Hyperspectral Image Recognition Using SVM Combined Deep Learning

Yifan Li[1], Junbao Li[2], Jeng-Shyang Pan[1,3,4]

[1] College of Computer Science and Technology, Shenzhen Graduate School, Harbin Institute of Technology, China
[2] Department of Electronics and Information Engineering, Harbin Institute of Technology, China
[3] College of Computer Science and Engineering, Shandong University of Technology, China
[4] School of Information Science and Engineering, Fujian University of Technology, China
lyf3f@163.com, junbaolihit@126.com, jspan@cc.kuas.edu.tw

Abstract

In this paper we present the conbination of deep learning and Support Vector Machine applied on the recognition of hyperspectal images. Hyperspectral image recognition is an essential problem in the practical hyperspectral imagery system. While deep learning is capable of reproducing feature vectors with great dimensions out of original data, it leads to great time cost and the Hugh phoenomenon. Such nonlinear problem is regarded as obstacles and kernel method appears to be a promising way to solve it. The performance of kernel-based learning system is influenced by the choices of kernel function and parameter greatly. We present the kernel learning method termed Support Vector Machine (SVM) applied on feature vectors supplied by deep learning upon hyperspectral image. The learning system is improved by adjusting the parameters and kernel functions to the data structure for improving performance on solving complex tasks. Experimental results validate the feasibility of the proposed methods.

Keywords: Deep learning, SVM, Hyperspectral image

1 Introduction

In recent decades, hyperspectral imageris have exibited great potential capability in remote sensing including target detection technology and spectral imaging technology. Among all hyperspectral applications, recognition various land coverings have been one of our essential concerns, which benefits from the great amount of spatial and spectral information, compared to other images. However, such amount of high spectral dimension leads to high dimension feature vectors which create onstacles in applying traditional image classification algorithms [1]. The Hughes phenomenon may come into being during classification procedures if training samples are outnumbered which happens a lot [2].

The main procedure of the hyperspectral image recognition can be divided into two parts: extracting essential features from enormous bands and designing suitable classifiers for significant classification accuracies. Unfortunately, the huge data size of the hyperspectral image is not only bad for detecting valuable information but also increases the difficulties of classifiers construction. These problems, coupled with other disadvantages such as the overfitting of classifiers caused by the noise, will seriously influence the classification performance [2].

In general, there are three categories of HSI spectral-spatial classifiers. First, many spectral-spatial classifications extract spatial and spectral features from HSI before performing classification. Spatial features based on morphological filters [3-6] are widely used in HSI classification; for example, Ghamisi et al. exploit spatial information using extended multi-attribute profiles (EMAPs) [7]. In order to use spatial features, some researches extract spatial features and spectral features, then use spatial and spectral information in a concatenation strategy; However, all those spatial features are handcrafts, which demanded human knowledge. Furthermore, more features mean higher dimensionality and make HSI classification a more time-consuming task [8].

Second, some spectral-spatial classifications take spatial information into the classifier during classification. Simultaneous orthogonal matching pursuit (SOMP) and simultaneous subspace pursuit (SSP) [9] incorporate the spatial correlation between neighboring samples through a classifier based on joint sparsity representation. This kind of methods gives neighboring samples the right of decision and can improve classification [8].

Third, several classification methods attempt to use spatial dependencies after classification in a decision rule or by spatial regularization. Tarabalka [10] proposed a spectral-spatial classification scheme based on the pixel-wise SVM classification, followed by

*Corresponding Author: Jeng-Shyang Pan; E-mail: jspan@cc.kuas.edu.tw
DOI: 10.3966/160792642019052003018

majority voting within the watershed regions. All those spectral-spatial classifications significantly improve classification results and can be used in succession [8].

Deep learning using neural networks have presented state-of-the-art performances in a wide range of tasks. A great many of the tasks use the softmax activation function for classification. Support vector machine is an widely used alternative to softmax for classification. Using SVMs in combination with convolutional nets have been proposed in the past as part of a multistage process. In particular, a deep convolutional net is first trained using supervised/unsupervised objectives to learn good invariant hidden latent representations. The corresponding hidden variables of data samples are then treated as input and fed into SVMs.

In this paper, we provide a solution to object classification on hyperspectral image using SVM combined deep learning, the following content is divided into 5 parts: related works; proposed method; experiment and result; and conclusion.

