Hyperspectral image classification based on multi-layer feature extraction

Yongfeng Qi¹ and Xujie Yang^{1,a}

^{1,a}College of Computer Science & Engineering, Northwest Normal University, 730070 Lanzhou, China

Abstract. The Hyperspectral image classification is an important issue, which has been pursued in recent year. The field of application involves many aspects of life. Hyperspectral images (HSIs) exhibit a limited number of labeled high-dimensional training samples, which limits the performance of some classification methods on feature extraction or feature reduction. In the paper, we propose a supervised method based on the PCA network (PCANet) and linear SVM for HSIs classification. We used PCANet (principal component analysis network) to learn the character features. We verified the influence of these parameters on the performance of PCANet by modifying the key parameters of the experiment. We carry out extensive experiments on India pines dataset. The results demonstrate that our method significantly outperforms PCA+KNN methods. And the results not only are optimistic but also the recognition rate can reach 94.29%. At last, we compared the experimental results of the same algorithm on different data sets and so on.

1 Introduction

With the rapid development of remote sensors, the acquisition and collection of hyperspectral data has become much easier and more affordable, making hyperspectral image analysis to be one of the most promising techniques in many practical applications, including precision agriculture, environmental monitoring, military surveillance, etc^[1]. PCANet is proposed in [2], mainly using Face Recognition and Handwritten digit recognition, and has made great success in printed document recognition, and it is a kind of deep learning framework by cascading PCA as in convolution neural network. Consequently, a great number of classification methods have been proposed in character recognition^[3], especially in scene character recognition^[4], but Hyperspectral images classification using PCANet is only a little in recent 2 years. HSIs classification focuses on extracting effective features, combining with complicated spectral, spatial information, classifier and so on. Many spectral feature-based methods for HSIs classification have been developed gradually and applied successfully. For example, Deep learning based character feature learning methods have drawn increasing interests from the community, but also some methods are applied based on the support vector machine (SVM)^[5,6]. Some are designed based on multinomial logistic regression^[7], using limited samples^[8] and a supervised method based on PCANet and Gaussian SVM^[9]. Shiqi Yu et al. [1] proposed a feature learning method using convolutional neural networks (CNN) for hyperspectral image classification, which draws significant attention recently. Although these classification methods can

make full advantages of the spectral information of HSIs, some noise points often found to be appeared on obtained classification maps, and classification is inconvenient. Therefore, the spatial information or the advanced feature should be taken into account. Recently, numerous works have been made to combine the spatial information with spectral information^[10].

However, data high-dimensionality and lack of labeled samples limit the performance of these methods on feature extraction or feature reduction for classification. It is critical yet challenging to fully excavate and use the information for HSIs classification. One characteristic of the PCANet model seem to challenge common wisdoms in building a deep learning network such as ConvNet^[11,12]. PCANet extracts image features during the training procedure is divided into four steps:

The first step is pattern preprocessing. It mainly adds zero to the each pattern, so that the dimension of the pattern after zero is integer. Only in this way, can we solve the first more conveniently.

The second step is cascade the principal component analysis as in convolution neural network. It is mainly to extract the feature of patterns and it contains two stage in this step. One is the original input image of the local characteristics of the sliding window is selected to complete the image feature selection operation, and then calculation the result, next the feature vector will be selected as the main filter of the stage, finally each image will have the filters for a convolution operation. The second stage is similar with the first stage and the input images of second stage come from the the first stage.

^a Corresponding author: xujieyang90@gmail.com

The third step is to perform a hash operation. The purpose is to keep the binarization processing after encoding translate into the decimal image.

The fourth step is to calculate block histogram and to feature extraction. The following will explain feature extraction in detail.

2 The proposed method

PCANet is a deep learning baseline for image classification that performs competitively when extracting features. In this paper, we proposed a new method that Hyperspectral image classification based on multi-layer feature extraction.

2.1 PCANet for HSIs

We used a two-level cascaded PCA of PCANet to learn the spectral features. The architecture of PCANet for HSIs is shown in Fig.1 and Fig.2^[2]. In the Fig.1 we first deal with the preprocessed hyperspectral image pixel vector two times, because the first time we deal with the noise, select the useful band, transform the high optical image data cube from the three-dimensional space to the two-dimensional space. After second processing, in fig.1, The red strip represents the input hyperspectral image pattern. By step 1, we can get a red and black strip. The black represents the number of zeros that we make up behind the red strip. Then step 2, we can turn the sample before step 1 into a square, and finally, step 3, get it. The input sample we need is I_i . The method we propose is zeros, truncation and square up.

