dataset show that the proposed model provides competitive performance.

1. INTRODUCTION

Spatial and spectral information are obtained simultaneously by hyperspectral remote sensing. Hyperspectral classification is one of the foremost tasks in remote sensing image analysis. And its application scope has now reached farther than ever owing to significant improvements of pattern recognition, statistics and other related technologies.

For characteristics of hyperspectral data, some data analysis techniques based on machine learning have been applied to classification during the past several years (Plaza et al. 2009). As an excellent method of machine learning, support vector machine (SVM) was applied to classification of hyperspectral data. It maximizes the margin in high-dimensional feature spaces using kernel methods for the samples. SVM-based classification methods were the state-of-the-art methods for a long time (Melgani and Bruzzone 2004). In addition, many techniques based on spectral–spatial have been proposed for hyperspectral image classification. A model based on markov random fields (MRFs) and SVM, fusing spectral and spatial features, was proposed for hyperspectral image classification (Tarabalka et al. 2010). Then, a spectral–spatial segmentation method based on subspace multinomial logistic and markov random is proposed (Li, Bioucas-Dias, and Plaza 2012). Recently, deep learning has been successfully applied in hyperspectral classification. In order to get deep features generated by the model of hyperspectral classification, a method based on stacked auto-encoders for the classification of hyperspectral data was proposed (Chen, Lin, et al. 2014). But it overlooked the spatial distribution patterns, due to flattening the spatial feature generated by PCA. Then an approach based on convolutional neural network was proposed for extracting spectral features, but it didn’t take the spatial information into account. Instead, convolutional neural network has been introduced for hyperspectral classification to generate spatial features (Yue et al. 2015). For the traditional deep convolutional neural networks can only extract spatial or spectral features of the same scale, a deep convolutional neural network (CNN) with spatial pyramid pooling was proposed to extract spatial features for hyperspectral image classification (Yue, J., et al. 2016).

In this paper, a novel deep convolutional neural network (CNN) is proposed, which extracts spectral-spatial information of hyperspectral images correctly. The proposed model, based on three-dimensional local convolutional filters and SPP, not only learns sufficient knowledge from the limited number of samples, but also has powerful generalization ability.

The following text is organized as follows. In section 2, we present architecture of the proposed method. Experiment results and discussions are given in section 3. In section 4, we make the conclusion.

2. METHODOLOGY

2.1 Three-dimensional convolution for hyperspectral image

In hyperspectral image processing field, 1D CNN or 2D CNN is usually applied to feature extraction (Hu W., et al. 2015). And CNN is used to extract spectral features or spatial features. When applied to HSI classification problems, it is crucial to capture spectral-spatial features in an end-to-end framework. Considering the characteristics of hyperspectral images, three-dimensional hyperspectral data is input into the proposed model. To take advantage of the information in hyperspectral image with more than one hundred of bands efficiently, the proposed model is based on three-dimensional local convolutional filters. As neighboring pixels of hyperspectral image is the input of model, three-dimensional local convolutional filters can learn spectral-spatial features in the same channel easily.

* Corresponding author

The value of a neuron is given and shown as follows:

$$v_{ij}^{xyz} = f\left(b_{ij} + \sum_{m=0}^{M_i-1}\sum_{p=0}^{P_i-1}\sum_{q=0}^{Q_i-1}\sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}\right) \quad (1)$$

where, v means the output variable in the feature map. P, Q is the size of kernel toward the spatial dimension respectively. And (p, q, r) are the indexes of kernel and m is the index of feature map. (x, y, z) are the indexes of feature map. w means the convolutional kernel parameter. i, j are the indexes of input layer and output layer respectively. M is the number of feature maps. b is the bias term.

Parametric rectified linear unit (PReLU) is selected as the activation function of three-dimensional local convolutional filters in this work.

Through 3D convolution, CNN can extract the spatial and spectral information of hyperspectral data simultaneously. The learnt spectral-spatial features are useful for classification.

2.2 Feature extraction with spatial pyramid pooling used 3D pooling

CNN has been introduced for hyperspectral classification to generate spatial features (Yue et al. 2015). When applied to hyperspectral image classification problems, the deep CNNs with traditional pooling can only extract features of the same scale. For this problem, a deep CNN with spatial pyramid pooling (SPP) was proposed for hyperspectral image classification to extract spatial information (Yue et al. 2016).