2 Related Works

Recently, fully-connected and convolutional neural networks have been trained to achieve state-of-the-art performance on a wide variety of tasks such as speech recognition, image classification, natural language processing, and bioinformatics. For classification tasks, most of these "deep learning" models employ the softmax activation function for prediction and minimize cross-entropy loss [11]. Support vector machine is an widely used alternative to softmax for classification [12]. Using SVMs (especially linear) in combination with convolutional nets have been proposed in the past as part of a multistage process. In particular, a deep convolutional net is first trained using supervised/unsupervised objectives to learn good invariant hidden latent representations [11]. The corresponding hidden variables of data samples are then treated as input and fed into linear (or kernel) SVMs [13]. Other papers have also proposed similar models but with joint training of weights at lower layers [14]

SVM is a kernel based classifier consisting in projecting data in a high dimensionality space by means of non-linear mapping function Φ and aiming at determining the optimal separator hyperplane by margin maximization. SVM is a supervised classifier. It has been proposed first for binary classification [15]. In Yichuan Tang's paper, they demonstrate a small but consistent advantage of replacing the soft-max layer with a linear support vector machine [11]. In Rafika Ben Salem's work, they proposed method exploited the performance of support vector machine (SVM) in the processing of data with high dimensionality. In their work, they employed all spectral information and two different spatial features which are Extended Multi-Attribute Profile (EMAP) and the mean of neighborhood pixels. Camps-Valls and al. [16] have investigated the combination of spectral and spatial kernels to get an accurate classification of hyperspectral images. Fauvel and al. [17] have registered a significant improvement according to the characterization of each pixel by a stocked vector that concatenate spectral and contextual information extracted by Morphological Profiles (MP). Kang and al. [18] have exploited the performance of edge-preserving filters to develop an accurate spectral-spatial classification outperforming the classification without filtering. Huang and al. [19] have proposed a multi-feature model aiming at constructing a SVM set combining multiple spectral and spatial features.

The spectrum data in database is collected in advance, so it has inconsistency between the spectrums with the data collection. The inconsistency can be considered the nonlinear changing. The relationships of between spectral curves are the classical nonlinear relationship. So the classification is the nonlinear and complex classification problem. Traditional classification methods are not effective to hyperspectral sensing data, among these machine learning methods kernel learning is a feasible and effective nonlinear classifier methods on hyperspectral sensing data. Kernel-based machine learning, which solves the problem of linear learning using kernel trick, is one important preprocessing method of hyperspectral sensing data. Recently, many linear methods are kernelized. For example, Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) [20-22] are kernelized to Kernel PCA (KPCA) and Kernel Discriminant Analysis (KDA) [23], respectively. The performances of these linear methods are improved because the kernel method is used to characterize the complicated nonlinear relationships. Researchers are developing the kernel learning methods. Baudat and Anouar [24], Liang and Shi [25], Lu [26], Chen [27] and Wang [28] developed a series of improved KDA methods. Some other researchers presented the alternative framework of KLPP to develop a framework of KPCA+LPP [29-31] for image recognition, radar target recognition and other researchers improved LPP with kernels [32-36]. Researchers optimized the parameters of kernel function to improve kernel-based learning [27, 33, 37]. These methods select the optimal kernel parameter from a set of discrete values, but the geometry structure of data distribution in the kernel-based mapping space is not changed. Xiong proposed a data-dependent kernel machine learning [34] and Amari presented the support vector machine classifier by modifying the kernel function [20]. In the previous works [27, 34], the authors presented data-dependent kernel for face recognition. Multiple kernel learning methods are developed to solve the kernel model selection problems [38-41]. In the work of Lingping Kong [42-43], their study in the Hierarchichal network and Genetic Algorithm have been inspiring for the design of our own neural network and our experiment.

3 Proposed Method

3.1 Motivation and Contribution

Deep learning based methods are widely used in machine learning. Deep learning learns hierarchical representation, and the higher layer represents increasingly abstract concepts and is increasingly invariant to transformations and scales [8].

For classification problems using deep learning techniques, it is standard to use the softmax or 1-of-K encoding at the top. For example, given 10 possible classes, the softmax layer has 10 nodes denoted by p_i, where i = 1, ..., 10. p_i specifies a discrete probability distribution, therefore, $\sum_{i}^{10} p_i = 1$.

Let h be the activation of the penultimate layer nodes, W is the weight connecting the penultimate layer to the softmax layer, the total input into a softmax layer, given by α, is

$$\alpha_i = \Sigma_k h_k W_k \tag{1}$$

then we have

$$p_i = \frac{\exp(\alpha_i)}{\sum_{j}^{10} \exp(\alpha_j)} \tag{2}$$

The predicted class i would be

$$\hat{i} = \underset{i}{\arg\max}\ p_i = \underset{i}{\arg\max}\ \alpha_i \tag{3}$$

To achieve a better result, we fused the SVM method as an advanced classifier. SVM Methods are supervised learning models with associated learning algorithms that analyse data and recognize patterns and are used for classification and regression analysis. The basic SVM takes a set of input features and predicts, for each given input, the possible class form, making it a nonprobabilistic binary linear classifier.

Given a training dataset $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ where $x_1 \in R$ and y_1 is either 1 or −1 indicating the class to which the point x_1 belongs, let $x = [x_1, x_2, ..., x_n]^T$. The construction of the hyperplane for a linearly separable problem is $w^T x + b = 0$, where w is the normal vector to the hyperplane and the parameter $b / \| w \|$ determines the offset of the hyperplane from the origin along the normal vector w. Thus, the margin between the hyperplane and the nearest point is maximized and can be posed as the following problem:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i \tag{4}$$

Subject to $y_i (w^T x_i + b) \geq 1 - \xi_i$, i = 1, 2, ..., n, $\xi_i \geq 0$

where C is a user-defined constant as the penalty parameter of the error term.

