

Hyperspectral image classification based on discriminative locality preserving broad learning system

Yonghe Chu, Hongfei Lin^{*}, Liang Yang, Dongyu Zhang, Yufeng Diao, Xiaochao Fan, Chen Shen

Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian
116024, China

Abstract: Recently, broad learning system (BLS) has been widely used for its simple, fast and excellent generalization ability in hyperspectral image (HSI) classification. However, how to implement a broad learning system for fine-grained classification of hyperspectral images with a few-shot setting is still a challenging problem. In this paper, we proposed a new method based on the discriminative locality preserving broad learning system (DPBLS) for hyperspectral image classification by exploiting the manifold structure between neighbouring pixels of hyperspectral image. To make full use of the spectral and spatial information of hyperspectral images, we firstly leverage edge-preserving filters to fuse both spectral and spatial features of hyperspectral image samples. Secondly, we introduce discriminative information and local manifold structure of samples into the broad learning system to enhance the discriminative ability of output weights and improve its performance on hyperspectral image classification task. In order to verify the performance of the framework proposed in this paper, we conducted experiments on four hyperspectral image datasets. experiment results show that the method we proposed is well-performed on hyperspectral image classification tasks.

Keywords: broad learning system; hyperspectral images; information fusion; local manifold structure.

1. Introduction

Hyperspectral image (HSI) classification makes fully use of hyperspectral images with characteristics of the unification of their graph and spectral, as well as the informative richness of their spectral, classifying each pixel in the image, and resulting to a high accuracy on the classification and identification of ground objects. This task is also an important component of earth observation. Classifying hyperspectral remote sensing images accurately plays a significant role in practical applications of various fields, such as fine agriculture [1], astronomy [2], environmental science [3] and others.

For solving hyperspectral image classification problem, researchers have proposed a large number of classification methods [4-13], such as support vector machine (SVM) [4], K-nearest neighbor method [5], decision tree [6], extreme learning machine (ELM) [7] and etc. Recently, joint spectral-spatial methods have been proved to be efficient and effective in merging dual spectral and spatial information to improve precision of classification [14]. The extraction of spatial information is generally based on an assumption that adjacent pixels in one image are similar. For example, support vector machine-markov random fields [15] method, joint sparse representation method [16], total variation (TV) method [17], and augmented morphology by extending spatial morphological profiles [18]. In order to exploit the joint spectral-spatial information, Wang et al. [19] applies the adaptive scalable kernel (ASK) to hyperspectral image

^{*} Corresponding author E-mail: hflin@dlut.edu.cn.

classification. Maggu et al. [20] applied transform learning to the hyperspectral image classification. Kang et al. [21] proposed the PCA-EPFs (PCA-EPFs) method, which used Edge-Preserving Features (EPFs) to extract spatial and spectral information from the hyperspectral images and then integrated both spatial and spectral information with the principal component analysis (PCA). These methods effectively improve the accuracy of classification by incorporating spatial information. The composite kernel (CK) method [22] abstracts spatial information of a target pixel by extracting spatial features (for example, the mean or variance of spectral) within a square neighbourhood. This kernel-based method not only obtains non-linear features from data, but also adds spatial information, reaching a high accuracy in classifying. In order to obtain spatial characteristics of HSI, Marconcini et al. [23] introduced Composite Kernels into SVM and proposed support vector machine with composite kernel (SVMCK), Zhou et al. [24] introduced Composite Kernels into ELM and proposed extreme learning machine with composite kernel (ELMCK) and kernel extreme learning machine with composite kernel (KELMCK). Unlike the CK method, which relies on spatial filters to extract spatial information, the spectral-spatial kernel (SSK) [25] method directly learns similarity features of samples in a high-dimensional kernel feature space. As a result, this method can reflect complex manifold of data samples hiding in the dimensional space.

Recently, Chen et al. [26] proposed a broad learning system and provided an efficient learning framework for machine learning and pattern recognition. Instead of expanding neurons in deep stack layers, this method expands the neurons with feature map nodes and enhancement nodes from a broad view and calculates output weights through a pseudo inverse. Compared with deep learning algorithms, BLS could provide an alternative method for data processing in large-scale, and its training process could be extended to incremental learning models without retraining after adding new nodes. Subsequently, Chen et al. proved the general approximation ability of BLS. This method has achieved excellent results in a lot of applications, such as time series prediction [27], regression [28-29], and image classification [30-34].

Kong et al. [27] applied BLS to hyperspectral image classification. Aiming at solving the limited quantity of hyperspectral images for model training, Kong et al. [30] proposed a semi-supervised broad learning system. Kong et al. [30] introduced a class probability framework into broad learning system and calculated the relationship matrix between labelled samples and unlabelled samples, then obtaining pseudo labels for unlabelled samples. Finally, the ridge regression theory is leveraged to calculate the output layer weights of the semi-supervised broad learning system. However, how to improve the fine-grained classification performance of the broad learning system for hyperspectral images with few labelled data remains a challenging problem. In this paper, we introduce the discriminative information and manifold structure information of hyperspectral image data into BLS from the perspective of manifold learning. Researchers have pointed out that manifold learning methods could effectively reveal the inherent geometric structure of data points [35-39]. Supposing that two samples x_1 and x_2 are from the same marginal distribution P_x , the conditional probabilities would be more similar on the scenario of the two point A and B are approaching each other. This assumption is called smoothing assumption in machine learning [35]. Therefore, it is possible to provide effective information for pattern recognition by mining the geometric structure from the data. Based on the explanation above, we propose a discriminative locality preserving broad learning system (DPBLS). This algorithm draws on the idea of local fisher discriminant analysis (LFDA) [40] and takes the

discriminant information of data samples and the information of local manifold structures into account, introducing a manifold regularization framework, which is constructed by that information, into BLS. Therefore, the output weights could learn more information about both discrimination and local manifold structure from samples, enhancing the ability of BLS on hyperspectral image classification tasks. The main contributions of the algorithm in this paper are as follows:

- (1) We propose a hyperspectral image classification framework based on information fusion of local discriminative broad learning.
- (2) We introduce discriminative information of data samples and the information of local manifold structure into the broad learning model to optimize the projection direction of output weights.

The rest of the paper is arranged as follows: Section 2 gives an introduction one broad learning system, while in Section 3 we propose an information fusion framework based on a discriminative locally preserving broad learning technique. Section 3 is mainly about the experimental analysis, and we make a conclusion about our work in the last section.

2. Related Works

Recently, deep learning algorithms can learn advanced features from data automatically over training. This allows the deep learning model to have a better understanding on the data itself, improving classification accuracy in tasks[41-42]. Hu et al. applied convolutional neural network (CNN) to the study of deep representation methods with spectral features, and its classification result is higher than other traditional SVM algorithms [43]. Subsequently, Romero et al. [44] learned unsupervised sparse features through CNN in a greedy layered manner and used them for the final pixel classification. Kang et al. [45] applied Gabor filtering and deep network for hyperspectral image classification, proposing Gabor filtering based deep network (GFDN). The Gabor filtering extracted spatial features from hyperspectral images, while the deep network is used to classify hyperspectral images with both the spatial features and the spectral features. Xu et al. [46] proposed an efficient deep learning based method, namely, random patches network (RPNNet) for HSI classification, which directly regards the random patches taken from the image as the convolution kernels without any training. Cheng et al. [9] combined spectral features and hierarchical convolutional spatial features, yielding a spectral-spatial feature representation. This yields a feature space in which similar classes are closer together and different classes are farther apart., which are then classified by a SVM. Zhou et al. [10] proposed an effective framework, named compact and discriminative stacked autoencoder (CDSAE), for HSI classification. The proposed CDSAE framework comprises two stages with different optimization objectives, which can learn discriminative low-dimensional feature mappings and train an effective classifier progressively. Chen et al. [47] proposed a CNN-based method extracting features in depth and established a deep finite element model based on a three-dimensional CNN to extract spatial spectrum features from hyperspectral remote sensing images. Then, Li et al. [48] let CNN to learn pixel pair features, thereby improving accuracy in classification tasks. Based on the Bayesian theory, Cao et al. [49] regarded classification results from CNN as likelihood probabilities and leveraged a priori generated by super-pixel segmentation, obtaining classification results eventually. Zhao et al. [50] referred to an image block idea based on pixels neighbourhoods and proposed an improved dimensionality reduction algorithm BLDE based on the locally

discriminative embedding algorithm (LDE). Subsequently, spectral features after dimensionality reducing and features abstracted by CNN were concatenated. And support vector machine and logistic regression classifiers are applied to the final classification.

However, deep network structure is complex and involves many hyperparameters, resulting in deep learning being very time-consuming during model training. In response to the problems above, Chen et al. [26] proposed a broad learning system. The broad learning system has been widely used for its simple structure, fast and good generalization ability. In this paper, based on these advantages of broad learning system, we propose a discriminative locality preserving broad learning system for hyperspectral image classification.

3. Broad learning system

Because the deep network structure is complex and involves a large number of hyper-parameters, Deep learning methods suffer from a time-consuming training process. To solve this problem, Chen et al. [26] proposed a broad learning system based on a feedforward neural network with one hidden layer. Compared with deep learning, in the BLS model, all weights for connecting in the hidden layer are randomly generated, and only the output weights between the hidden layer and the output layer are solved by the least square method. The BLS structure is demonstrated in Figure 1.

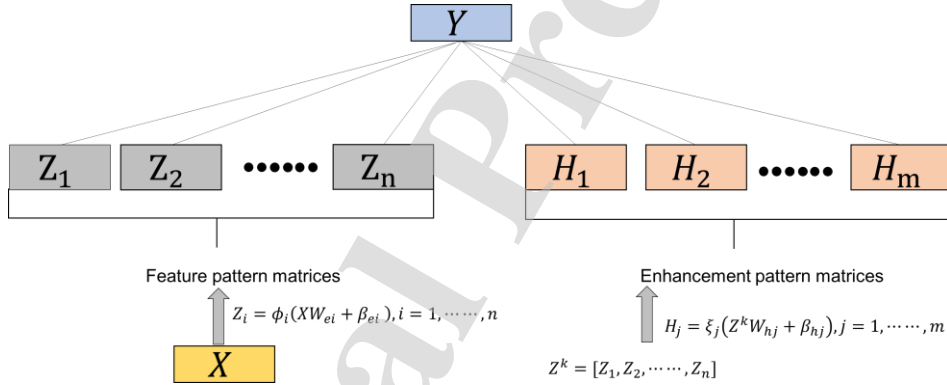


Figure 1. The structure of BLS

In figure 1, $X = [x_1^T, x_2^T, \dots, x_N^T]^T \in R^{N \times M}$ is the training sample, $Y = [y_1^T, y_2^T, \dots, y_N^T]^T \in R^{N \times C}$, T is the expected output matrix. where $x_i = (x_{i1}, x_{i2}, \dots, x_{iM}) \in R^M$, $y_i = (y_{i1}, y_{i2}, \dots, y_{iC}) \in R^C$, N, M are the total number of data samples and the dimension of the data samples respectively, while C is the number of categories of data samples. Z_1, Z_2, \dots, Z_n is the feature node matrix, H_1, H_2, \dots, H_m is the enhanced node matrix, and W is the output weight matrix between the hidden layer and the output layer. The training set X is firstly passed through the k -group feature map to form the k -group feature node matrix Z_1, Z_2, \dots, Z_n :

$$Z_i = \phi_i(XW_{ei} + \beta_{ei}) \in R^{N \times q}, i = 1, 2, \dots, n. \quad (1)$$

In formula (1), ϕ_i is a freely selectable activation function, and q is the number of feature nodes corresponding to each group of feature maps. $W_{ei} \in R^{M \times q}$ and $\beta_{ei} \in R^{N \times q}$ are randomly generated input weight matrix and bias matrix, respectively.