SPP uses multi-level pooling windows rather than a single window size to pool the feature maps, which is more robust to object distortions and it also can pool features of different scales. Instead, SPP can also generate features of different scales by different sizes of pooling windows in this paper.

To be more robust to object distortions and generate features of different scales, the top pooling layer after the top convolutional layer is replaced by an SPP layer.

Consider the output feature maps of the top convolutional layer which has a size of $n \times m \times w$. Then a certain number of different sizes of pooling windows are chosen as SPP. For example, three sizes of 3D-pooling window are chosen and they are $n \times m \times w$, $\frac{n}{2} \times \frac{m}{2} \times \frac{w}{2}$ and $\frac{n}{3} \times \frac{m}{3} \times \frac{w}{3}$. As the strides of the pooling are the same as the pooling window sizes, SPP can generate features with three sizes ($1 \times 1 \times 1$, $2 \times 2 \times 2$, $3 \times 3 \times 3$).Then these features are flatten and concatenated into one vector which has the size of 1×36. The vector generated by SPP is input of fully connected layer.

Figure 1 illustrates an example of three-level 3D-SPP. The size of the feature map is a × a × a (a is twice as much as b and a is three times as much as c). In each part, the responses of each filter are pooled using max pooling. The outputs of the SPP are different dimensional vectors, respectively.

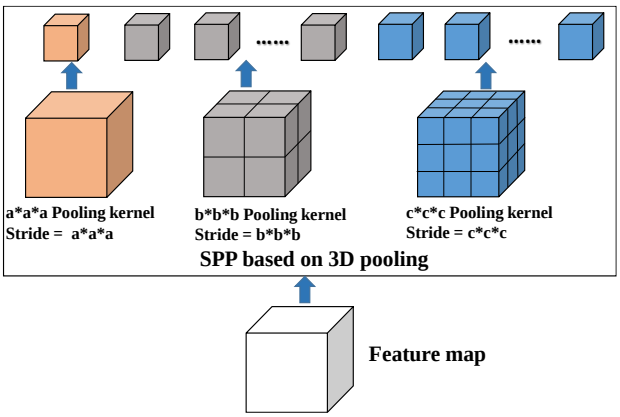


Figure 1. An example of three-level pyramid pooling used 3D pooling. The configuration of the figure is for a network whose feature map size of the top convolutional layer is and the output of the spatial pyramid pooling is {3 × 3 × 3, 2 × 2×2, 1 × 1×1}.

2.3 Joint spectral–spatial classification framework

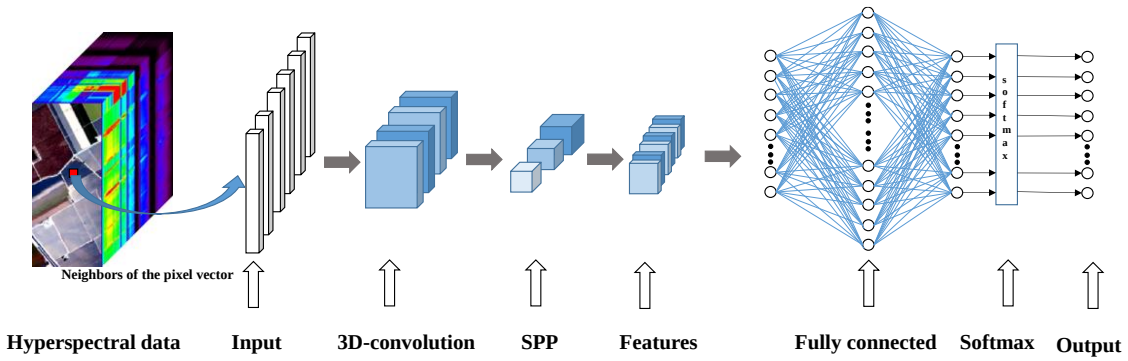


Figure 2. Flowchart of our proposed deep CNN model.

With 3D convolution and SPP based on 3D pooling, a deep convolutional neural network model is built, which is illustrated in Figure 2. Three-dimensional hyperspectral data is input to the proposed model. And the 3D CNN is designed to extract the spectral-spatial features of hyperspectral data, and three different sizes of three-dimensional pooling windows are chosen as the SPP to generate different scales of features. As the input of the fully connected network, these features are concatenated into one-dimensional data. Then the fully connected neural network with activation function called the hyperbolic tangent function fuses the extracted features. At the end of the framework, we choose a logistic regression named softmax to produce the final classification map.

