



Article

Classification of Hyperspectral Images with Robust Regularized Block Low-Rank Discriminant Analysis

Baokai Zu ^{1,2} , Kewen Xia ^{1,*}, Wei Du ³, Yafang Li ⁴, Ahmad Ali ¹  and Sagnik Chakraborty ⁵

¹ School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China; zubaokai@163.com (B.Z.); aali_forever@hotmail.com (A.A.)

² Computer Science Department, Worcester Polytechnic Institute, Worcester, MA 01609, USA

³ College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China; weidu@webmail.hzau.edu.cn

⁴ Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China; yafangli@bjut.edu.cn

⁵ Post Doctoral Fellow, School of Energy and Environmental Engineering, Hebei University of Technology, Tianjin 300401, China; sagnik11@yahoo.com

* Correspondence: kwxia@hebut.edu.cn

Received: 22 April 2018; Accepted: 17 May 2018; Published: 24 May 2018



Abstract: Classification of Hyperspectral Images (HSIs) has gained attention for the past few decades. In remote sensing image classification, the labeled samples are insufficient or hard to obtain; however, the unlabeled ones are frequently rich and of a vast number. When there are no sufficient labeled samples, overfitting may occur. To resolve the overfitting issue, in this present work, we proposed a novel approach for HSI feature extraction, called robust regularized Block Low-Rank Discriminant Analysis (BLRDA), which is a robust and efficient feature extraction method to improve the HSIs' classification accuracy with few labeled samples. To reduce the exponentially growing computational complexity of the low-rank method, we divide the entire image into blocks and implement the low-rank representation for each block respectively. Due to the symmetric matrix requirements for the regularized graph of discriminant analysis, the k -nearest neighbor is applied to handle the whole low-rank graph integrally. The low-rank representation and the k NN can maximally capture and preserve the global and local geometry of the data, respectively, and the performance of regularized discriminant analysis feature extraction can be apparently improved. Extensive experiments on multi-class hyperspectral images show that the proposed BLRDA is a very robust and efficient feature extraction method. Even with simple supervised and semi-supervised classifiers (nearest neighbor and SVM) and randomly given parameters, the feature extraction method achieves significant results with few labeled samples, which shows better performance than similar feature extraction methods.

Keywords: hyperspectral image; regularized block low-rank discriminant analysis; semi-supervised discriminant analysis; low-rank representation; feature extraction

1. Introduction

With the advancement of remotely-sensed hyperspectral imaging instruments, Hyperspectral Images (HSIs) have gained widespread attention throughout the globe. HSI contains enriched information due to the presence of various bands (almost more than hundreds) [1]. HSI provides the comprehensive spectral information of the materials' physical properties. This ubiquitous technique is applied in agricultural monitoring [2,3], forestry [4], ecosystem monitoring [5], mineral identification [6,7], environmental pollution monitoring [8] and urban growth analysis [9,10]. For the

HSIs' classification, good results usually entail many labeled samples. The images are deemed as high-dimensional points that lie in or nearly in low-dimensional manifolds. Recently, numerous dimensionality reduction algorithms have proposed preserving the local manifold structure of the data, such as Locally Linear Embedding (LLE) [11], Isomap [12] and Laplacian Eigenmap [13]. The main limitation is the lack of sufficient labeled training data since identifying and labeling samples comprise a daunting task. The main drawback of the machine learning approach is overfitting [14,15] due to the limited number of training data when dealing with small sample problems in highly dimensional and nonlinear cases, which is referred to as the "Hughes phenomenon" [16,17].

To resolve the issue above, semi-supervised learning was proposed to utilize both labeled data and the information conveyed by the marginal distribution of the unlabeled samples to boost the algorithmic performance [18–20]. Graph embedding means that each node is mapped to a low-dimensional feature vector and tries to maintain the connection between vertices [20]. Cai et al. proposed Semi-supervised Discriminant Analysis (SDA), which is a novel method that takes advantage of labeled and unlabeled samples [21]. Nevertheless, the semi-supervised learning algorithmic performance relies profoundly on the graph construction process. The k -nearest neighbor [12] and Locally Linear Embedding (LLE) neighbors [13,22], as well as other traditional methods depend mainly on the pair of Euclidean distances. However, these methods are not robust to noise as a result of using Euclidean distances to find the pairwise weights. Yan et al. proposed the l_1 graph for SSL (Semi-supervised learning) [23]. Zhu et al. presented a series of novel semi-supervised learning approaches arising from graph representation [18,24,25]. Belkin and Niyogi [26] proposed a regression function that fits the label of the labeled data while maintaining the smoothness of the data. Jebara et al. provided a b -matching graph for SSL, which ensures that each node has the same number of edges in the balanced graph [22]. Zhou et al. [18] conducted semi-supervised learning with local and global consistency.