The randomly generated W_{ei} and β_{ei} are fine-tuned by a sparse autoencoder to obtain a sparse representation of the input data. Merging all feature node moments to calculate the total feature node matrix:

$$Z^k = [Z_1, Z_2, \dots, Z_n] \in R^{N \times nq}. \quad (2)$$

Z^k undergoes m -group enhancement transforms to form m -group "enhancement node" matrix H_1, H_2, \dots, H_m :

$$H_j = \xi_j(Z^k W_{hj} + \beta_{hj}) \in R^{N \times r}, j = 1, 2, \dots, m. \quad (3)$$

among them, r represents the number of enhanced nodes corresponding to each group of enhanced transformations; $W_{hj} \in R^{nq \times r}$ and $\beta_{hj} \in R^{N \times r}$ are randomly generated input weight matrix and bias matrix, respectively. ξ_j denotes as the activation function.

Splice all enhanced node matrices to obtain the overall enhanced node matrix:

$$H^m = [H_1, H_2, \dots, H_m] \in R^{N \times mr}. \quad (4)$$

The predicted value of the BLS model can be expressed by the following function:

$$\hat{Y} = [Z_1, Z_2, \dots, Z_n, H_1, H_2, \dots, H_m]W = AW. \quad (5)$$

In formula (5), $A = [Z^n, H^m] \in R^{N \times (nq + mr)}$, W is output weight matrix between hidden layer and output layer.

In order to make the error between the predicted value \hat{Y} and the real value Y as small as possible, we need to find a suitable W . Therefore, the optimization function of the BLS model is as follows:

$$\arg \min_W F_{BLS} = \|Y - \hat{Y}\|^2 + \frac{1}{2} \lambda \|W\|^2 = \|Y - AW\|^2 + \lambda \|W\|^2. \quad (6)$$

In formula (6), the first term is the training error, and the second term is to minimize the output weight, which can avoid over fitting of the BLS model. λ is the regularization parameter. We derive the formula (6) and let $\frac{\partial F_{BLS}}{\partial W} = 0$.

W is calculated as follows:

$$W = (A^T A + \lambda I)^{-1} A^T Y. \quad (7)$$

where I is the identity matrix. Through the process above, we could figure out the optimal W , as the matrix $(A^T A + \lambda I)$ generally nonsingular.

4. Proposed method

In this section, we make an introduction of the overall framework. The HSI classification of information fusion based on discriminative locality preserving broad learning system is shown in Figure 2. Firstly, we apply a feature fusion method [37] to fuse spatial-spectrum joint features, the

hyperspectral images obtained using the spectral-spatial method [37] have 30 features. In Figure 2, to reduce the computing cost spent in the following feature extraction and reduce the image noise of the original HSI, the dimension of the HSI is first reduced with a simple band averaging method, For the dimension-reduced hyperspectral images we use EPF to extract the spatial information of the hyperspectral images, then obtaining their spectral-spatial features. Secondly, we construct a locally discriminative broad learning algorithm. Finally, the fused hyperspectral image is as an input to a discriminative local broad learning model for classifying.

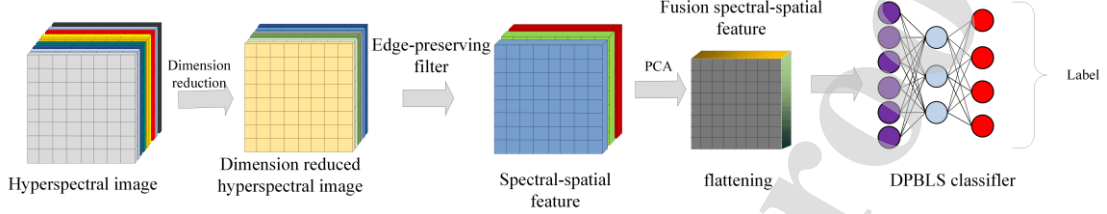


Figure 2. HSI classification based on discriminative locality preserving broad learning system

4.1. features fusion

We use the method of [21] for joint spectral-spatial feature fusion. Firstly, the dimension of an original image is reduced by the PCA, while extracting the joint spectral-spatial features by multiple edge filters. Secondly, the PCA algorithm is used to fuse the joint spectral-spatial features obtained by those filters.

With the K-dimensional hyperspectral data obtained by PCA, the EPFs of the HSI could be easily calculated and stacked together as follows:

$$F_k^x = DTRF(\hat{I}_k, \delta_s^x, \delta_r^x), \quad x=1, \dots, X, k=1, \dots, K. \quad (8)$$

$$F = \{F^1, \dots, F^X\} \quad (9)$$

In formula (8), DTRF refers to the domain transform filtering operation with different parameter settings. \hat{I} is a remote sensing image representation obtained by PCA dimensionality reduction, δ_s^x and δ_r^x are the x -th parameter setting adopted for the domain transform recursive filter. Specifically, for the t -th band of the dimension-reduced band, the EPFs are firstly obtained in case of different parameter settings: $(\delta_s^x, \delta_r^x) \in \{(\delta_s^1, \delta_r^1), \dots, (\delta_s^X, \delta_r^X)\}$. Then, the resulting features are stacked together according to (9). Through formulas (8) and (9), we could obtain multiple joint spectral-spatial features obtained by edge-preserving filtering. Then we use PCA to perform feature fusion on the obtained joint spectral-spatial features. Specifically given by formula (10):

$$P = PCA(F, L), \quad (10)$$

where L is the number of principal components preserved in P and F is the stacked EPFs.

4.2. discriminative locality preserving broad learning system

4.2.1. Motivation of discriminative locality preserving broad learning system

During recent years, researchers have pointed out that manifold learning methods can effectively reveal the inherent geometric structure of data points. Supposing that two samples x_1 and x_2 have a common trait of a same marginal distribution P_x , the conditional probabilities $P(y|x_1)$ and $P(y|x_2)$ would be more similar in the scenario of these two points x_1 and x_2 were

approaching each other. The above assumptions are called smoothing assumptions in machine learning. Therefore, it is possible to provide effective information for pattern classification by mining the geometric structure between the data. Recently, researchers have proposed several manifold learning algorithms, Yang et al.[35] incorporated discriminative analysis and $\ell_{2,1}$ -norm minimization into a joint framework for unsupervised feature selection. Different from existing unsupervised features election algorithms, their algorithm selects the most discriminative feature subset from the whole feature set in batch mode. Afterwards, Yang et al. [39] proposed a new semi-supervised algorithm called ranking with Local Regression and Global Alignment (LRGA) to learn a robust Laplacian matrix for data ranking. In LRGA, for each data point, a local linear regression model is used to predict the ranking scores of its neighboring points. Sugiyama proposed the local fisher discriminant analysis (LFDA) algorithm based on Fisher Discriminant Analysis (FDA) [51], considering the local discriminant information and local geometric structure information of the data samples. Inspired by the aforementioned manifold learning algorithms, we introduce local discriminant information and local discriminant geometric structure information of data samples into the broad learning system. In this paper, we use the LFDA algorithm to construct the manifold regularization framework and then introduce the framework into the broad learning system.

In this section, we propose a discriminative locality preserving broad learning system (DPBLS) based on manifold learning. We firstly introduce the basic principles of the local fisher discriminant analysis (LFDA) algorithm. Secondly, the manifold regularization framework is defined on the basis of the LFDA algorithm. Finally, we introduce the manifold regularization framework into DPBLS for system optimization.

4.2.1. Local fisher discriminant analysis

FDA is one of the commonly used techniques for dimensionality reduction. The formulated description of FDA is as followed. $X = [x_1^T, x_2^T, \dots, x_N^T]^T \in R^{N \times M}$ denotes as the training set,

while $Y = [y_1^T, y_2^T, \dots, y_N^T]^T \in R^{N \times C}$ is the expected output matrix, where $x_i = (x_{i1}, x_{i2}, \dots, x_{iM}) \in R^M$,

$y_i = (y_{i1}, y_{i2}, \dots, y_{iC}) \in R^C$. N , M and C represent the total number of samples in the training set,

the dimension of each sample, and the number of categories for samples respectively. Defining S_W as the intra-class divergence matrix and S_B as the inter-class divergence matrix:

$$S_W = \sum_{c=1}^C \sum_{j=1}^{N_c} (x_j^{(c)} - m^{(c)})^T (x_j^{(c)} - m^{(c)}). \quad (11)$$

$$S_B = \sum_{c=1}^C N_c (m^{(c)} - m)^T (m^{(c)} - m). \quad (12)$$

In formulas (11) and (12), N_c is the total number of samples in the c -th class, and $x_j^{(c)}$ is the

c -th sample in the j -th class. $m^{(c)}$ is the mean vector of the c -th class; m is the mean vector of all samples, and C means the total number of categories of the dataset. LDA has the following

optimization guidelines:

$$L_{LDA} = \max_U \frac{Tr(U^T S_B U)}{Tr(U^T S_W U)}. \quad (13)$$

Formula (12) obtains the projection transformation matrix U by the Lagrange multiplier method and calculates the corresponding low-dimensional expression of X by $Y = XU$.

Local fisher discriminant analysis is improved by the FDA. Firstly, S_W and S_B are transformed into:

$$\tilde{S}_W = \frac{1}{2} \sum_{i,j=1}^N W_{ij}^w (x_i - x_j) (x_i - x_j)^T. \quad (14)$$

$$\tilde{S}_B = \frac{1}{2} \sum_{i,j=1}^N (x_i - x_j) (x_i - x_j)^T. \quad (15)$$

among them, N is the total number of samples.

$$W_{ij}^w = \begin{cases} \frac{1}{N_c}, & \text{if } c(x_i) = c(x_j) \\ 0, & \text{if } c(x_i) \neq c(x_j) \end{cases} \quad (16)$$

$$W_{ij}^w = \begin{cases} \frac{1}{N} - \frac{1}{N_c}, & \text{if } c(x_i) = c(x_j) \\ \frac{1}{N}, & \text{if } c(x_i) \neq c(x_j) \end{cases} \quad (17)$$

among them, $c(x_i)$ represents the category to which the sample x_i belongs. LFDA introduced the idea of LPP algorithm with extended S_W, S_B into local intra-class divergence matrix and local inter-class divergence matrix, respectively. Keeping the formulas (14) and

(15) unchanged, adjusting W_{ij}^w and W_{ij}^b by:

$$W_{ij}^w = \begin{cases} \frac{A_{ij}}{N_c}, & \text{if } c(x_i) = c(x_j) \\ 0, & \text{if } c(x_i) \neq c(x_j) \end{cases} \quad (18)$$

$$W_{ij}^w = \begin{cases} A_{ij} \left(\frac{1}{N} - \frac{1}{N_c} \right), & \text{if } c(x_i) = c(x_j) \\ 0, & \text{if } c(x_i) \neq c(x_j) \end{cases} \quad (19)$$

among them, A_{ij} represents the adjacency weights of x_i and x_j , generally calculated

by a thermal kernel function:

$$A_{ij} = \exp \left(- \frac{\|x_i - x_j\|^2}{t} \right). \quad (20)$$

In formula (20), t is a thermonuclear parameter.

LFDA has the following optimization guidelines:

$$L_{LFDA} = \max_{\tilde{U}} \frac{Tr(\tilde{U}^T \tilde{S}_B \tilde{U})}{Tr(\tilde{U}^T \tilde{S}_W \tilde{U})}. \quad (21)$$

Formula (21) obtains the projection transformation matrix \tilde{U} by the Lagrange multiplier method and then calculates the corresponding low-dimensional expression of X by $Y = X\tilde{U}$.

4.2.2. Manifold regularization framework

We define the manifold regularization framework on the basis of the LFDA algorithm. Considering the singularity of the manifold regularization framework, we apply the maximum marginal criterion [52] to solve problems above. The manifold regularization frame loss function is shown in formula (22):

$$J = \min_{\tilde{U}} Tr(\tilde{U}^T \tilde{S}_W \tilde{U} - \tilde{U}^T \tilde{S}_B \tilde{U}). \quad (22)$$

4.2.3. Discriminative locality preserving broad learning system

In this subsection, we would demonstrate how to introduce the manifold regularization framework into the BLS model. In the BLS algorithm, $Y = AW$, we replace \tilde{U} with W in the manifold regularization framework, The DPBLS algorithm model can be written as follows:

$$\arg \min_W L_{DPBLS} = \|Y - AW\|^2 + \lambda_1 Tr(W^T \tilde{S}_W W - W^T \tilde{S}_B W) + \lambda_2 \|W\|^2. \quad (23)$$

In formula (23), $\|Y - AW\|^2$ is the training error and $Tr(W^T \tilde{S}_W W - W^T \tilde{S}_B W)$ is the manifold regularization frame. $\|W\|^2$ is the l_2 -norm regularization term, λ_1 and λ_2 are regularization parameters to balance the impact of these two terms.