In our framework, spatial pyramid pooling for three-dimensional local convolutional filters can generate features of different

scales to make our proposed model learn these spectral-spatial features easily. Moreover, three groups of features are generated for extracting the spatial-spectral information effectively. Then a sufficient number of fully connected neurons is set to fuse these extracted features. To prevent overfitting, dropout is introduced to fully connected network. The network temporarily stops updating some weights of hidden nodes in network and retains these weights during the training, which can be seen as reducing redundant connections of the network structure randomly.

3. EXPERIMENTS

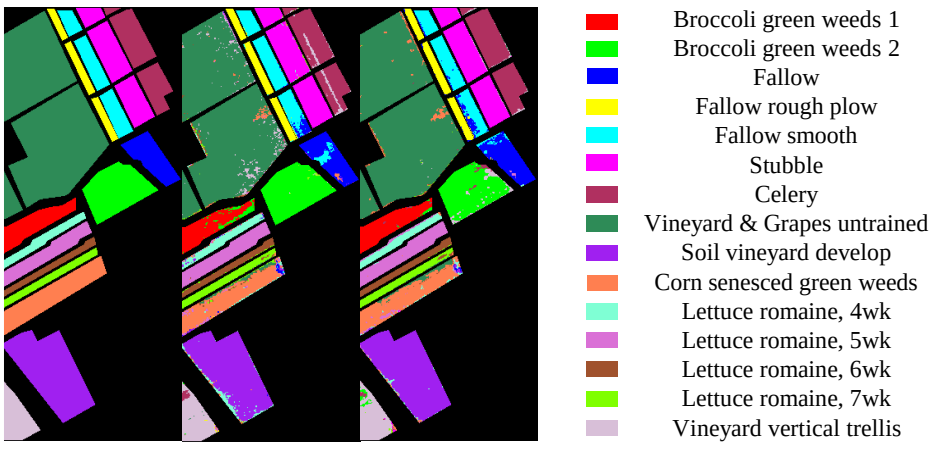
To evaluate the performance of the proposed method, the Salinas Valley scene dataset is used in our work.

The Salinas scene dataset collected by the AVIRIS sensor illustrates an area over Salinas Valley, California, with a spatial resolution of 3.7 m. The image comprises 512×217 pixels with 224 bands. There are also 15 different classes, and the numbers of training and testing samples are listed in Table 1. For the data, 180 labeled pixels per class for training and all other pixels in the ground truth map for test. Surrounding 3×3 neighboring pixels for training convolutional filters are cropped from these two experimental datasets to learn the spatial and spectral features. Because of the existence of 3×3 neighboring pixels, the labeled samples are randomly selected in the middle of category areas. To avoid overfitting, the number of the training samples is also increased by four times by mirroring the training samples across the horizontal, vertical, and diagonal axis.

Figure 3 illustrates the corresponding classification maps obtained with our proposed method and a traditional 2D deep CNN only based on spectral features. Furthermore, overall accuracy and kappa coefficient are calculated by confusion matrices to quantify the performance of the proposed deep CNN. Overall accuracies, individual classification accuracies and kappa coefficient obtained for these two different classification methods are listed in Table 2. Compared with a traditional 2D deep CNN only based on spectral features, the producer accuracies of these classes named fallow, celery and soil vineyard develop are higher obviously. It shows that our proposed model are sensitive to these classes and can get these features easily. The overall accuracy of the traditional deep CNNs is 92.4634%, and the kappa coefficient is 0.9101. Instead, the overall accuracy of our proposed method is 94.2596%, and the kappa coefficient is 0.9312. It is obvious that our proposed method has better performance than a traditional 2D deep CNN using Salinas Valley data set.