Low-rank representation is proposed to construct an undirected graph (LR-graph) [27–29], which jointly receives the graph of all of the data according to the low-rank constraint [27–29]. The graph regards as finite a group of samples in which one sample is associated with a vertex and the weight represents two connecting vertexes' similarity [20,30,31]. In [32], Zhang et al. introduced the HSI restoration Low-Rank-based Recovery Method (LRMR), which can remove various noises. Veganzones et al. revealed an approach that partitions the image into patches and resolves the fusion issue of each patch by low-rank representation individually [33]. According to LRR (Low-rank Representation) and the Learned Dictionary (LD), a hyperspectral anomaly detector has been put forward, which assumes that the hyperspectral image can disintegrate into a low-rank and a sparse matrix, respectively, representing the background and anomalies [34]. To resolve the failure to preserve spatial information problem, a Tensor Sparse and Low-rank Graph for Discriminant Analysis (TSLGDA) is proposed [35]. In [36], He et al. suggested a spatial-spectral mixed-noise removal method for the HSI. The proposed Sparse and Low-rank Graph-based Discriminant Analysis (SLGDA) incorporates sparsity and low-rank to simultaneously preserve both local and global information [30]. Qi et al. presented a multi-task joint sparse and low-rank representation model for high-resolution satellite image interpretation [37].

The calculation of the LRR graph is a grave concern; because the growing data points subsequently issue the pixels in HSIs, the computational complexity of LRR grows exponentially. To resolve the issues above, we explore the low-rank structure via block theory. Figure 1 shows the formulation of the proposed regularized block low-rank discriminant analysis feature extraction for HSI classification. As shown in Figure 1, we preprocess the hyperspectral image first by the Image Fusion and Recursive Filtering feature (IFRF), which removes the redundant information simultaneously [17]. The IFRF feature method is to select better bands, which eradicates the noise and redundant information concurrently. However, our goal is to improve the graph adjacency between the pixels for the regularization term of regularized discriminant analysis. Consequently, inspired by the LRR algorithm, we aim to exploit the low-rank structure. LR uses all samples as the dictionary, where each sample is represented as a linear combination of the dictionary. The low-rank hypothesis is a global constraint,

which guarantees that data points in the same subspace are clustered in the same class [29]. When all samples distribute into independent subspaces, the coefficients of the low-rank assumption reveal the membership of these samples: the within-cluster affinities are dense, and the between-cluster affinities are all zeros.