We derive the formula (23) $\partial L_{DPBLS} / \partial W = 0$:

$$\begin{aligned} \frac{\partial L_{DPBLS}}{\partial W} &= \frac{\partial (\|Y - AW\|^2 + \lambda_1 Tr(W^T \tilde{S}_W W - W^T \tilde{S}_B W) + \lambda_2 \|W\|^2)}{\partial W} \\ &= \frac{\partial (Tr[(Y - AW)^T (Y - AW)] + \lambda_1 Tr(W^T \tilde{S}_W W - W^T \tilde{S}_B W) + 2\lambda_2 Tr(W^T W))}{\partial W} \\ &= \frac{\partial (Tr(W^T A^T AW - 2W^T A^T Y + Y^T Y) + Tr(W^T \tilde{S}_W W - W^T \tilde{S}_B W) + 2\lambda_2 Tr(W^T W))}{\partial W} \\ &= (2A^T AW - 2A^T Y) + 2\lambda_1 (\tilde{S}_W - \tilde{S}_B)W + 2\lambda_2 W \end{aligned} \quad (24)$$

We order $\frac{\partial L_{DPBLS}}{\partial W} = 0$, according to formula (24), we can obtain the following formula:

$$(A^T AW - A^T Y) + \lambda_1 (\tilde{S}_W - \tilde{S}_B)W + \lambda_2 W = 0. \quad (25)$$

then the output weight matrix is calculated according to formula (25):

$$W = [A^T A + \lambda_1 (\tilde{S}_W - \tilde{S}_B) + \lambda_2 I]^{-1} A^T Y. \quad (26)$$

where I is the identity matrix.

Based on the above derivation, the steps of the hyperspectral image classification framework based on information fusion of local discriminative broad learning system are as follows:

Hyperspectral image classification of information fusion based on discriminative locality preserving broad learning system

Input: Hyperspectral image X ;

Output: w ;

Initialize: X

1. PCA-EPF

-
- (1) Obtain \hat{I} based on PCA for X
 - (2) Generate F based on EPF for \hat{I}
 - (3) Obtain P by Eq.(10)
 - (4) Generate training set, testing set based on EPF for P
 - End
 2. DPBLS
 - (5) for $i=0:n$ do
 - (6) randomize W_{ei} , β_{ei} , $i=1,2,\dots,n$;
 - (7) Calculating $Z_i = \phi_i(XW_{ei} + \beta_{ei}) \in R^{N \times q}, i=1,2,\dots,n$;
 - End
 - (8) Set the feature mapping group $Z^k = [Z_1, Z_2 \dots, Z_n] \in R^{N \times nq}$;
 - (9) for $j=0:m$ do
 - (10) randomize W_{hj} , β_{hj} , $j=1,2,\dots,m$;
 - (11) Calculating $H_j = \xi_j(Z^n W_{hj} + \beta_{hj}) \in R^{N \times r}, j=1,2,\dots,m$;
 - End
 - (12) Set $A = [Z^n, H^m] \in R^{N \times (nq+mr)}$;
 - (13) Calculating the manifold regularization framework by means of equations (11) - (22);
 - (14) Calculating the output weight matrix via equations (23)-(26);
 - End
 - return W
-

5. Experiments

In this section, in order to verify the effectiveness of the proposed algorithm, we use four hyperspectral image datasets for experiments. All the experiments are performed using MATLAB 2015b on a computer with an Intel(R) Core(TM) CPU i7-6700 and 8GB of RAM.

5.1. Experimental dataset

We carry out experiments on four commonly used datasets on hyperspectral images¹: the Indian Pines dataset, the Pavia of University dataset, the Salinas dataset and KSC dataset.

The Indian Pines dataset was acquired by an AVIRIS spectrometer in the Indian Pines area in western North Indiana. This dataset consists of 145×145 pixels with 220 spectral channels. The spectral range and the spatial resolution are 0.4-2.5 μ m and 20m. After removing 20 water absorption bands, the remaining 200 bands were for training, testing and validating. Figures 3 (a) and 3 (b) show a pseudo-color image and a ground truth image of Indian Pines. The number of training samples and test samples of the dataset are given in Table 1. Each feature category

¹ The four HSI benchmarks are available from: [http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral Remote Sensing Scenes](http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes).

randomly selects 100 samples as the training set, while the remaining are as the test set.

The Pavia of University image dataset was obtained by the ROSIS-03 optical sensor to capture the urban area of the University of Pavia. The scene has 610×340 pixels and the spatial resolution is as high as 1.3m. In the Pavia of University dataset, the original number of spectral bands is 115, 12 noise spectral bands are removed, and the remaining 103 spectral bands are retained in the experiment. Figures 4 (a) and 4 (b) show the pseudo-color image and ground truth image of Pavia of University. The number of training samples and test samples of the data set are demonstrated in Table 2. Each feature category randomly selects 100 samples as the training set and the remaining are as the test set.

The Salinas dataset was acquired in 1998 by the AVIRIS sensor in the Salinas Valley, California. Each image in this dataset is in size of 512×217 pixels and has a high spatial resolution of 3.7m pixels. There are 224 spectral bands in the Salinas dataset, and 204 of them have been retained in our experiment because of 20 spectral bands for water absorption. In addition, there are 16 feature categories in the dataset. Figures 5 (a) and 5 (b) show a Salinas' pseudo-color image and its ground truth image. The number of samples for training and testing of the dataset is given in Table 3. Each feature category randomly selects 100 samples as the training set, while the remaining are as the test set.

The KSC dataset consists of hyperspectral remote sensing images of the Kennedy Space Center in Florida, USA, as shown in Figure 6. The ground spatial resolution of this remote sensing image is known to be 18m, with 614×512 pixels and a spectral range of 400~ 2500 nm with 224 bands, and the remaining 176 bands were pre-processed for experimental classification studies. The dataset contains 13 classes of known ground objects.

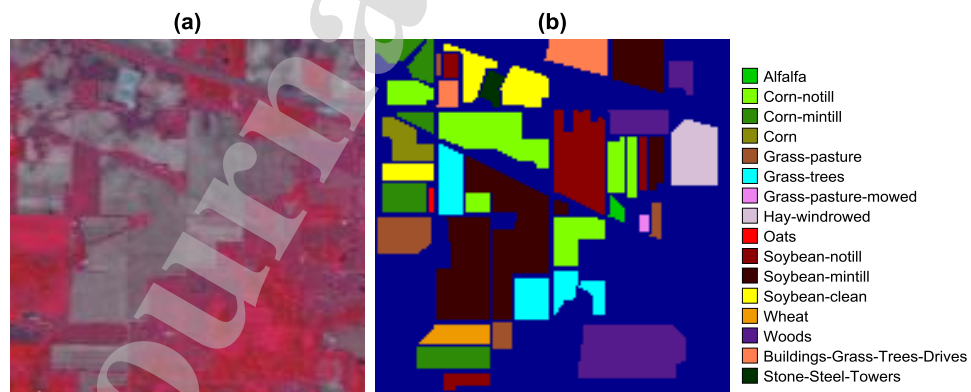


Figure 3. Indian Pines. (a) Colour map (b) Ground truth image

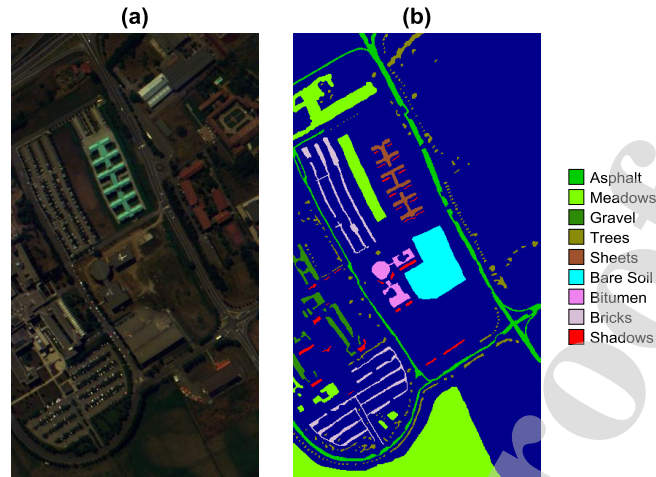


Figure 4. Pavia of University. (a) Colour map (b) Ground truth image

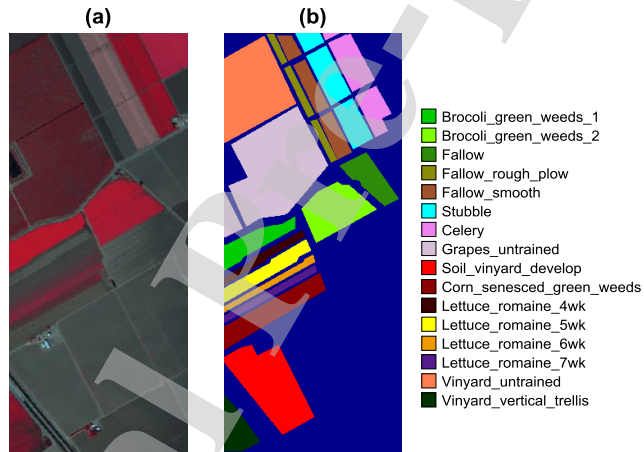


Figure 5. Salinas. (a) Color map (b) Ground truth image

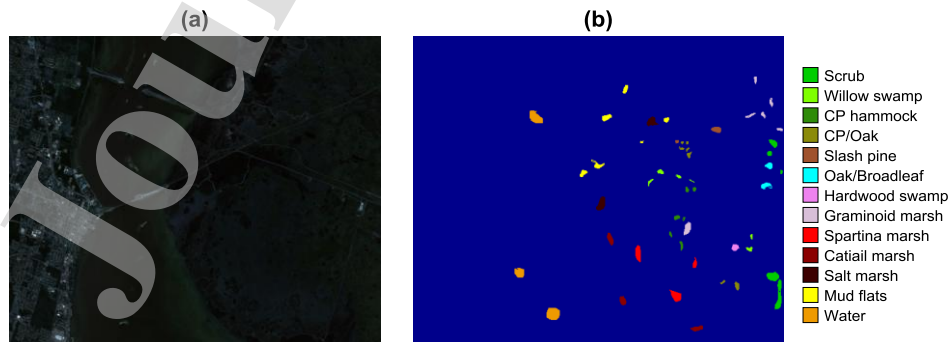


Figure 6. KSC. (a) Color map (b) Ground truth image

Table 1. Indian Pines dataset description

NO.	Class	Numbers of samples	Training-set	Testing-set
1	Alfalfa	46	23	23
2	Corn-notill	1428	100	1328
3	Corn-mintill	830	100	730
4	Corn	237	100	137
5	Grass-pasture	483	100	383
6	Grass-trees	730	100	630
7	Grass-pasture-mowed	28	14	14
8	Hay-windrowed	478	100	378
9	Oats	20	10	10
10	Soybean-notill	972	100	872
11	Soybean-mintill	2455	100	2355
12	Soybean-clean	593	100	493
13	Wheat	205	100	105
14	Woods	1265	100	1165
15	Buildings-Grass-Trees-Drives	386	100	286
16	Stone-Steel-Towers	93	47	46

Table 2. Univerisy of Pavia dataset description

NO.	Class	Numbers of samples	Training	Testing
1	Asphalt	6631	100	6531
2	Meadows	18649	100	18549
3	Gravel	2099	100	1999
4	Trees	3064	100	2964
5	Sheets	1345	100	1245
6	Bare Soil	5029	100	4929
7	Bitumen	1330	100	1230
8	Bricks	3682	100	3582
9	Shadows	947	100	847