Number	Class	Training	Test
1	Broccoli green weeds 1	180	1829
2	Broccoli green weeds 2	180	3546
3	Fallow	180	1796
4	Fallow rough plow	180	1214
5	Fallow smooth	180	2498
6	Stubble	180	3779
7	Celery	180	3399
8	Vineyard & Grapes untrained	180	18359
9	Soil vineyard develop	180	6023
10	Corn senesced green weeds	180	3098
11	Lettuce romaine, 4wk	180	888
12	Lettuce romaine, 5wk	180	1747
13	Lettuce romaine, 6wk	180	736
14	Lettuce romaine, 7wk	180	890
15	Vineyard vertical trellis	180	1627
	Total	2700	51429

Table 1. Number of training and test samples used in the Salinas scene dataset.



(a) (b) (c)

Figure 3. RGB composition maps resulting from classification for the Salinas scene dataset. From left to right: (a) ground truth, (b) a traditional deep CNNs, and (c) the proposed method.

Class	2D deep CNN	Our proposed method
Broccoli green weeds 1	83.43%	**95.52%**
Broccoli green weeds 2	**98.90%**	89.51%
Fallow	86.25%	**90.81%**
Fallow rough plow	**99.02%**	97.79%
Fallow smooth	**88.03%**	86.02%
Stubble	98.28%	**98.28%**
Celery	90.11%	**96.73%**
Grapes & Vineyard untrained	94.25%	**96.97%**
Soil vineyard develop	92.53%	**97.29%**
Corn senesced green weeds	75.98%	**79.76%**
Lettuce romaine, 4wk	**85.46%**	80.65%
Lettuce romaine, 5wk	**100.00%**	99.54%
Lettuce romaine, 6wk	94.16%	**97.96%**
Lettuce romaine, 7wk	**94.84%**	90.46%
Vineyard vertical trellis	**93.39%**	90.45%
Overall Accuracy	92.4634%	**94.2596%**
Kappa coefficient	0.9101	**0.9312**

Table 2. Overall accuracies, individual classification accuracies and kappa coefficient obtained for different classification methods when applied to the AVIRIS Salinas scene hyperspectral data set.

4. CONCLUSION

In this paper, we proposed a novel deep CNN for spectral-spatial classification of hyperspectral data. To be more robust to object distortions and generate features of different scales, spatial pyramid pooling is introduced into three-dimensional local convolutional filters for hyperspectral classification. When applied to the AVIRIS Salinas scene hyperspectral data set, overall accuracies and kappa coefficient is obtained for different classification methods. The overall accuracy of the traditional deep CNN is 92.4634%, and the kappa coefficient is 0.9101. Instead, the overall accuracy of our proposed method is 94.2596%, and the kappa coefficient is 0.9312. Compared with a traditional 2D deep CNN only based on spectral features, the proposed method could achieve higher accuracy using the Salinas scene dataset. Research on our proposed model for other widely used hyperspectral remote sensing datasets is our future work.

ACKNOWLEDGMENTS

This work was supported by the National Key Technologies R&D Program (Grant No. 2017YFC0602104 and No. 2016YFB0500505), National High Technology Research and Development Program (Grant No. 2016YFF0103604), National Natural Science Foundation of China (Grant No. 41402293), and Program for Changjiang Scholars and Innovative Research Team (Grant No. IRT0705).

REFERENCES

Chen, Y., Lin, Z., Zhao, X., Wang, G., and Gu, Y., 2014. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 7(6), pp. 2094-2107.

Hu, W., Huang, Y., Wei, L., Zhang, F., and Li, H., 2015. Deep convolutional neural networks for, hyperspectral image classification. *Journal of Sensors*, 2015(2), pp. 1-12.

Li, J., Bioucas-Dias, J. M., and Plaza, A., 2012. Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields. *IEEE Transactions on Geoscience & Remote Sensing*, 50(3), pp. 809-823.

Melgani, F., and Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience & Remote Sensing*, 42(8), pp. 1778-1790.

Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., and Camps-Valls, G., et al., 2009. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113(1), pp. S110-S122.

Tarabalka, Y., Fauvel, M., Chanussot, J., and Benediktsson, J. A., 2010. SVM - and MRF-based method for accurate classification of hyperspectral images. *IEEE Geoscience & Remote Sensing Letters*, 7(4), pp. 736-740.

Yue, J., Mao, S., and Li, M. 2016. A deep learning framework for hyperspectral image classification using spatial pyramid pooling. *Remote Sensing Letters*, 7(9), pp. 875-884.

Yue, J., Zhao, W., Mao, S., and Liu, H., 2015. Spectral–spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sensing Letters*, 6(6), pp. 468-477.