In Fig.2 we use the pattern I_i , it represents the input Hyperspectral image pattern data and W is the filter parameter. In the input layer, image I_i is divided into patches utilizing the overlapping strategy. L_1 and L_2 are the number of selected principal components in the two-level cascaded PCA separately. In the first stage of cascaded PCA, the front L_1 blocks of PCA from the patches are chosen to convolve with the original image, I_i . This produces L_1 images, which act as the input data of the second stage. In the second stage of cascaded PCA, the structures are designed similarly as the first stage. The front L_2 blocks are also used to convolve with the input data in second stage. Thus, the obtained $L_1 \times L_2$ images represent the input for the last layer. While the histogram blocks are extracted from the output images of the second stage of cascaded PCA. Finally, the features f_i of I_i are learned. A detail description on the PCANet is given in ^[2] and ^[9].

Pixel Vector(X_i)

Fig.1. The architecture of PCANet for HSIs

Fig.2. The architecture of PCANet for HSIs

2.2 HIS Classification

The proposed method of Hyperspectral image classification based on multi-layer feature extraction contains four main steps: Preprocessing, Sample selection, Spectral features extraction, Linear SVM classification and so on. The diagram of the proposed PCANet- Linear SVM to classify HSIs is shown in Fig.3.

Fig.3. Hyperspectral image classification flow chart

3 Main steps

Suppose that a pattern has p characteristics, p is a pixel vector, and we added q number of zeros at each the end of pattern, so that the pattern after zero is integer $p + q$, and the specific steps are as follows formula:

$$\text{def: } p, q, n \in \mathbb{Z}^+; \\ n = \sqrt{p + q}, \quad \text{s.t. } p, q, n \in \mathbb{Z}^+ \quad (1)$$

Only in this way, can we make pixel vector p became a square, its border length and width is n .

3.1. Feature extraction

3.1.1 The first stage of PCA

Suppose that N input training sample images of size $n \times n$ are given, and the patch size is $k_1 \times k_2$. First of all, the local features are selected for each training image by using the patch, then subtract patch mean from each and obtain^[3]:

$$\bar{X}_i = [\bar{x}_{i,1}, \bar{x}_{i,2}, \dots, \bar{x}_{i,\tilde{m}\tilde{n}}] \quad (2)$$

Where, $\bar{x}_{i,j}$ is a mean-removed patch, Constructing the same matrix for all input images and putting them together to get:

$$X = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N] \in \mathbb{R}^{k_1, k_2 \times \tilde{m}\tilde{n}} \quad (3)$$

If the number of filter in the i layer is L_i , PCA reduce the reconstruction error to meet the requirements of the following formula:

$$\min_{V \in \mathbb{R}^{k_1, k_2}} \|X - VV^T X\|_F^2, \text{ s. t. } V^T V = I_{L_1} \quad (4)$$

Where I_{L_1} is identify matrix of size $L_1 * L_1$, The solution is known as the L_1 principal eigenvectors of XX^T . The PCA filters are therefore expressed as:

$$W_l^1 = \text{mat}_{k_1, k_2}(q_l(XX^T)) \in \mathbb{R}^{k_1, k_2}, \quad l = 1, 2, \dots, L_1 \quad (5)$$

3.1.2 The second stage of PCA

The second stage of PCA process is similar with the first, Just it is noting that the second stage input is output of the first. Assume that l th filter output of first stage be^[3]:

$$I_i^l = I_i * W_l^1, \quad i = 1, 2, \dots, N, \quad (6)$$

Then as the same as the first stage, we select the local feature for each image by using the patch, then subtract patch mean from each and obtain:

$$\bar{Y} = [\bar{y}_{i,l,1}, \bar{y}_{i,l,2}, \dots, \bar{y}_{i,l,\tilde{m}\tilde{n}}] \in \mathbb{R}^{k_1, k_2 \times \tilde{m}\tilde{n}}, \quad (7)$$

next collect all the average value of the matrix after the first l as output filter:

$$Y^l = [\bar{Y}_1^l, \bar{Y}_2^l, \dots, \bar{Y}_N^l] \in \mathbb{R}^{k_1, k_2 \times \tilde{m}\tilde{n}}, \quad (8)$$

To together all the filters outputs as

$$Y = [Y^1, Y^2, \dots, Y^{L_1}] \in \mathbb{R}^{k_1, k_2 \times \tilde{m}\tilde{n}}, \quad (9)$$

and select the first L_2 main feature vector as the filter. Finally, each image will have the filters for a convolution operation respectively.