Table 3. Salinas dataset description

NO.	Class	Numbers of samples	Training	Testing
1	Brocoli_green_weeds_1	2009	100	1909
2	Brocoli_green_weeds_2	3726	100	3626
3	Fallow	1976	100	1876
4	Fallow_rough_plow	1394	100	1294
5	Fallow_smooth	2678	100	2578
6	Stubble	3959	100	3859
7	Celery	3579	100	3479
8	Grapes_untrained	11271	100	11171
9	Soil_vinyard_develop	6203	100	6103
10	Corn_senesced_green_weeds	3278	100	3178

11	Lettuce_romaine_4wk	1068	100	968
12	Lettuce_romaine_5wk	1927	100	1827
13	Lettuce_romaine_6wk	916	100	816
14	Lettuce_romaine_7wk	1070	100	970
15	Vinyard_untrained	7268	100	7168
16	Vinyard_vertical_trellis	1807	100	1707

Table 4. KSC dataset description

NO.	Class	Numbers of samples	Training	Testing
1	Scrub	761	20	1909
2	Willow swamp	243	20	3626
3	CP hammock	256	20	1876
4	CP/Oak	252	20	1294
5	Slash pine	161	20	2578
6	Oak/Broadleaf	229	20	3859
7	Hardwood swamp	105	20	3479
8	Graminoid marsh	431	20	11171
9	Spartina marsh	520	20	6103
10	Catail marsh	404	20	3178
11	Salt marsh	419	20	968
12	Mud flats	503	20	1827
13	Water	927	20	816

5.2. Comparative Experiment

To verify the performance of the DPBLS on hyperspectral images classification, we choose support vector machine with composite kernels (SVMCK) [22], extreme learning machine with composite kernels (ELMCK) [24], kernel extreme learning machine with composite kernels (KELMCK) [24], Adaptive scalable kernel (ASK) [19], PCA based EPFs (PCA-EPF) [21], Gabor Filtering Based Deep Network (GFDN) [45], Label-Consistent Transform Learning (LCTL) [20], BLS [26] and other algorithms as comparison. In the experiments, both SVMCK and KELMCK take Gaussian RBF kernel as the kernel function, while ELMCK uses it as the activation function. The penalty factor of SVMCK and the regularization coefficient of ELMCK and KELMCK are selected from $\{10^0, \dots, 10^5\}$, and the parameters of Gaussian RBF kernel are chosen from

$\{2^{-4}, 2^{-3}, \dots, 2^3, 2^4\}$. For SVMCK, ELMCK and KELMCK, we leverage triple cross-validation and grid search to select the optimal parameter values. For the parameters of ASK, PCA-EPF, GFDN, and LCTL algorithms, we refer to the corresponding literature. For fairness of experiments, we use the spectral-spatial joint feature of information fusion as the input of BLS. For the BLS algorithm and the DPBLS algorithm in this paper, we use triple cross-validation and grid search method for parameter optimization. The regularization parameter λ of the BLS algorithm is selected from $\{10^{-5}, 10^{-4}, \dots, 10^{-1}, 10^0\}$. The regularization parameters λ_1 and λ_2 of DPBLS proposed in this

paper are from $\{10^{-5}, 10^{-4}, \dots, 10^{-1}, 10^0\}$. We use overall accuracy (OA), average accuracy (AA), and

Kappa coefficient as evaluation metrics, where OA represents the ratio of the number of correctly categorized pixels to the total number of pixels, AA represents the mean value of the sum of all category accuracies, and Kappa coefficient represents the ratio of error reduction produced by categorization versus completely random categorization. The experimental results of different algorithms on the Indian Pines, Pavia of University, Salinas and KSC datasets are given in Table 5-8. The classification results of different algorithms on the four datasets are shown in Figure 6-9.

Table 5. Classification results of different algorithms on the Indian Pines dataset (%)

NO	SVMCK	ELMCK	KELMCK	ASK	LCTL	PCA-EPF	GFDN	BLS	DPBLS
1	100.00	100.00	100.00	100.00	100.00	100.00	99.57	97.39	98.70
2	93.18	94.10	96.02	97.51	92.28	93.06	96.87	97.22	97.17
3	96.11	95.81	97.84	99.02	96.05	98.47	98.88	98.23	98.81
4	99.34	100.00	99.93	94.81	99.05	94.06	100.00	99.42	100.00
5	98.22	98.75	99.56	98.37	99.85	99.07	98.38	98.96	99.97
6	99.44	99.78	99.81	99.22	97.81	99.95	99.86	99.89	99.98
7	98.57	100.00	100.00	97.57	99.86	100.00	95.71	100.00	100.00
8	99.95	100.00	100.00	99.48	99.81	100.00	99.97	99.92	100.00
9	100.00	100.00	100.00	96.03	100.00	100.00	100.00	100.00	100.00
10	93.38	96.69	97.32	92.96	95.95	96.78	97.99	98.76	96.51
11	90.70	87.79	92.62	98.92	94.95	98.13	97.08	93.44	97.15
12	96.43	96.23	97.71	98.25	98.44	97.92	98.32	97.77	98.95
13	99.81	100.00	100.00	99.72	99.81	100.00	99.71	100.00	100.00
14	98.20	99.74	99.74	99.35	99.98	99.97	99.52	97.83	99.90
15	99.51	99.93	100.00	98.65	99.13	98.01	99.58	97.76	99.30
16	99.78	100.00	99.57	98.49	98.26	98.49	99.57	86.30	99.13
OA	95.00	94.94	96.84	98.04	96.57	97.63	98.22	96.98	98.29
AA	97.66	98.05	98.76	98.02	98.26	98.37	98.81	97.68	99.10
Kappa	94.23	94.17	96.35	97.73	96.03	97.26	97.94	96.51	98.02
Time(s)	149.15	9.80	199.95	74.78	16.56	18.69	225.19	9.73	11.50

Table 6. Classification results of different algorithms on the University of Pavia dataset (%)

NO	SVMCK	ELMCK	KELMCK	ASK	LCTL	PCA-EPF	GFDN	BLS	DPBLS
1	94.08	86.74	92.62	89.26	95.94	99.07	96.94	97.45	97.79
2	96.67	97.66	98.49	99.00	97.97	99.84	96.88	99.03	99.46
3	91.46	92.81	92.41	94.43	98.56	98.30	99.55	99.58	99.63
4	96.37	95.23	98.11	67.16	94.41	96.82	96.91	97.19	96.96
5	100.00	100.00	100.00	99.33	97.04	94.73	99.96	99.78	99.89
6	96.82	98.26	98.21	99.35	99.80	100.00	98.82	99.99	99.97
7	98.95	100.00	99.96	89.75	98.72	99.83	99.89	99.98	100.00
8	90.95	86.97	93.77	92.05	95.26	96.79	98.45	98.43	98.48
9	97.65	93.87	96.99	79.29	94.83	98.18	99.81	98.37	99.16
OA	95.71	94.77	96.88	93.35	97.34	98.98	97.62	98.78	99.03
AA	95.89	94.62	96.73	89.96	96.95	98.17	98.58	98.87	99.04
Kappa	94.31	93.06	95.85	91.20	96.47	98.64	96.86	98.37	98.70

Time(s)	75.13	22.92	87.31	497.03	109.45	24.86	203.80	31.31	33.03
---------	-------	-------	-------	--------	--------	-------	--------	-------	-------

Table 7. Classification results of different algorithms on Salinas dataset (%)

NO	SVMCK	ELMCK	KELMCK	ASK	LCTL	PCA-EPF	GFDN	BLS	DPBLS
1	99.76	100.00	99.99	100.00	100.00	100.00	99.93	100.00	100.00
2	99.76	99.96	99.92	100.00	99.95	99.99	99.96	99.99	99.99
3	99.71	99.99	100.00	98.81	100.00	99.91	99.96	99.99	100.00
4	97.89	99.04	99.41	95.53	100.00	97.11	99.86	99.95	99.98
5	99.51	99.45	99.48	99.76	99.02	100.00	99.26	97.73	98.99
6	99.54	100.00	99.97	99.94	99.90	99.98	99.90	99.82	99.87
7	99.33	99.96	99.41	99.91	99.79	99.82	99.64	99.97	99.98
8	91.60	89.89	93.19	99.87	98.35	99.87	93.94	94.73	99.67
9	99.67	100.00	99.97	99.99	100.00	99.99	99.77	99.99	100.00
10	98.24	98.96	99.40	99.99	99.66	99.82	99.00	99.14	99.70
11	99.88	100.00	99.99	100.00	100.00	99.83	99.43	99.46	99.96
12	99.97	99.91	100.00	99.93	99.12	99.97	99.76	97.12	99.95
13	99.71	100.00	99.96	100.00	98.64	99.95	99.71	97.98	99.36
14	99.68	99.62	99.76	98.22	99.15	98.47	99.21	98.33	99.51
15	93.12	87.63	97.15	98.41	99.61	97.18	97.26	99.01	99.67
16	99.60	99.54	99.81	100.00	100.00	100.00	99.98	100.00	100.00
OA	96.91	96.02	98.02	99.57	99.43	99.41	98.13	98.39	99.78
AA	98.56	98.37	99.21	97.46	99.57	99.49	99.16	98.95	99.79
Kappa	96.56	95.55	97.80	99.52	99.37	99.34	97.92	98.20	99.76
Time(s)	212.16	41.56	332.40	406.07	122.96	30.77	263.64	29.60	30.92

Table 8. Classification results of different algorithms on KSC dataset (%)

NO	SVMCK	ELMCK	KELMCK	ASK	LCTL	PCA-EPF	GFDN	BLS	DPBLS
1	98.80	99.50	97.84	99.26	96.91	99.95	95.16	95.41	98.04
2	86.55	78.03	84.84	99.65	95.70	99.42	84.95	95.96	99.10
3	93.98	96.69	94.36	98.59	99.83	100.00	87.92	100.00	100.00
4	79.31	59.61	71.72	99.87	98.75	97.31	94.84	96.34	97.46
5	76.38	79.08	79.86	95.25	92.70	96.93	95.68	93.97	97.52
6	89.62	92.34	92.44	91.08	99.86	87.80	97.42	100.00	100.00
7	95.88	100.00	97.76	100.00	100.00	99.77	98.61	100.00	100.00
8	90.15	95.67	90.34	98.56	100.00	99.78	84.46	100.00	100.00
9	99.12	98.88	98.90	100.00	100.00	99.82	91.85	98.12	100.00
10	92.97	84.69	93.62	100.00	96.80	99.37	95.27	98.70	99.51
11	94.66	95.62	95.76	95.38	100.00	98.66	93.23	100.00	100.00
12	89.54	60.43	82.17	100.00	80.25	97.05	93.12	93.00	97.83
13	94.94	96.23	98.85	100.00	99.98	100.00	90.50	100.00	100.00
OA	92.60	89.37	92.77	98.62	96.88	98.59	92.12	97.83	99.23
AA	90.64	87.42	90.65	98.28	96.98	98.14	92.69	97.81	99.19

Kappa	91.73	88.16	91.94	98.46	96.52	98.43	91.32	97.58	99.14
Time(s)	12.92	3.27	13.94	488.42	67.25	25.14	173.52	27.50	28.50

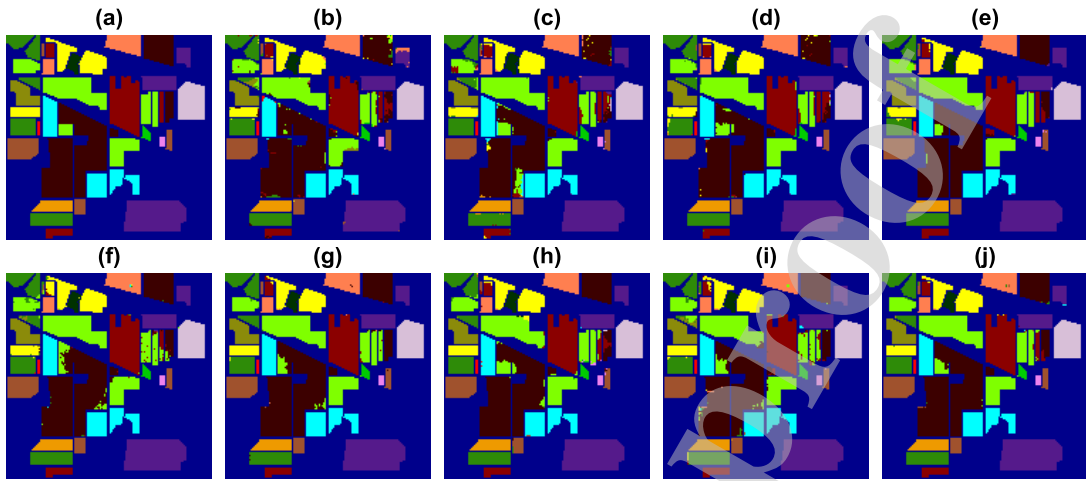


Figure 6. Classification results of different algorithms on the Indian Pines dataset.(a) Ground truth map(b) SVMCK, (c) ELMCK, (d) KELMCK,(e) ASK, (f) LCTL, (g) PCA-EPF, (h) GFDN, (i) BLS, (j) DPBLS.