In machine learning, one-class classification, also known as unary classification, tries to identify objects of a specific class amongst all objects, by learning from a training set containing only the objects of that class. This is different from and more difficult than the traditional classification problem, which tries to distinguish between two or more classes with the training set containing objects from all the classes. One-class SVM is an unsupervised algorithm that learns a decision function for novelty detection: classifying new data as similar or different to the training set.

Most deep learning methods for classification using fully connected layers and convolutional layers have used softmax layer objective to learn the lower level parameters. In this paper, we substitute softmax with single-class SVM as classifier. Single-class SVM takes the output of the third layer of the backpropogate procedure of deep learning as input which are vectors with 2000 dimensions. It is proved that SVM combined deep learning (DLSVM) is capable of recognize materials better than using deep learnning only.

3.2 Steps and Procedure

Based on our work flow design in Figure 1, our first procedure is to delete some noise band in order to improve the accuracy. Secondly, in virtue of the lack of pixels of some materials, we eliminated two kinds of materials from our experiment.

Finishing the above steps, we have obtained sufficient high-quality data to begin recognition. Firstly, we utilize deep learning transforming original data vectors into vectors of 2000 dimentions; then we divide the vectors of each material into training samples and test samples according to a certain ratio; finally, the vectors are imported into single-class SVM for recognition and we can achieve its accuracy.

3.3 Discussion

Through such procedure, we managed to take advantage of the neural network to obtain ten times the amount of feature value than original ones while combining it with SVM for its ability of classification. The previous process dig into the hidden pattern of different materials making it easier to conduct classification at the risk of Hughes danger which can be avoided by applying SVM that makes high dimensional vector classification possible.

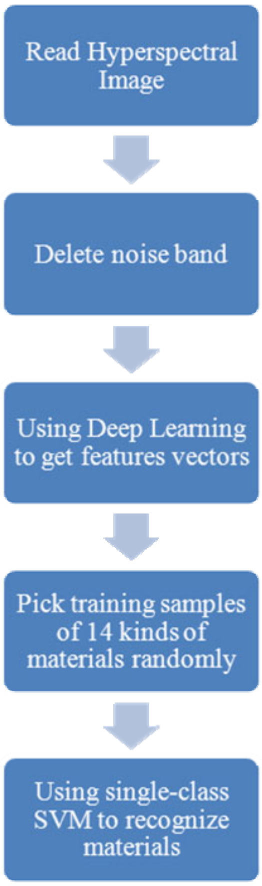


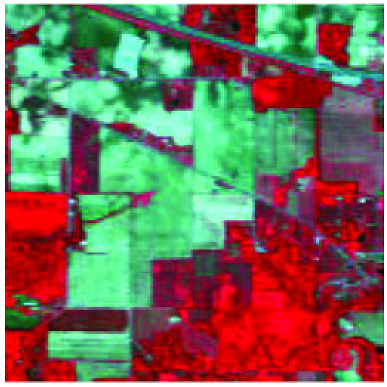
Figure 1. Work flow of SVM combined deep learning on hyperspectral image

4 Experiment and Result

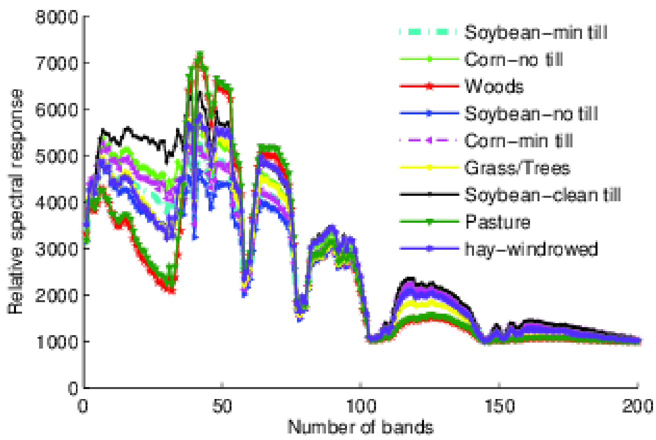
4.1 Experiment Sighting and Database

Indian Pines dataset is collected under various spectral and spatial resolutions. The spetral curves denote the different remote sensing environments with Airborne platform. The data cube have 224 bands of spectral resolution through 0.4-2.5μ m range and it has the spatial resolution of 20m per pixel. We removed the noisy and water-vapor absorption bands and 200 bands of images are used in the experiments. The whole scene consists of 145×145 pixels and 16 classes of interested objects with the size ranging from 20 to 2468 pixels, 9 classes are used in the experiments. One example is shown in Figure 2.