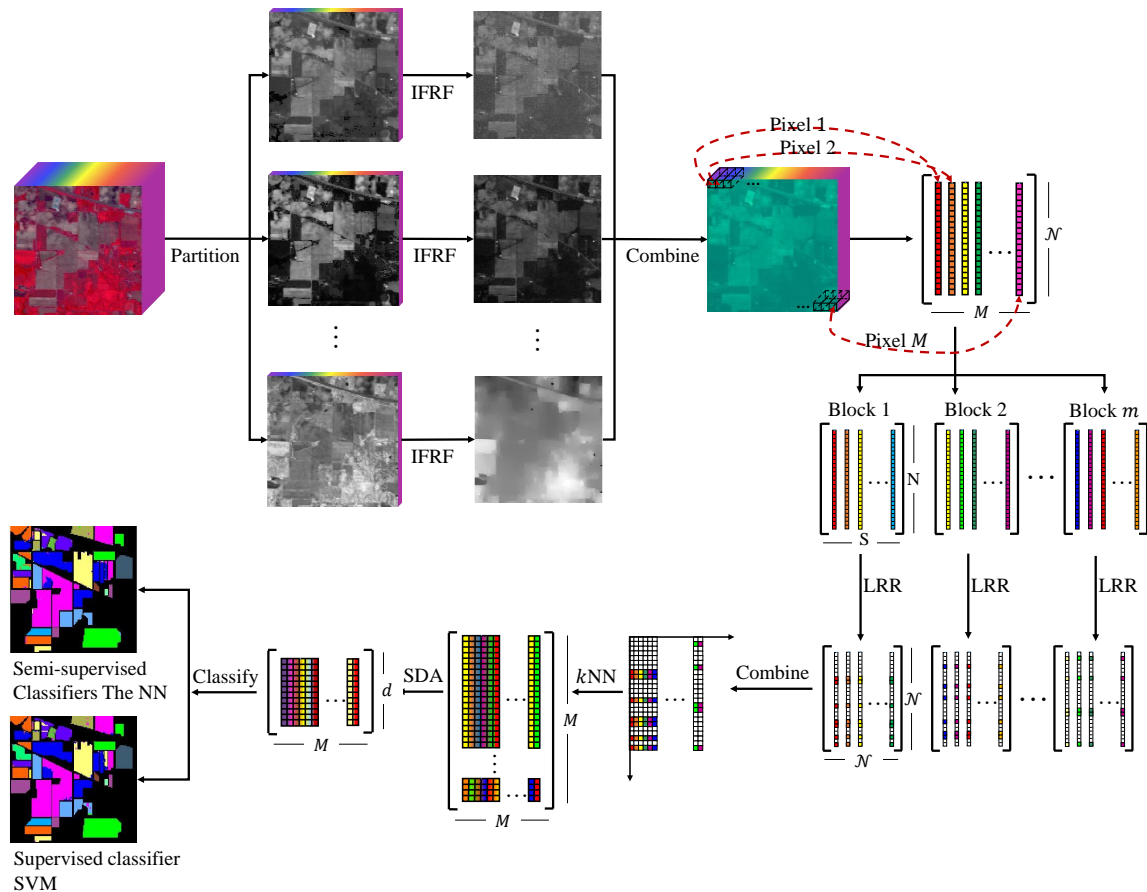


Figure 1. Formulation of the proposed regularized block low-rank discriminant analysis feature extraction for Hyperspectral Image (HSI) classification. IFRF, Image Fusion and Recursive Filtering; SDA, Semi-supervised Discriminant Analysis.

Afterward, divide the processed image into blocks (subsets) by pixels and implement the low-rank representation on each block image. Eventually, we combine the subsets' feature representation with a complete feature graph. Further, the k -nearest neighbor is applied to handle the integral low-rank graph to satisfy the symmetric matrix requirements for the regularized graph of discriminant analysis. Additionally, at the same time, k -nearest neighbor preserves the local information of the image [38]. Furthermore, process the semi-supervised discriminant analysis for feature extraction, which takes advantage of the labeled samples and the distribution of the whole samples. Finally, implement the supervised and semi-supervised classifier methods. We perform comprehensive experiments on several real multi-class HSIs. We summarize the main contributions of the paper in the following paragraphs.

- From the inspiration of LRR and the semi-supervised discriminant analysis algorithm, we proposed a robust feature extraction method, regularized block low-rank discriminant analysis. The block LRR solves the growing computational complexity problem of LRR, which captures the global structure of the data.
- After image fusion and recursive filtering, we implement our proposed regularized block low-rank discriminant analysis feature extraction method. The k NN approach simultaneously addresses

two issues such as SDA's symmetric requirements and capturing the local information of the HSIs. Consequently, the BLRDA method maximally preserves the local geometry of the data.

- Extensive experiments on several multi-class HSIs demonstrate that our proposed BLRDA method is a novel feature extraction method not yet evident in the literature. It can enhance the performance of hyperspectral image classification significantly. Given simple supervised and semi-supervised classifiers, the feature graph achieves significant performance in the HSIs' classification with few labeled samples. Moreover, the BLRDA feature extraction method is much more robust than similar methods.

We organize the remaining part of our paper as follows. We start with a preliminary introduction in Section 2. Section 3 deciphers the HSI classification based on regularized block low-rank discriminant analysis. To check the reproducibility of our proposed method, several experiments on real-world hyperspectral images are carried out in Section 4. In Section 5, we provide the discussion. Conclusions are presented in Section 6.

2. Preliminary

2.1. Regularization

To avoid the overfitting issue, regularization is implemented. Regularization means adding some rules (restrictions) to the objective function that needs training, which amends the solution space and reduces the possibility of finding the wrong solution, so that it does not inflate itself. The constraint is interpreted as a priori knowledge (the regularization parameter is equivalent to introducing the prior distribution to the parameter).

The regularization constraint has a guiding role. When optimizing the error function, it tends to choose the gradient reduction direction. Subsequently, the final solution tends to conform to a priori knowledge (such as the general l -norm a priori knowledge, indicating that the original problem is more likely to be relatively simple and tends to produce sparse parameters).