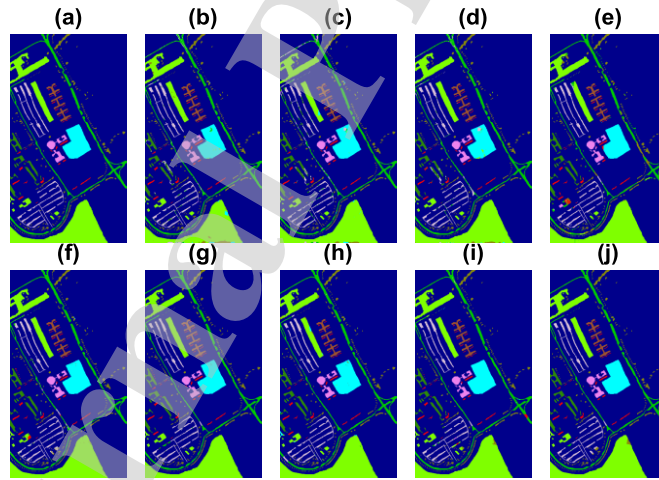


Figure 7. Classification results of different algorithms on the University of Pavia dataset. (a) Ground truth map(b) SVMCK, (c) ELMCK, (d) KELMCK, (e) ASK, (f) LCTL, (g) PCA-EPF, (h) GFDN, (i) BLS, (j) DPBLS.

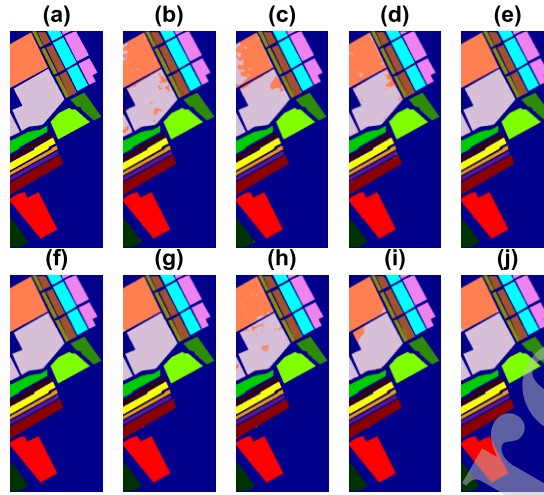


Figure 8. Classification results of different algorithms on the Salinas dataset. (a) Ground truth map(b) SVMCK, (c) ELMCK, (d) KELMCK, (e) ASK, (f) LCTL, (g) PCA-EPF, (h) GFDN, (i) BLS, (j) DPBLS.

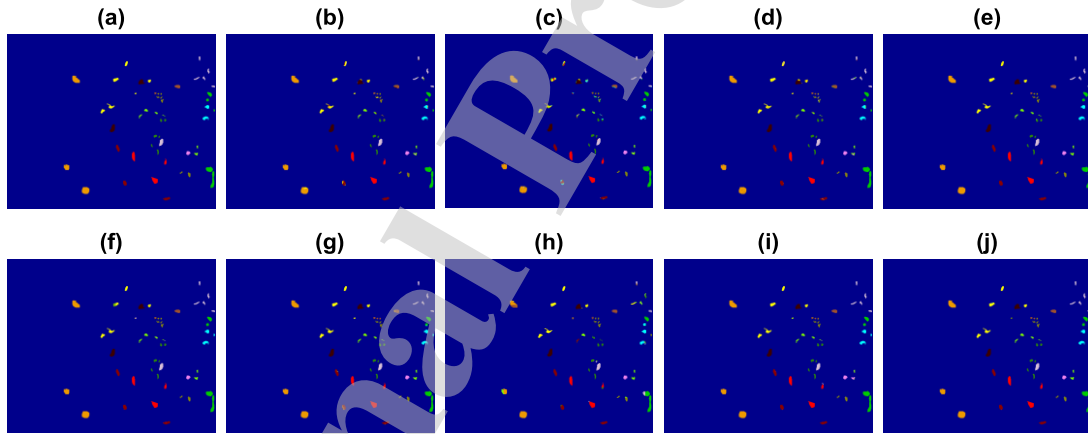


Figure 9. Classification results of different algorithms on the KSC dataset. (a) Ground truth map(b) SVMCK, (c) ELMCK, (d) KELMCK, (e) ASK, (f) LCTL, (g) PCA-EPF, (h) GFDN, (i) BLS, (j) DPBLS.

In order to verify the effectiveness of the algorithm proposed in this paper, we performed experiments on Indiana Pines, University of Pavia and Salinas respectively, and SVMCK, ELMCK, KELMCK, ASK, PCA-EPF, GFDN, LCTL, BLS are used for comparison experiments. Table 5 shows the classification results of different algorithms on Indiana Pines data, and Figure 6 shows the classification results of different algorithms. From Table 5, we can see that the accuracy of the algorithm in this paper reaches 100% in the categories of 'Corn', 'Grass-pasture-mowed', 'Hay-windrowed', 'Oats', 'Soybean-clean'. The accuracy of the proposed algorithm DPBLS on the three indicators of OA, AA and Kappa is better than other algorithms. The bold type in the table indicates the best results. From table 5, we can observe that the accuracy

of the algorithm in this paper is better than other algorithms in multiple categories. SVMCK, ELMCK, and KELMCK combine the spatial information and spectral information of remote sensing images through a combination kernel method for classification. From the classification result graph in Figure 5, we can see that the three algorithms SVMCK, ELMCK, and KELMCK have achieved good classification results. ASK applies the adaptive scalable kernel to hyperspectral image classification, it can achieve an excellent status in removing insignificant details and defending crucial features. From the classification results in table 5, we can see that the ASK algorithm has also achieved good classification results. The PCA-EPF algorithm uses PCA and edge-preserving filtering methods to fuse spatial spectrum joint features, and then uses SVM to classify. From the classification results in table 5, we can see that PCA-EPF has achieved good classification results. In the comparison experiment, we also conducted a comparison experiment with the deep learning methods GFDN. From the results in table 5, we can observe that the accuracy of the classification results of the algorithm in this paper is better than other algorithms on the three indicators OA, AA, and Kappa better than GFDN.

Table 6 shows the classification results of different algorithms on the University of Pavia data. From table 6, we observe that the OA of this algorithm is 99.03%, AA is 99.04%, and Kappa coefficient is 98.70%. The results of the three indicators of the algorithm in this paper are better than other algorithms. The accuracy rate in multiple categories such as 'Gravel', 'Trees', 'Bitumen', 'Bricks' is better than other algorithms. 100% accuracy in the two categories of 'Bitumen'. Figure 7 shows the classification results of different algorithms on the University of Pavia.

Table 7 shows the classification results of different algorithms on Salinas data. From table 7, we can see that the classification results of the algorithm in this paper on OA, AA, Kappa are significantly better than other algorithms. 100% accuracy in four categories: 'Brocoli_green_weeds_1', 'Fallow', 'Soil_vinyard_develop', and 'Vinyard_vertical_trellis'. Figure 8 shows the classification results of different algorithms on Salinas data. From Figure 8 we observe that our algorithm has achieved good results.

Table 8 gives the classification results of the different algorithms on the KSC dataset, and from Table 8 we can observe that the results of our model outperformed those of the other algorithms on the three metrics, namely, OA, AA, Kappa. Figure 9 demonstrates the classification results of different algorithms.

In order to verify the computational efficiency of our algorithm, the running times of different algorithms are given in Tables 5-8. Taking both the classification results and running times into consideration, the proposed algorithm can be used as an effective method on hyperspectral image classification.

Through the above analysis, our algorithm has achieved good results in hyperspectral image classification tasks, which verifies the effectiveness of our algorithm. The reason for the effectiveness of this algorithm is that the manifold structure information and discriminant information of the training samples are used to provide prior knowledge for the classification of hyperspectral images and enhance the classification performance of hyperspectral images.

5.3. Statistical Analysis

We perform significance test on the three indicators of AA, OA, and Kappa for the experimental results in Table 5-8, and draw a boxplot of the three indicators of AA, OA, and Kappa. Significance test aims to tell whether there is a difference between the results of two different

methods and whether the difference is significant. In the experiment, significance level is set as 0.05. When the estimated p-value is lower than 0.05, the performance difference between the two compared methods is statistically significant. Box plots on the Indiana Pines, University of Pavia, Salinas and KSC dataset are shown in Figure 10-13. Significance test results on the Indiana Pines, Pavia of University, and Salinas data are given in table 9-12.