The hyperspectral unmixing algorithms proposed in this work have been tested using the public domain Indian Pines hyper-spectral dataset which has been previously used in many different studies. This image was obtained from the AVIRIS imaging spectrometer at Northern Indiana on June 12, 1992 from a NASA ER2 flight at high altitude with ground pixel resolution of 17 meters. The dataset comprises 145*145 pixels and 220 bands of sensor radiance without atmospheric correction. It contains two thirds of agriculture, and one third of forest, two highways, aril lane and some houses. Ground truth determines sixteen different classes. Water absorption bands (104-108, 150-163 and 220) were removed, obtaining a 200 band spectrum at each pixel. Figure 3 indicates the distribution of different kinds of materials on Indian Pines and Table 1. records the amount of samples of different materials.



(a) Three band false color composite



(b) Spectral signatures

Figure 2. One example of Indian Pines data

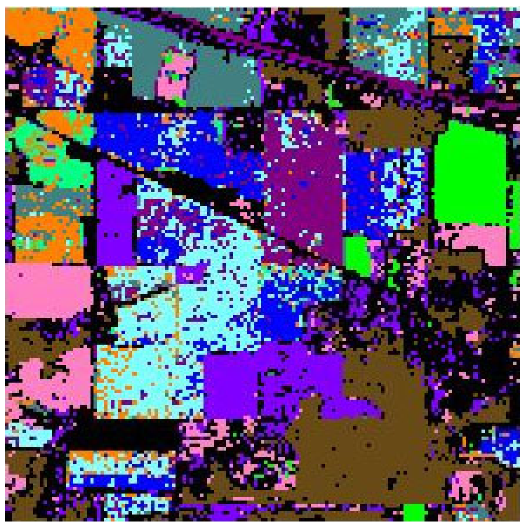


Figure 3. Indian Pines

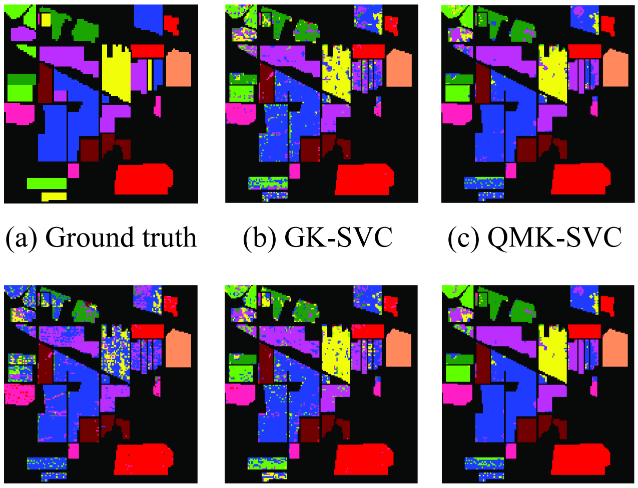
Table 1. Groundtruth classes for the Indian Pines and their sample numbers

Num	Class	Samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93

4.2 Algorithm Evaluaiton

The cross-validation method is applied to select the procedure parameters. The polynomial kernel and Gaussian kernel are chosen as the basic kernels. In different application systems, we can choose the type of kernel functions of the offline training system. We implement the optimized multi-kernel learning on Indian Pines dataset and AVIRIS Salinas Dataset, compared with Support Vector Classifier (SVC), Kernel Sparse Representation Classifier (KSRC), SVM [44], RMKL-SVM [45] and POL-KSRC [46]. The classification results denote that the feasibility in the intelligent hyperspectral imaging instrument.

Firstly, we implement some experiments to evaluate the proposed multiple quasiconformal kernel based on Support Vector Classifier (SVC), Kernel Sparse Representation Classifier (KSRC) on data classification. We test the single kernel and quasiconformal multikernels for kernel classifiers on SVC and KSRC, that is, PK-SVC: Polynomial Kernel-SVC, GK-SVC:Gaussian Kernel-SVC, QMK-SVC: Quasiconformal Multi-kernels Based SVC, PK- KSRC: Polynomial Kernel- KSRC, GK- KSRC: Gaussian Kernel- KSRC, QMK- KSRC: Quasiconformal Multi-kernels Based KSRC. The experimental results are shown in Figure 4. Compared with the truth, QMK-SVC perofor better than GK-SVC, and QMK-KSRC perofor better than GK-KSRC. So the proposed method is more fitted to the hyperspectral sensing data analysis in the intelligent hyperspectral imaging instrument. In the practical applications, the multiple bands of hyperspectral data are changed to the single band of image. Thus, for the intelligent hyperspectral instrument the multiple bands of sensing data are changed to the single data.



(a) Ground truth (b) GK-SVC (c) QMK-SVC

(d) PK- KSRC (e) GK- KSRC (f) QMK- KSRC

Figure 4. Performance basic kernel and optimized kernel on classification on Indian Pines data

For the quantitative comparison, we implement some experiments using the averaged accuracy to evaluate the performance of the algorithms. The experimental results are shown in Table 2, Table 3 and Table 4. For the SVC, QMK-SVC performs better than PK-SVC and GK-SVC. For the KSRC, QMK-KSRC outperforms PK-KSRC and GK-KSRC. On the multiple kernels, Gaussian kernel and Polynomial kernel are as the basic kernels for the combination. The classification results denote that the feasibility in the intelligent hyperspectral imaging instrument. And in the practical application the multiple bands of hyperspectral data are changed to the single band of image. Thus, for the intelligent hyperspectral instrument the multiple bands of sensing data are changed to the single data.