3.1.3 Hash operation

The matrices handled by PCA are done binarization processing, and the result is a matrix that consist zero and one. Then we view the vector of binary bit as a decimal number, so that an output matrix can convert into a single integer image. The calculation formula is as follows^[3]:

$$T_i^l = \sum_{l=1}^{L_2} 2^{l-1} H(I_i^l * W_l^2) \quad (10)$$

3.1.4 Histogram operation

The core of PCANet is to use PCA to learn multi stage filter, so the effectiveness of the filter selection is essential to the generation of the characteristics of the expression. Then divide each image into B blocks and the size of the block matrix is obtained. Next calculate the histogram of each block, and then transformed the histogram into a vector. Finally we collect all the vectors

together to get the characteristic vector of each training image.

4 Experiments

4.1 Effects of number of patch

The experimental test data set for the data Indian Pines, including the test data 7390 and the training data of 1844. Each image size is 15×15 pixels. In the data sampling, using our sampling method. A large number of experiments are conducted to study the effect of different parameters on the performance of the experiment. One of the first to verify the impact of the size of the slider on the experiment, the recognition rate of the relevant parameters of the following table 1. According to table data, in the case of other parameters unchanged, the large size of the recognition rate is relatively high. But when the block size is increased to nine, the recognition performance is slightly reduced. Because the image local feature extraction by sliding window, sliding window is too small, incomplete extraction of local feature, PCA reconstruction could appear some errors, so recognition rate could down.

Table1. Experimental data

Num Stages	2	2	2	2
Patch Size	3	5	7	9
Num filter	[8 8]	[8 8]	[8 8]	[8 8]
HistBlock	[7 7]	[7 7]	[7 7]	[7 7]
Overlap rate	0.5	0.5	0.5	0.5
Recognition rate	75.64%	86.29%	86.56%	85.99%

The following polyline graph is shown in Fig. 4, the network layers are 2, the number of filters is 8, the HistBlock is 7, and the block Overlap rate is 0.5. However, the recognition rate of hyperspectral remote sensing images with different block sizes is as follows:

Fig.4. Recognition rate graph

4.2 Effects of the number of filters

In the training process of PCA, each image at first stage perform convolution operation with each filters and result is the input image of the second stage to train. Then each image perform convolution operation with each filter at the second stage respectively. The experimental results are shown in the following table. From the table we can know that the number of the filter at two stages are eight and sixteen, the recognition rate is

the best, at the same time, the training time is the short. Because filter is generated by the the feature vector, and each image needs to perform convolution operation with those filters, so the number of filters will certainly have an impact on the experiment. The more the number of filters, the greater the amount of data, the amount of computation is complex.

Table 2.Experimental data

Num Stages	2	2	2	2
Patch Sizes	7	7	7	7
Num Filter	[8 8]	[16 8]	[8 16]	[16 16]
Hist Block	[7 7]	[7 7]	[7 7]	[7 7]
Overlap Rate	0.5	0.5	0.5	0.5
Recognition Rate	86.56%	87.90%	88.59%	88.53%
PCA Train Times(s)	218.74	238.85	284.03	442.51

4.3 Effects of the number of PCA train layer

We know that PCANet has two layer of the PCA processing stage, and images feature extraction are done in the two stages. In order to verify the effect of the number of PCA layers on the experiment, we do the experiment on the structure of two layer PCA processing, three layer PCA processing and four layer PCA processing respectively. The results of the experiment are shown in the following TABLE 3. We can know that the recognition is better than others when the number of PCA layer is three. It also shows that the number of PCA training layer does not mean the higher the recognition rate.

Table 3.Experimental data

Num Stages	2	3	4
Patch Size	7	7	7
Num Filter	[8 8]	[8 8 8]	[8 8 8 8]
Hist Block	[7 7]	[7 7]	[7 7]
Overlap Rate	0.5	0.5	0.5
Recognition Rate	86.56%	89.68%	89.99%

The following broken line graph is shown in Fig. 5. The network layers are different, the sampling blocks are all 7, the number of filters is 8, the histblock is 7, and the block overlap rate is 0.5. However, the recognition rate of hyperspectral remote sensing image is different in the case of different network layers, as shown in Fig. 5 below.