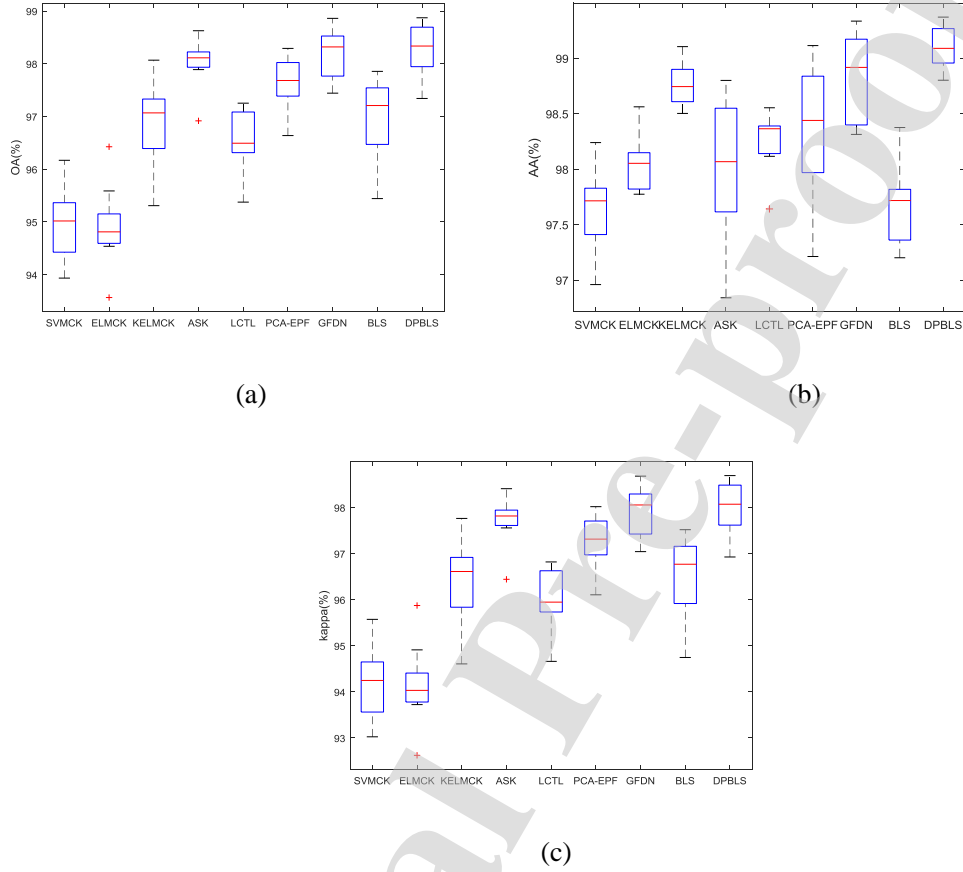
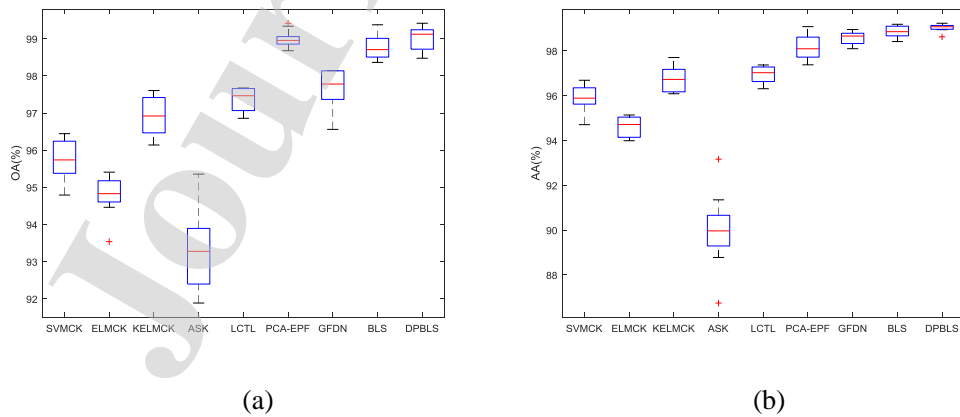
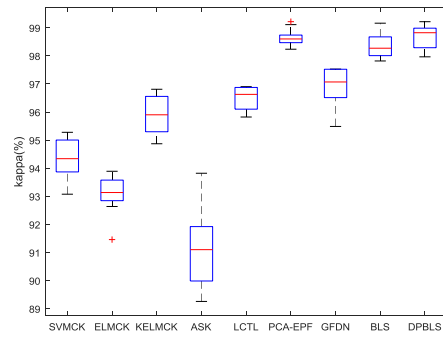


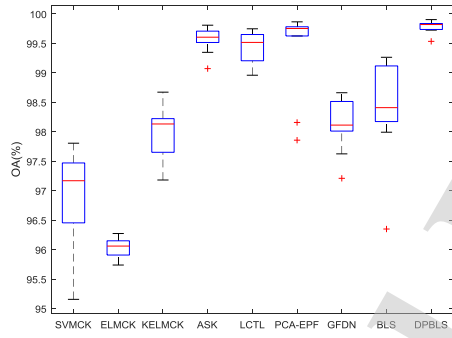
Figure 10. Boxplots of the OA, AA and Kappa (%) of SVMCK, ELMCK, KELMCK, ASK, LCTL, PCA-EPF, GFDN, BLS, DPBLS on the Indiana Pines dataset.



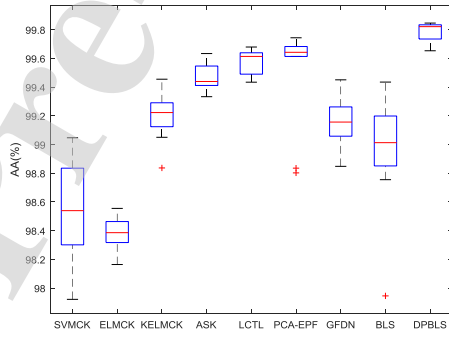


(c)

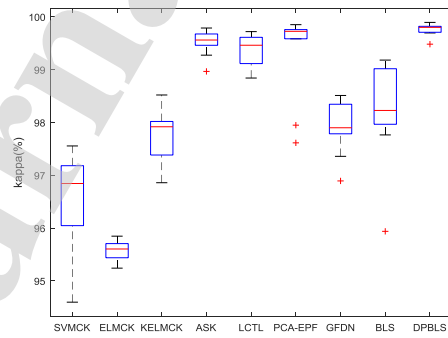
Figure 11. Boxplots of the OA,AA and Kappa (%) of SVMCK, ELMCK, KELMCK, ASK, LCTL, PCA-EPF, GFDN, BLS, DPBLS on the Univerisy of Pavia dataset.



(a)



(b)



(c)

Figure 12. Boxplots of the OA,AA and Kappa (%) of SVMCK, ELMCK, KELMCK, ASK, LCTL, PCA-EPF, GFDN, BLS, DPBLS on the Salinas dataset.

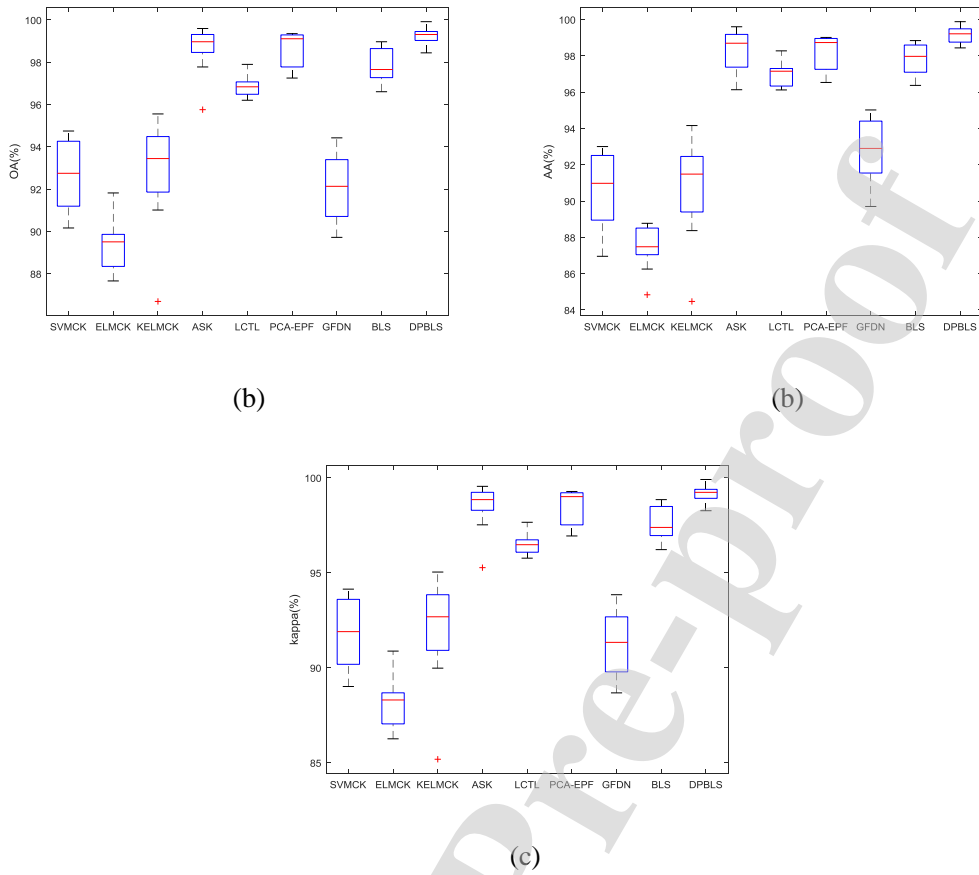


Figure 13. Boxplots of the OA, AA and Kappa (%) of SVMCK, ELMCK, KELMCK, ASK, LCTL, PCA-EPF, GFDN, BLS, DPBLS on the KSC dataset.

Table 9. DPBLS and eight other algorithms' p-values on the Indiana Pines dataset. (Bold indicates that at $p = 0.05$, the algorithm in this paper has statistically significant).

Methods	OA	AA	Kappa
SVMCK	3.4×10^{-7}	2.3×10^{-7}	3.5×10^{-7}
ELMCK	2.4×10^{-7}	2.0×10^{-6}	2.3×10^{-7}
KELMCK	1.0×10^{-2}	2.1×10^{-2}	1.1×10^{-2}
ASK	2.2×10^{-1}	7.0×10^{-4}	2.2×10^{-1}
LCTL	1.7×10^{-21}	1.2×10^{-24}	8.4×10^{-5}
PCA-EPF	3.6×10^{-2}	8.2×10^{-4}	3.5×10^{-2}
GFDN	7.3×10^{-1}	1.1×10^{-1}	7.4×10^{-1}
BLS	4.8×10^{-2}	2.4×10^{-6}	4.8×10^{-2}

Table 10. DPBLS and eight other algorithms' p-values on the University of Pavia dataset. (Bold indicates that at $p = 0.05$, the algorithm in this paper has statistically significant).

Methods	OA	AA	Kappa
SVMCK	3.6×10^{-9}	1.6×10^{-8}	3.5×10^{-9}
ELMCK	7.0×10^{-9}	6.3×10^{-10}	6.3×10^{-9}
KELMCK	2.6×10^{-6}	8.3×10^{-7}	2.5×10^{-6}
ASK	1.8×10^{-7}	3.9×10^{-8}	1.7×10^{-7}

LCTL	7.2×10^{-24}	2.9×10^{-23}	2.3×10^{-7}
PCA-EPF	6.7×10^{-1}	2.8×10^{-4}	6.7×10^{-1}
GFDN	3.8×10^{-4}	4.5×10^{-2}	3.7×10^{-4}
BLS	7.3×10^{-2}	1.2×10^{-1}	7.4×10^{-2}

Table 11. DPBLS and eight other algorithms' p-values on the Salinas dataset. (Bold indicates that at $p = 0.05$, the algorithm in this paper has statistically significant).

Methods	OA	AA	Kappa
SVMCK	1.1×10^{-6}	2.6×10^{-6}	1.1×10^{-6}
ELMCK	2.0×10^{-12}	1.3×10^{-10}	2.1×10^{-12}
KELMCK	3.2×10^{-6}	2.1×10^{-5}	3.1×10^{-6}
ASK	2.0×10^{-2}	1.2×10^{-6}	2.0×10^{-2}
LCTL	1.53×10^{-24}	8.3×10^{-29}	1.0×10^{-3}
PCA-EPF	1.6×10^{-1}	3.7×10^{-2}	1.6×10^{-1}
GFDN	7.0×10^{-7}	1.6×10^{-6}	6.9×10^{-7}
BLS	8.1×10^{-4}	1.5×10^{-4}	8.0×10^{-4}

Table 12. DPBLS and eight other algorithms' p-values on the KSC dataset. (Bold indicates that at $p = 0.05$, the algorithm in this paper has statistically significant).

Methods	OA	AA	Kappa
SVMCK	1.3×10^{-7}	3.2×10^{-7}	1.3×10^{-7}
ELMCK	1.2×10^{-10}	2.6×10^{-10}	1.2×10^{-10}
KELMCK	4.6×10^{-5}	1.2×10^{-10}	4.5×10^{-5}
ASK	1.2×10^{-1}	5.9×10^{-2}	1.1×10^{-1}
LCTL	4.2×10^{-22}	5.5×10^{-21}	3.5×10^{-7}
PCA-EPF	7.0×10^{-2}	1.3×10^{-2}	7.1×10^{-2}
GFDN	2.2×10^{-7}	1.2×10^{-6}	2.2×10^{-7}
BLS	3.0×10^{-4}	3.6×10^{-4}	3.1×10^{-4}

Figure 10-13 shows the box plot results of different algorithms. In the box plot, the edges of the box are 25% and 75%, respectively. Lines that extend vertically from the box represent variability beyond the upper and lower quartiles. Outliers are displayed in red "+". Table 9-12 shows the significance test results of OA, AA, and Kappa on the Indiana Pines, University of Pavia, and Salinas datasets. P-values was set to 0.05 when we performed significance test. When the p-values of the two comparison algorithms are less than 0.05, we consider that the difference between the two algorithms is significantly different. From table 9, we can observe that the significance test results of this algorithm and SVMCK, ELMCK, KELMCK, ASK, PCA-EPF, LCTL, BLS and other algorithms are in significant differences. From Table 11 and Table 12, we could also make a conclusion that compared with other algorithms, the DPBLS algorithm still has significant differences in most cases. The significance test results in Table 9-12 verify the effectiveness and efficiency of our algorithm in hyperspectral image classification.