Table 2. Performance of SVC on the Indian Pines data (%)

Class	1	2	3	4	5	6
PK-SVC	49.3	58.7	96.4	39.2	65.8	93.6
GK-SVC	78.0	73.6	99.1	76.9	80.5	97.1
Class	7	8	9	10	11	12
PK-SVC	62.9	85.3	100	65.8	72.3	58.4
GK-SVC	79.7	89.8	99.7	83.6	86.0	80.7

Table 3. Performance of KSRC on the Indian Pines data (%)

Class	1	2	3	4	5	6
PK-KSRC	51.8	59.6	96.1	49.1	78.5	93.8
GK-KSRC	77.8	76.4	99.1	75.5	79.0	97.4
Class	7	8	9	10	11	12
PK-KSRC	62.8	84.7	100	67.5	75.2	60.7
GK-KSRC	82.7	88.7	100	83.9	86.3	81.1

Table 4. Performance on the Pavia University data (%)

Class	POL-SVM	RBF-SVM	SMKL-SVM	RMKL-SVM	POL - KSRC	RBF - KSRC
p1	83.97	84.46	84.25	88.24	84.94	85.43
p2	85.25	91.42	90.26	93.14	86.23	92.46
p3	70.24	74.15	74.67	78.24	71.25	75.17
p4	87.25	90.42	90.26	89.26	88.26	91.42
p5	96.25	97.67	97.25	97.27	97.23	98.67
p6	70.45	78.78	77.48	84.24	71.46	79.71
p7	69.35	71.15	70.16	76.22	69.91	72.17
p8	76.24	81.53	80.25	82.95	77.26	82.51
p9	98.57	99.46	99.26	99.22	98.86	99.98

The experimental results on hyperspectral image databases show that the optimized multiple kernels-based machine learning achieves better performance than those of other methods on the hyperspectral data analysis for the intelligent hyperspectral imaging instrument. The kernel-based learning machine is to solve the data mapping through the selection of function and parameters of kernels. The optimized multiple kernels are combined to more precisely characterize the hyperspectral sensing data from the intelligent hyperspectral instrument for improving the performance of solving complex visual learning tasks. The experimental results show that the proposed framework outperforms others on hyperspectral data analysis for the intelligent hyperspectral instrument. In the practical application the multiple bands of hyperspectral data are changed to the single band of image. Thus, for the intelligent hyperspectral instrument the multiple bands of sensing data are changed to the single data. The classification results denote that the feasibility in the intelligent hyperspectral imaging instrument. The performance of the practical application is evaluated with the classification result. In the system, the cross-validation method is applied to select the procedure parameters, and the polynomial kernel and Gaussian kernel are chosen as the basic kernels. The type of kernel function in the offline training system is chosen through cross-validation method.

4.3 Perfomance Evaluation

In the part of deep learning, the iteration times is 2000; the nodes of the tree layer are 500, 500, 2000. In the SVM part we use the LIBSVM to solve this optimization problem and the parameters applied are default values. Due to the amount of the data and the effectiveness of the method, the ratio between training samples and test samples within the class is set up as 10:34.

We have conducted experiments on Indian Pines using deep learning solely with iteration number of 1, 10, 200, 2000, 3000, 4000. The Accuracy have reached the top when the iteration number is 2000 which is 45.25%. Considering that deep learning's goal is to classify all the materials at the same time, it is understandable that the accuracy is hard to be improved. However, our goal is to recognize different kinds of materials from hyperspectral images, it is effectively unnecessary to classify them all at the same time. In other words, single classification is capable of satisfying our requirement.

The recognition accuracy result of 14 classes of materials using deep learning and single-class SVM are persented in Figure 5.

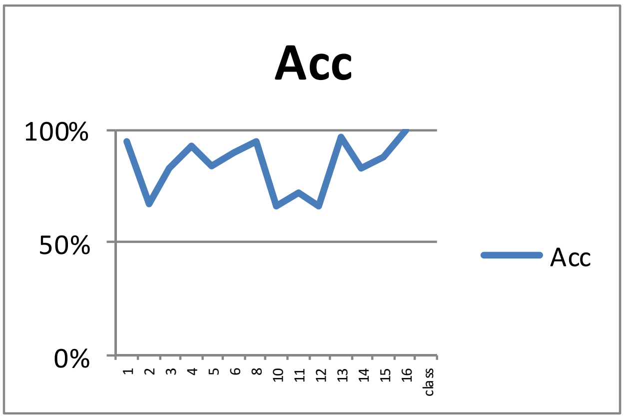


Figure 5. Recognition accuracy of single-class SVM

Some conclusions can be obtained from Figure 6, First, DLSVM perform better result than those reported previously: simple Euclidean and LOOC-based method [47-48]. The average accuracy have reached 83.99%. Comparing to the accuracy of deep learning, DLSVM have accomplished a great improvement.