Fig.5.Recognition rate graph

4.4 Classification of hyperspectral images by PCANet

In order to verify the classification of Hyperspectral images by PCANet, we find a series of Hyperspectral images which are made up of 3 (In there, At least three species of ground, otherwise there is no need to classify.) to t ($t \in \mathbb{Z}^+$ & $t > 2$) of these kinds of ground objects from different places, each of ground object has different curves, a total of many many samples. 20% sample pictures for training and 80% samples for testing.

Owing to these samples are influenced by many factors, such as illumination. In order to verify the PCANet can adapt to the Hyperspectral remote sensing image samples, we compare the experimental results with the original experimental results. The results of the experiment are shown in the TABLE IV. We can know that the recognition rate of Pavia University is 94.13%, which just high 7.57% than Indian Pines data. From the experimental data shows that the recognition effect is good, at the same time, it also shows that PCANet has a good adaptability to different Hyperspectral remote sensing image data.

Table4.Experimental

	Indian Pines	Pavia University
Num Stages	2	2
Patch Sizes	5	5
Num Filters	[8 8]	[8 8]
Hist Block	[7 7]	[7 7]
Overlap Rate	0.5	0.5
Recognition Rate	86.29%	89.48%

The content of the above table is the average classification accuracy of the test samples. At the same time, in order to illustrate the performance of PCANet, we use the same algorithm to test the performance of the algorithm with different data sets. Many experiments show that the overall classification effect on Pavia University dataset is better than that on Indian Pines dataset. Experiments on Pavia University dataset show that the highest recognition accuracy is 92.83% when the network layer number is 2, the block size is 7, the number of filters is 8, and the overlap rate is 0.5. Similarly, the highest recognition accuracy is 94.29% when the network layer number is 2, the block size is 5, the number of filters is 8, and the overlap rate is 0.5. Then, we use different algorithms for the same dataset, and we illustrate its superiority through

comparative experiments. The recognition rate of PCA+KNN can reach 71%, because PCA + KNN algorithm firstly reduces the dimension of hyperspectral remote sensing data by PCA algorithm, and then classifies hyperspectral image patterns by KNN algorithm. The recognition rate of PCANet is about 90%, and the recognition rate of PCANet and PCA + KNN is obviously different, which fully demonstrates the reliability of PCANet in hyperspectral image classification.

5 Conclusions

In this paper, a new Hyperspectral image classification method is proposed. We have compared the experimental data, verified the several factors affecting the identification of PCANet and the adaptability of PCANet, at the same time, we also showed that the reliability of the Hyperspectral image recognition by the PCANet. Then we would consider using big Hyperspectral image data sets to train. The intention is that I hope to achieve better classification accuracy and performance.

On the other hand, further research can be extracted and classified from two dimensions of spectral feature and space spectrum feature. With the condition was allowed, the high spectral image data set could be used so that the features of hyperspectral image data can be extracted better so as to get better classification results.

Acknowledgment

This work was supported by Gansu Provincial Higher Education Research Project, China (No. 2016A-004), Gansu Science and Technology Plan Project, China (No. 18JR3RA097).

References

1. S. Yu, C. Xu, and S. Jia, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, 219 :88–98, (2017)
2. T. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, Y. Ma, "PCANet: A Simple Deep Learning Baseline for Image Classification?", *IEEE Transactions on Image Processing*, vol.24, pp:5017-5032, (2015)
3. R. Liu, and T. Lu, "Character Recognition Based on PCANet," 2016 15th International Symposium on Parallel and Distributed Computing (ISPD), pp:364-367, (2016)
4. C. Chen, D. Wang, and H. Wang, "Scene Character Recognition Using PCANet," *ICIMCS 2015*:65:1-65:4, (2015)
5. F. Melgani, and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, (2004)
6. M. Chi, and L. Bruzzone, "Semisupervised classification of hyperspectral images by SVMs optimized in the primal," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no.6, pp. 1870-1880, (2007)
7. J. Li, J.M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression," *IEEE Geosci. Remote Sens. Lett.*, vol.10, no.2, pp.318–322, (2013)
8. B. Pan, Z. Shi, and X. Xu, "MugNet: Deep learning for hyperspectral image classification using limited samples," *ISPRS*, vol.145, Part A, pp.108-119, (2018)
9. F. Wang, R. Zhang, and Q. Wu, "Hyperspectral image classification based on PCA network," *IEEE 2016 8th WHISPERS*, pp:1–4, (2016)
10. M. Fauvel, Y. Tarabalka, J.A. Benediktsson, J. Chanussot, and J.C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol.101, no.3, pp.652–675, (2013)
11. K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition," in *ICCV*, (2009)
12. A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural network," in *NIPS*, (2012)