5.4. Classification results under different training samples

For illustrating how different number of samples effects classification abilities of SVMCK, ELMCK, KELMCK, ASK, LCTL, PCA-EPF, GFDN, BLS and DPBLS, we conducted experiments on the Indiana Pines dataset, University of Pavia dataset, Salinas dataset and KSC datasets respectively. For the Indiana Pines dataset, we randomly selected 6%, 7%, 8%, 9%, and 10% of samples in each category as training samples, and the remaining of each category are for testing. On the other three datasets, these numbers are set to 1%, 2%, 3%, 4%, and 5%. As for the evaluation, we use the overall classification accuracy and Kappa coefficient as metrics. The experimental results are shown in Figures 14-17.

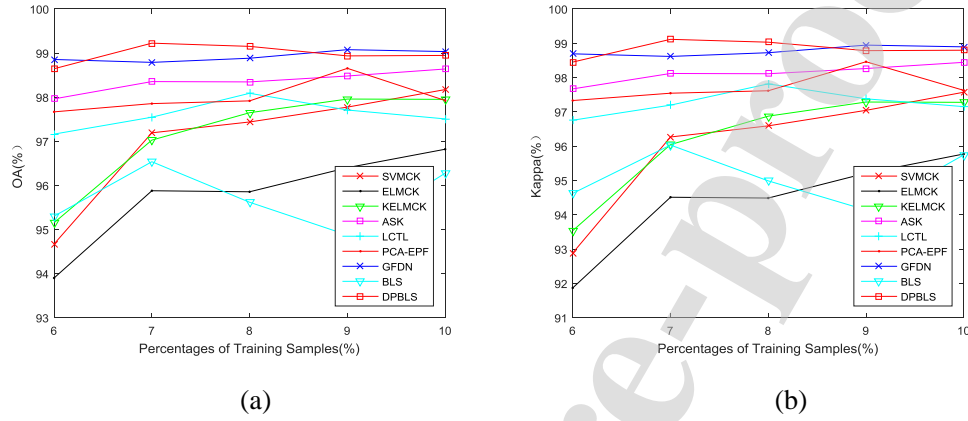


Figure 14. Classification accuracies with different number of training samples. OA and Kappa on the Indian Pines dataset.

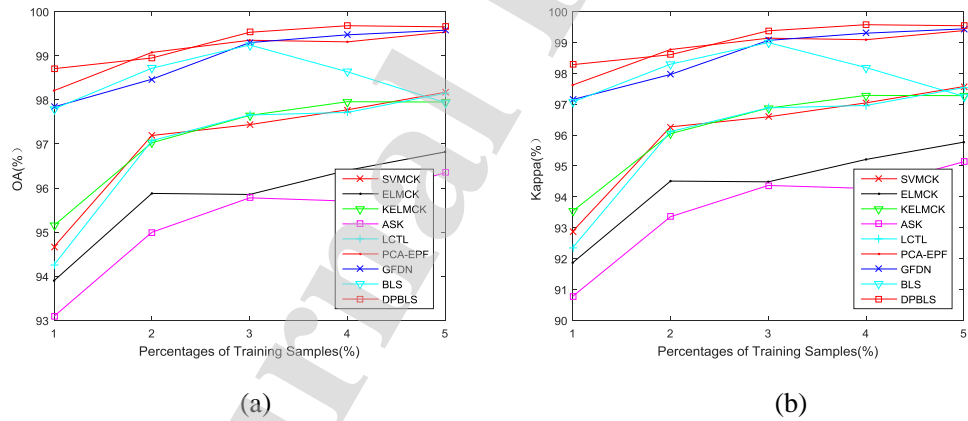


Figure 15. Classification accuracies with different number of training samples. OA and Kappa on the University of Pavia dataset.

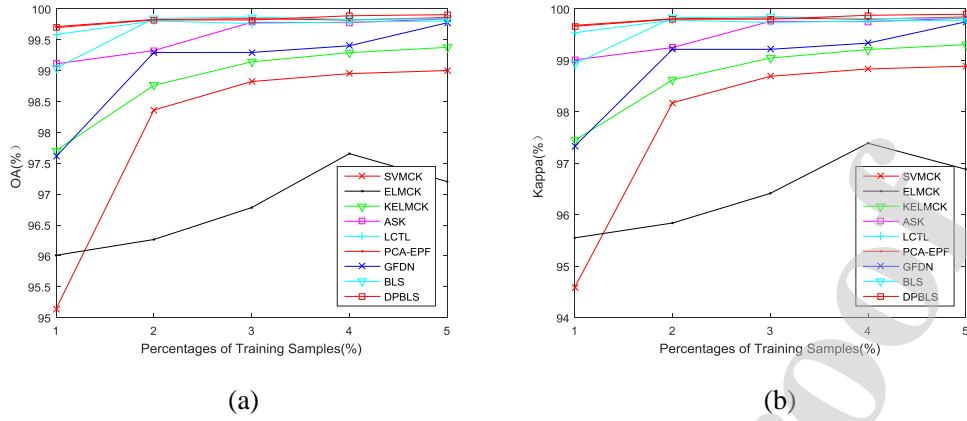


Figure 16. Classification accuracies with different number of training samples. OA and Kappa on the Salinas dataset.

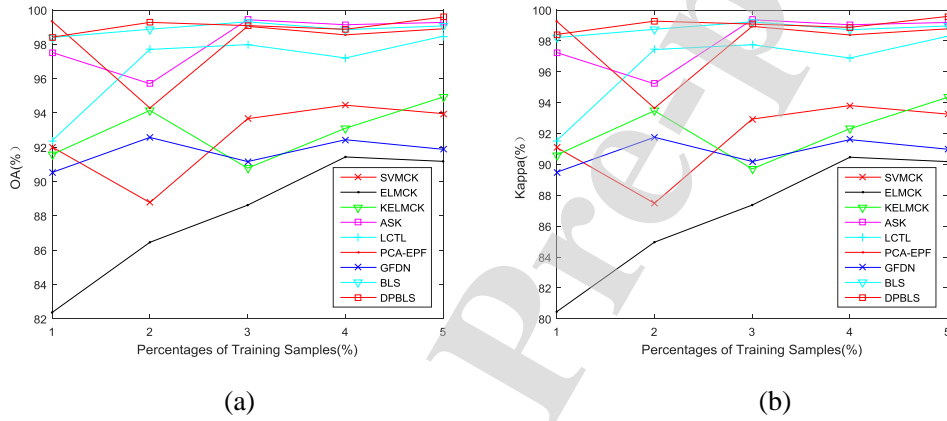


Figure 17. Classification accuracies with different number of training samples. OA and Kappa on the KSC dataset.

From the experimental results in Figure 14-17, we could observe that under different numbers of training samples, the results (OA and Kappa) of the proposed algorithms in this paper are better than other algorithms in most cases. This also verifies that in case of limited hyperspectral image training samples, using both the manifold structure information and discriminant information of data samples as prior information in the DPBLS could improve the discriminative ability of output weights in the BLS.

5.5. Ablation study

In order to further verify the performance of the proposed algorithm, we analyze the results of both BLS and DPBLS algorithms in cases of with and without feature fusion. For Indian Pines, we randomly selected 10% of the total sample in each category as the training set, with the remaining sample as a test set. For the University of Pavia dataset, we randomly selected 1% of the total sample in each category as the training set and the remaining samples as the test set. For the Salinas dataset, we randomly select 0.5% of the total sample in each category as the training set, and the remaining sample is used as the test set. For the KSC dataset, we randomly select 1% of the total sample in each category as the training set, and the remaining sample of the as a test set.

The experiments were repeated ten times for each algorithm, and then these results were averaged to obtain the final results, which are given in Tables 13-16.

Table 13. Classification results of different algorithms on the Indian Pines dataset (%)

	Spectral classifiers		Spectral-spatial classifiers	
	BLS	DPBLS	BLS	DPBLS
OA	76.52	80.79	98.00	99.14
AA	76.65	80.31	98.29	99.08
Kappa	72.77	77.88	97.71	99.02

Table 14. Classification results of different algorithms on the University of Pavia (%)

	Spectral classifiers		Spectral-spatial classifiers	
	BLS	DPBLS	BLS	DPBLS
OA	75.86	80.92	95.60	98.83
AA	59.11	69.39	93.09	97.75
Kappa	67.43	74.15	94.14	98.45

Table 15. Classification results of different algorithms on the Salinas (%)

	Spectral classifiers		Spectral-spatial classifiers	
	BLS	DPBLS	BLS	DPBLS
OA	83.35	87.97	99.44	99.84
AA	88.27	91.78	99.33	99.68
Kappa	81.50	86.58	99.38	99.83

Table 16. Classification results of different algorithms on the KSC (%)

	Spectral classifiers		Spectral-spatial classifiers	
	BLS	DPBLS	BLS	DPBLS
OA	78.30	82.97	91.73	99.10
AA	73.53	77.52	91.75	99.16
Kappa	75.87	81.00	90.79	99.00

Tables 13-16 show the experimental results for both the BLS and DPBLS algorithms on the four datasets, which are divided into two conditions: one is in case of using spectral-spatial features, while the other is in case of using only spectral features. From tables 13-16 we see that the experimental results of both the BLS and DPBLS algorithms in the first case are significantly better than those in the latter case. We can also observe that the experimental results of the proposed algorithm are significantly better than those of BLS, further proving that the validity of our algorithm.

5.6. Parameter sensitivity and selection

In this part, we discuss the sensitivity of the regularization parameters λ_1 and λ_2 to the DPBLS algorithm under different values. We set the number of feature nodes of DPBLS to 100, the size of the feature node window to 10, and the number of enhancement nodes to 1000. The regularization parameters λ_1 and λ_2 are $\{10^{-10}, 10^{-9}, \dots, 10^{-1}, 10^0\}$, respectively. As for the four datasets, we

randomly select 100 samples in each category for model training, while the remaining in each category are for testing. Figure 15 demonstrates the classification results of the overall accuracy of the regularization parameters λ_1 and λ_2 on three hyperspectral image datasets with different values. From Figure 15 we could find that the overall accuracy rate shows an increase with the growth of the regularization parameter λ_1 on the Indian Pines dataset. While on the Pavia University dataset, the overall accuracy of DPBLS is reduced by the increasing of regularization parameter λ_2 . These changes mean that the DPBLS is with a high stability on the Pavia University dataset.