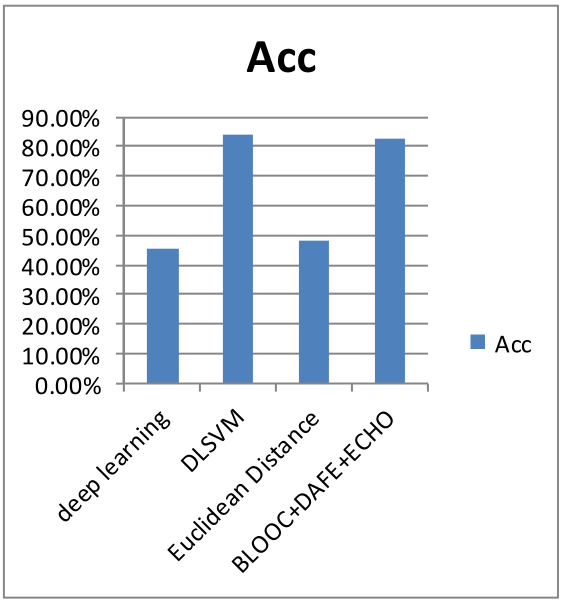


Figure 6. Recognition accuracy of different classifiers

4.4 Discussion

Judging from Figure 5, the performance of our method varies upon different materials. However, most materaials reached accuracy higher than 80%, few reached higher than 95% using training data against testing data at the ratio of 29.4%. Proving our method to be an effective way of classifying hyperspectral images and a promising algorithm for further work.

5 Conclusion

In conclusion, we have shown that DLSVM works better than softmax on standard dataset Indian Pines and the result exhibits the efficiency of DLSVM upon recognition of hyperspectral images. Besides, the substitution from from softmax to SVM appears to be smooth and steady. Further research could be focused on multiclass SVM formulations and other hyperspectral images.

References

[1] Y. Cui, J. Q. Wang, S. B, Liu, L. G. Wang, Hyperspectral Image Feature Reduction Based on Tabu Search Algorithm, *Journal of Information Hiding & Multimedia Signal Processing*, Vol. 6, No. 1, pp. 154-162, April, 2015

[2] J. S. Pan, Y. F. Li, J. B. Li, L. Li, P. W. Tsai, Q. Su, W. Cui, Hyperspectral Imagery Classification Based on Kernel Principal Component Analysis, *Tenth International Conference on Innovative Computing, Information and Control*, Dalian, China, 2015, pp. 20-22

[3] P. Ghamisi, J. A. Benediktsson, J. R. Sveinsson, Automatic Spectral-spatial Classification Framework Based on Attribute Profiles and Supervised Feature Extraction, *IEEE Transactions on Geoscience & Remote Sensing*, Vol. 52, No. 9, pp. 5771-5782, September, 2014.

[4] P. Ghamisi, J. A. Benediktsson, G. Cavallaro, A, Plaza, A. Automatic Framework for Spectral-spatial Classification Based on Supervised Feature Extraction and Morphological Attribute Profiles, *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, Vol. 7, No. 6, pp. 2147-2160, June, 2014.

[5] J. Y. Li, H. Y. Zhang, L. P. Zhang, Supervised Segmentation of Very High Resolution Images by the Use of Extended Morphological Attribute Profiles and a Sparse Transform, *IEEE Geoscience & Remote Sensing Letters*, Vol. 11, No. 8, pp. 1409-1413, August, 2014.

[6] B. Song, J. Li, M. D. Mura, P. J. Li, A. Plaza, J. M. Bioucas-Dias, J. A. Benediktsson, J. Chanussot, Remotely Sensed Image Classification Using Sparse Representations of Morphological Attribute Profiles, *IEEE Transactions on Geoscience & Remote Sensing*, Vol. 52, No. 8, pp. 5122-5136, August, 2014.

[7] P. Ghamisi, M. D. Mura, J. A. Benediktsson, A Survey on Spectral-spatial Classification Techniques Based on Attribute Profiles, *IEEE Transactions on Geoscience & Remote Sensing*, Vol. 53, No. 5, pp. 2335-2353, May, 2015.

[8] S. Ma, J. Geng, H. Y. Wang, Hyperspectral Image Classification via Contextual Deep Learning, *Eurasip Journal on Image & Video Processing*, Vol. 2015, No. 1, pp. 20, July, 2015.

[9] Y. Chen, N. M. Nasrabadi, T. D. Tran, Hyperspectral Image Classification via Kernel Sparse Representation, *IEEE Transactions on Geoscience & Remote Sensing*, Vol. 51, No. 1, pp. 217-231, January, 2013.