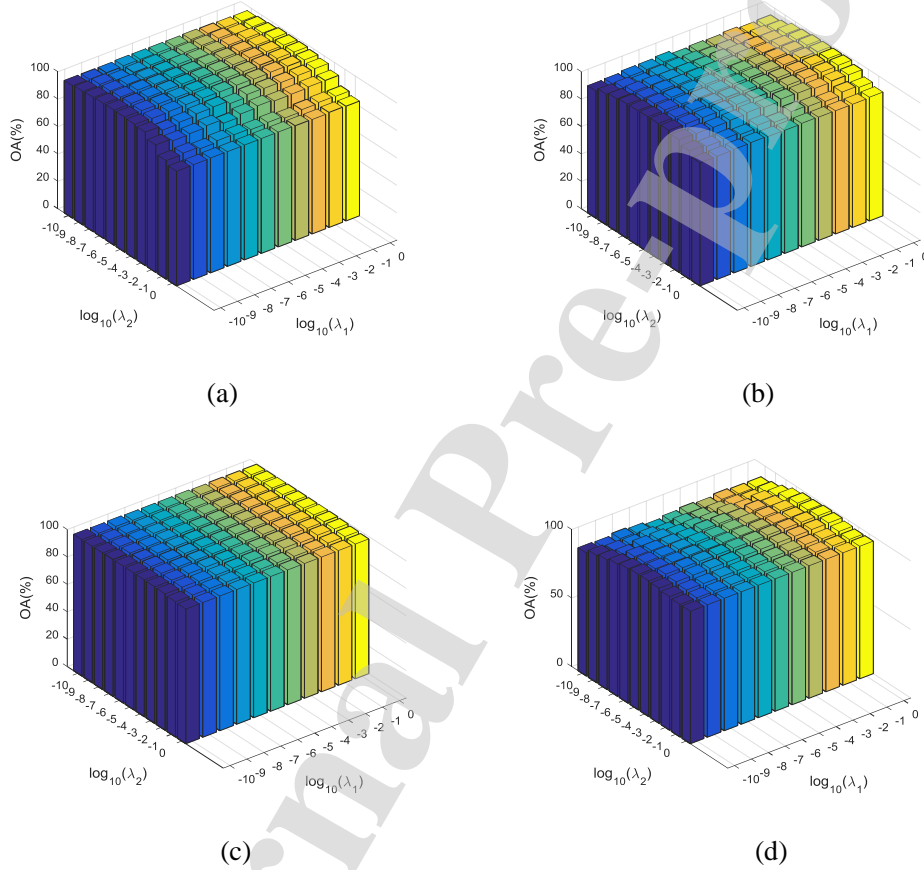


Figure 15. Effect of different regularization parameters on DPBLS on hyperspectral image dataset (%): (a) Indian Pines; (b) Pavia University; (c) Salinas; (d) KSC.

6. Conclusions

In this paper, we introduce broad learning system into hyperspectral image classification tasks based on advantages of BLS, proposing a hyperspectral image classification framework based on discriminative local broad learning and information fusion techniques. Due to the limited labeled samples in hyperspectral image classification, BLS learning is not sufficient. Our method introduces the discriminative information and local manifold structure of the data samples into the BLS model to enhance the classification ability of the BLS model in hyperspectral images. We perform experiments on four hyperspectral image datasets, and setting traditional machine learning algorithms and deep learning algorithms as comparisons. Traditional machine learning

algorithms include SVMCK, ELMCK, KELMCK and others, while for deep learning algorithms we choose GFDN. Experimental results demonstrate that our algorithm is superior to other algorithms on four hyperspectral image datasets.

Acknowledgment

This work is partially supported by grant from the Natural Science Foundation of China (No. 61632011, 61572102, 61702080, 61602079), the Fundamental Research Funds for the Central Universities (NO. DUT18ZD102, DUT17RC(3)016).

References

- [1] Lee M A, Huang Y, Yao H, et al. Determining the Effects of Storage on Cotton and Soybean Leaf Samples for Hyperspectral Analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(6): 2562-2570.
- [2] Hege E K, O'Connell D, Johnson W, et al. *Hyperspectral Imaging for Astronomy and Apace Surveillance*. International Society for Optics and Photonics, 2004, 5159: 380-391.
- [3] Campsvalls G, Tuia D, Bruzzone L, et al. Advances in Hyperspectral Image Classification: EartMonitoring with Statistical Learning Methods. *IEEE Signal Processing Magazine*, 2014, 31(1): 45-54.
- [4] Melgani F, Bruzzone L. Classification of Hyperspectral Remote Sensing Images With Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing*, 2004, 42(8): 1778-1790.
- [5] Blanzieri E, Melgani F. Nearest Neighbor Classification of Remote Sensing Images With the Maximal Margin Principle. *IEEE Transactions on Geoscience and Remote Sensing*, 2008, 46(6): 1804-1811.
- [6] Delalieux S, Somers B, Haest B, et al. Heathland Conservation Status Mapping Through Integration of Hyperspectral Mixture Analysis and Decision Tree Classifiers. *Remote Sensing of Environment*, 2012: 222-231.
- [7] Su H, Cai Y, Du Q, et al. Firefly-Algorithm-Inspired Framework With Band Selection and Extreme Learning Machine for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 10(1): 309-320.
- [8] Zhan K, Wang H, Xie Y, et al. Albedo Recovery for Hyperspectral Image Classification. *Journal of Electronic Imaging*, 2017, 26(4): 043010.
- [9] Cheng G, Li Z, Han J, et al. Exploring Hierarchical Convolutional Features for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(11): 6712-6722.
- [10] Zhou P, Han J, Cheng G, et al. Learning Compact and Discriminative Stacked Autoencoder for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(7): 4823-4833.
- [11] Barman B, Patra S. Variable Precision Rough Set Based Unsupervised Band Selection Technique for Hyperspectral Image Classification. *Knowledge-Based Systems*, 2020, 193: 105414.
- [12] Yang L, Yang S, Li S, et al. Coupled Compressed Sensing Inspired Sparse Spatial-spectral LSSVM for Hyperspectral Image Classification. *Knowledge Based Systems*, 2015: 80-89.
- [13] Li S, Wu H, Wan D, et al. An Effective Feature Selection Method for Hyperspectral Image

Classification Based on Genetic Algorithm and Support Vector Machine. *Knowledge Based Systems*, 2011, 24(1): 40-48.

[14] Fauvel M, Benediktsson J A, Chanussot J, et al. Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 2008, 46(11): 3804-3814.

[15] Moser G, Serpico S B. Combining Support Vector Machines and Markov Random Fields in an Integrated Framework for Contextual Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, 51(5): 2734-2752.

[16] Chen Y, Nasrabadi N M, Tran T D, et al. Hyperspectral Image Classification Using Dictionary-Based Sparse Representation. *IEEE Transactions on Geoscience and Remote Sensing*, 2011, 49(10): 3973-3985.

[17] Sun L, Wu Z, Liu J, et al. Supervised Spectral-Spatial Hyperspectral Image Classification With Weighted Markov Random Fields. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(3): 1490-1503.

[18] Benediktsson J A, Palmason J A, Sveinsson J R, et al. Classification of Hyperspectral Data From Urban Areas Based On Extended Morphological Profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 2005, 43(3): 480-491.

[19] Wang J, Liu B, He Y, et al. Adaptive Scalable Kernel For Hyperspectral Image Classification. *Journal of Electronic Imaging*, 2019, 28(1): 013051.

[20] Maggu J, Aggarwal H K, Majumdar A. Label-consistent Transform Learning For Hyperspectral Image Classification. *IEEE Geoscience and Remote Sensing Letters*, 2019, 16(9): 1502-1506.

[21] Kang X, Xiang X, Li S, et al. PCA-Based Edge-Preserving Features for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(12): 7140-7151.

[22] Campsvals G, Gomezchova L, Munozmari J, et al. Composite Kernels for Hyperspectral Image Classification. *IEEE Geoscience and Remote Sensing Letters*, 2006, 3(1): 93-97.

[23] Marconcini M, Campsvals G, Bruzzone L, et al. A Composite Semisupervised SVM for Classification of Hyperspectral Images. *IEEE Geoscience and Remote Sensing Letters*, 2009, 6(2): 234-238.

[24] Zhou Y, Peng J, Chen C L, et al. Extreme Learning Machine With Composite Kernels for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(6): 2351-2360.

[25] Liu J, Wu Z, Wei Z, et al. Spectral-spatial Kernel Sparse Representation for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2013, 6(6): 2462-2471.

[26] Chen C L, Liu Z. Broad Learning System: An Effective and Efficient Incremental Learning System Without the Need for Deep Architecture. *IEEE transactions on neural networks and learning systems*, 2018, 29(1): 10-24.

[27] Han M, Feng S, Chen C L, et al. Structured Manifold Broad Learning System: A Manifold Perspective for Large-Scale Chaotic Time Series Analysis and Prediction. *IEEE Transactions on Knowledge and Data Engineering*, 2019, 31(9): 1809-1821.

[28] Feng S, Chen C L. Fuzzy Broad Learning System: A Novel Neuro-Fuzzy Model for Regression and Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 2020, 50(2): 414-424.

- [29] Jin J, Chen C L. Regularized robust Broad Learning System for Uncertain Data Modeling. *Neurocomputing*, 2018: 58-69.
- [30] Kong Y, Wang X, Cheng Y, et al. Hyperspectral Imagery Classification Based on Semi-Supervised Broad Learning System. *Remote Sensing*, 2018, 10(5).
- [31] Jin J, Liu Z, Chen C L, et al. Discriminative Graph Regularized Broad Learning System for Image Recognition. *Science in China Series F: Information Sciences*, 2018, 61(11).
- [32] Zhang T, Chen R, Yang X, et al. Rich Feature Combination for Cost-Based Broad Learning System. *IEEE Access*, 2019: 160-172.
- [33] Fan J, Wang X, Wang X, et al. Incremental Wishart Broad Learning System for Fast PolSAR Image Classification. *IEEE Geoscience and Remote Sensing Letters*, 2019, 16(12): 1854-1858.
- [34] Guo H, Sheng B, Li P, et al. Multiview High Dynamic Range Image Synthesis Using Fuzzy Broad Learning System. *IEEE Transactions on Systems, Man, and Cybernetics*, 2019: 1-13.
- [35] Huang G, Song S, Gupta J N, et al. Semi-Supervised and Unsupervised Extreme Learning Machines. *IEEE Transactions on Systems, Man, and Cybernetics*, 2014, 44(12): 2405-2417.
- [36] Yang Y, Xu D, Nie F, et al. Image Clustering Using Local Discriminant Models and Global Integration. *IEEE Transactions on Image Processing*, 2010, 19(10): 2761-2773.
- [37] Yan S, Xu D, Zhang B, et al. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. *IEEE transactions on pattern analysis and machine intelligence*, 2006, 29(1): 40-51.
- [38] Yang Y, Shen H T, Ma Z, et al. L2, 1-norm Regularized Discriminative Feature Selection for Unsupervised[C]//Twenty-Second International Joint Conference on Artificial Intelligence. 2011.
- [39] Yang Y, Nie F, Xu D, et al. A Multimedia Retrieval Framework Based on Semi-supervised Ranking and Relevance Feedback. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 34(4): 723-742.
- [40] Sugiyama M. Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis. *Journal of Machine Learning Research*, 2007: 1027-1061.
- [41] Cao F, Guo W. Cascaded Dual-scale Crossover Network for Hyperspectral Image Classification. *Knowledge-Based Systems*, 2020, 189: 105122.
- [42] Gao H, Yang Y, Lei S, et al. Multi-branch Fusion Network for Hyperspectral Image Classification. *Knowledge-Based Systems*, 2019, 167: 11-25.
- [43] Hu W, Huang Y, Wei L, et al. Deep Convolutional Neural Networks for Hyperspectral Image Classification. *Journal of Sensors*, 2015, 2015(2015): 1-12.
- [44] Romero A, Gatta C, Campsvals G, et al. Unsupervised Deep Feature Extraction for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(3): 1349-1362.
- [45] Kang X, Li C, Li S, et al. Classification of Hyperspectral Images by Gabor Filtering Based Deep Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(4): 1166-1178.
- [46] Xu Y, Du B, Zhang F, et al. Hyperspectral Image Classification via a Random Patches Network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018: 344-357.
- [47] Chen Y, Jiang H, Li C, et al. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(10): 6232-6251.
- [48] Li W, Wu G, Zhang F, et al. Hyperspectral Image Classification Using Deep Pixel-Pair

Features. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(2): 844-853.

[49] Cao J, Chen Z, Wang B. Deep Convolutional Networks With Superpixel Segmentation for Hyperspectral Image Classification[C]//2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016: 3310-3313.

[50] Zhao W, Du S. Spectral-Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(8): 4544-4554.

[51] Mika S, Ratsch G, Weston J, et al. Fisher Discriminant Analysis With Kernels[C]//Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop.1999: 41-48.

[52] Li H, Jiang T, Zhang K, et al. Efficient and Robust Feature Extraction by Maximum Margin Criterion. IEEE transactions on neural networks and learning systems, 2006, 17(1): 157-165.

Credit Author Statement

We declare that we have no conflicts of interest.

Journal Pre-proof

Conflict of Interest Statement

We declare that we have no conflicts of interest.

Journal Pre-proof