[10] Y. Tarabalka, Classification of Hyperspectral Data Using Spectral-spatial Approaches. Numerical and Experimental Investigation of Plasticity (slip) Evolution in Notched Single Crystal Superalloy Specimens, Ph.D. Thesis, University of Iceland and Grenoble Institute of Technology, Reykjavík, Iceland, 2011.

[11] Y. C. Tang, *Deep Learning Using Linear Support Vector Machines*, Computer Science, 2013.

[12] B. E. Boser, I. M. Guyon, V. N. Vapnik, A Training Algorithm for Optimal Margin Classifiers, *The Workshop on Computational Learning Theory*, Pittsburgh, PA, 1992, pp. 144-152.

[13] F. J. Huang, Y. Lecun, Large-scale Learning with SVM and Convolutional for Generic Object Categorization, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, NY, 2006, pp. 284-291.

[14] J. Nagi, G. A. Di caro, A. Giusti, F. Nagi, L. M. Gambardella, Convolutional Neural Support Vector Machines: Hybrid Visual Pattern Classifiers for Multi-robot Systems, *2012 11th International Conference on Machine Learning and Applications*, Boca Raton, FL, 2012, pp. 27-32.

[15] R. B. Salem, K. S. Ettabaa, M. A. Hamdi, A. Spectral-spatial Classification of Hyperspectral Images Using Different Spatial Features and Composite Kernels, *International Image Processing, Applications and Systems Conference*, Sfax, 2015, pp. 1-7.

[16] G. Camps-Valls, J. Calpe, Composite Kernels for Hyperspectral Image Classification, *IEEE Geoscience & Remote Sensing Letters*, Vol. 3, No. 1, pp. 93-97, January, 2006.

[17] M. Fauvel, J. Chanussot, J. A. Benediktsson, J. R. Sveinsson, Spectral and Spatial Classification of Hyperspectral Data Using Svms and Morphological Profiles, *IEEE Transactions on Geoscience & Remote Sensing*, Vol. 46, No. 11, pp. 3804-3814. August, 2008.

[18] X. D. Kang, S. T. Li, J. A. Benediktsson, Spectral-spatial Hyperspectral Image Classification with Edge-preserving Filtering, *IEEE Transactions on Geoscience & Remote Sensing*, Vol. 52, No. 5, pp. 2666-2677, April, 2014.

[19] X. Huang, L. P. Zhang, An Svm Ensemble Approach Combining Spectral, Structural, and Semantic Features for the Classification of High-resolution Remotely Sensed Imagery, *IEEE Transactions on Geoscience & Remote Sensing*, Vol. 51, No. 1, pp. 257-272, January, 2013.

[20] S. Amari, S. Wu, Improving Support Vector Machine Classifiers by Modifying Kernel Functions, *Neural Networks*, Vol. 12, No. 6, pp. 783-789, October, 1999.

[21] J. Li, L. Yu, S. Sun, Refined Kernel Principal Component Analysis Based Feature Extraction, *Chinese Journal of Electronics*, Vol. 20, No. 3, pp. 467-470, October, 2011.

[22] P. N. Belhumeur, J. P. Hespanha, D. Kriegman, Eigenfaces vs.

Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711-720, January, 1997.

[23] Y. Ming-Hsuan, Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods, *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, Washington D.C., 2002, pp. 215-220.

[24] G. Baudat, F. E. Anouar, Generalized Discriminant Analysis Using a Kernel Approach, *Neural Computation*, Vol. 12, No. 10, pp. 2385-2404, November, 2000.

[25] Z. Z. Liang, P. F. Shi, Uncorrelated Discriminant Vectors Using a Kernel Method, *Pattern Recognition*, Vol. 38, No. 2, pp. 307-310, Feburary, 2005.

[26] J. W. Lu, K. Plataniotis, A. N. Venetsanopoulos, Face Recognition Using Kernel Direct Discriminant Analysis Algorithms, *IEEE Transactions on Neural Networks*, Vol. 14, No. 1, pp. 117-226, Feburary, 2003.

[27] J. S. Pan, J. B. Li, Z. M. Lu, Adaptive Quasiconformal Kernel Discriminant Analysis, *Neurocomputing*, Vol. 71, No. 13-15, pp. 2754-2760, August, 2008.

[28] L. Wang, K. L. Chan, P. Xue, A Criterion for Optimizing Kernel Parameters in KBDA for Image Retrieval, *IEEE Trans. Systems, Man and Cybernetics-Part B: Cybernetics*, Vol. 35, No. 3, pp. 556-562, July, 2005.

[29] J. B. Li, J. S. Pan, S. C. Chu, Kernel class-wise Locality Preserving Projection, *Information Sciences*, Vol. 178, No. 7, pp. 1825-1835, April, 2008.

[30] G. Y. Feng, D. W. Hu, D. Zhang, Z. T. Zhou, An Alternative Formulation of Kernel LPP with Application to Image Recognition, *Neurocomputing*, Vol. 69, No. 13-15, pp. 1733-1738, August, 2006.

[31] H. L. Xiong, M. N.s. Swamy, M. O. Ahmad, Optimizing the Kernel in the Empirical Feature Space, *IEEE Transactions on Neural Networks*, Vol. 16, No. 2, pp. 460-474, April, 2005.

[32] Q. Zhu, Reformative Nonlinear Feature Extraction Using Kernel MSE, *Neurocomputing*, Vol. 73, No. 16-18, pp. 3334-3337, October, 2010.

[33] J. Huang, P. C. Yuen, W. S. Chen, J. H. Lai, Kernel Subspace LDA with Optimized Kernel Parameters on Face Recognition, *2004 Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, Seoul, South Korea, 2004, pp. 327-332.

[34] J. Cheng, Q. S. Liu, H. Q. Lu, Y. W. Chen, Supervised Kernel Locality Preserving Projections for Face Recognition, *Neurocomputing*, Vol. 67, pp. 443-449, August, 2005.

[35] H. T. Zhao, S. Y. Sun, Z. L. Jing, J. Y. Yang, Rapid and Brief Communication: Local Structure Based Supervised Feature Extraction, *Pattern Recognition*, Vol. 39, No. 8, pp. 1546-1550, August, 2006.

[36] X. D. Xie, B. H. Li, X. D. Chai, A Framework of Quasiconformal Mapping-Based Kernel Machine with Its Application to Hyperspectral Remote Sensing, *Measurement*, Vol. 80, pp. 270-280, November, 2015.

[37] W. S. Chen, P. C. Yuen, J. Huang, D. Q. Dai, Kernel Machine-Based One-Parameter Regularized Fisher Discriminant Method for Face Recognition, *IEEE Trans. Systems, Man and Cybernetics-Part B: Cybernetics*, Vol. 35, No. 4, pp. 658-669, September, 2005.

[38] D. Tuia, G. Camps-Valls, G. Matasci, M. Kanevski, Learning Relevant Image Features with Multiple-kernel Classification, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48, No. 10, pp. 3780-3791, November, 2010.

[39] N. Subrahmanya, Y. Shin, Sparse Multiple Kernel Learning for Signal Processing Applications, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 5, pp. 788-798, May, 2010.

[40] S. Sonnenburg, G. Rätsch, C. Schäfer, B. Schölkopf, Large Scale Multiple Kernel Learning, *Journal of Machine Learning Research*, Vol. 7. pp. 1531-1565, July, 2006.

[41] M. Kloft, lp-Norm Multiple Kernel Learning, *Journal of Machine Learning Research*, Vol. 12, pp. 953-997, January, 2011.

[42] J. S. Pan, L. K. Ping, T. W. Sun, P. W. Tsai, V. Snášel, α-Fraction First Strategy for Hierarchical Model in Wireless Sensor Networks, *Journal of Internet Technology*, Vol. 19, No. 6, pp. 1717-1726, November, 2018

[43] J. S. Pan, L. K. Ping, T. W. Sun, P. W. Tsai, V. Snášel, A Clustering Scheme for Wireless Sensor Networks Based on Genetic Algorithm and Dominating Set, *Journal of Internet Technology*, Vol. 19, No. 4, pp. 1111-1118, January, 2018

[44] A. Rakotomamonjy, F. R. Bach, S. Canu, Y. Grandvalet, Simple MKL, *Journal of Machine Learning Research*, Vol. 9, pp. 2491-2521. November, 2008

[45] Y. F. Gu, C. Wang, D. You, Y. H. Zhang, S. Z. Wang, Y. Zhang, Representative Multiple Kernel Learning for Classification in Hyperspectral Imagery, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 50, No. 7, pp. 2852-2865. July, 2012

[46] H. L. Chen, C. C. Huang, X. G. Yu, X. Xu, X. Sun, G. Wang, S. J. Wang, An Efficient Diagnosis System for Detection of Parkinson's Disease using Fuzzy Knearest Neighbor Approach, *Expert Systems with Applications*, Vol. 40, No. 1, pp. 263-271, January, 2013.

[47] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, G. Trianni, Recent Advances in Techniques for Hyperspectral Image Processing, *Remote Sensing of Environment*, Vol. 113, No. 1, pp. S110-S122. September, 2009.

[48] T. Saldju, D. Landgrebe, *Classification of High Dimensional Data with Limited Training Samples*, Ph.D. Thesis, School of Electrical Engineering and Computer Science, Purdue University, West Lafayette, IN, 1998.

Biographies

Yifan Li is doctoral candidate at Harbin Institute of Technology Department: Innovative Information Industry Research Center of Computer Science and technology. Research Field: Deep learning, Kernel Learning, Hyperspectral image processing.

Junbao Li is professor and Ph.D. supervisor at Harbin Institute of Technology. Department: Electronics and Information Engineering. Research Field: Image processing and Pattern Recognition, Artificial Intelligent.

Jeng-Shyang Pan is professor and Ph.D. supervisor at Fujian University of Technology. Department: School of Information Science and Engineering. Research Field: Information Security, Cloud Computing, Internet of Things, Multi-Media Signal Processing and Recognition.