Regularization has roughly two functions. From the model modification point of view, it is used to balance the two terms in the learning process such as bias-variance, fitting ability-generalization ability, loss function-generalization ability and empirical risk-structural risk. From the model solution point of view, regularization provides the possibility of a unique solution. Eventually, least squares fitting may produce an infinite solution, but adding l_1 or l_2 regularization terms can lead to unique solutions.

In practice, overfitting [14,15] may happen due to limited training samples [16]. To address the overfitting issue, semi-supervised learning was proposed to utilize both labeled samples and unlabeled samples to enhance algorithmic performance. The priori assumption of consistency is the key of SSL, which indicates that nearby points may have the same label or similar embeddings [18]. Cai et al. proposed Semi-supervised Discriminant Analysis (SDA), which takes advantage of both labeled and unlabeled samples [21].

The semi-supervised learning algorithmic performance relies profoundly on the graph construction processes. Hence, inspired by low-rank representation, we aim to improve the semi-supervised discriminant analysis feature extraction method via a novel regularized low-rank-based graph term. Furthermore, we narrated a brief outline of two important techniques, semi-supervised discriminant analysis and the low-rank representation algorithm.

2.2. Overview of Semi-supervised Discriminant Analysis

Semi-supervised Discriminant Analysis (SDA) is derived from Linear Discriminant Analysis (LDA). LDA is a supervised method that minimizes within-class covariance while maximizing between-class covariance. The objective function of LDA is as follows:

$$\mathbf{a}_{opt} = \arg \max_{\mathbf{a}} \frac{\mathbf{a}^T S_b \mathbf{a}}{\mathbf{a}^T S_w \mathbf{a}} \quad (1)$$

where S_b is defined as the between-class scatter and S_w is the within-class scatter matrix.

Overfitting may occur when there are not enough training samples. A possible solution to deal with overfitting could be learned on both labeled and unlabeled data by imposing a normalizer [21]. Applying the notion of the regularized graph, a smoothness penalty is incorporated into the objective function of linear discriminant analysis, and the overfitting problem is solved. The labeled samples in the SDA algorithm are to maximize the different classes' separability. The graph adjacency of all of the samples is used to estimate the fundamental geometric information, which is called the regularized graph.

Given a set of samples $[\mathbf{x}_1, \dots, \mathbf{x}_m, \mathbf{x}_{m+1}, \dots, \mathbf{x}_{m+l}]$, where $N = m + l$, the first m samples are labeled as $[y_1, \dots, y_m]$, and the remaining l samples are unlabeled. There are c classes. A rejection matrix \mathbf{a} is obtained here in the SDA method [21], which can present the prior assumption of consistency according to a regularization term.

$$\mathbf{a}_{opt} = \arg \max_{\mathbf{a}} \frac{\mathbf{a}^T S_b \mathbf{a}}{\mathbf{a}^T S_t \mathbf{a} + \alpha J(\mathbf{a})} \quad (2)$$

Here, S_t is defined as the total class scatter matrix. α is the parameter that balances the complexity and empirical loss of the model. The regularized graph $J(\mathbf{a})$ controls the learning complexity of the hypothesis family.

Given a set of examples, S_{ij} is the graph adjacency that represents the model of the relationships of nearby data points. Typically, the k -nearest neighbor graph is used to calculate the neighbor points' relationship, where an edge between the nearest neighbor is the weight of each other.

The model can easily incorporate the priori knowledge in the regularization term $J(\mathbf{a})$, as follows:

$$\begin{aligned} J(\mathbf{a}) &= \sum_{ij} \left(\mathbf{a}^T \mathbf{x}_i - \mathbf{a}^T \mathbf{x}_j \right)^2 S_{ij} \\ &= 2 \sum_i \mathbf{a}^T \mathbf{x}_i \mathbf{D}_{ii} \mathbf{x}_i^T \mathbf{a} - 2 \sum_{ij} \mathbf{a}^T \mathbf{x}_i S_{ij} \mathbf{x}_j^T \mathbf{a} \\ &= 2 \mathbf{a}^T \mathbf{X} (\mathbf{D} - \mathbf{S}) \mathbf{X}^T \mathbf{a} \\ &= 2 \mathbf{a}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{a} \end{aligned} \quad (3)$$

The column (or row, since \mathbf{S} is symmetric) sum of \mathbf{S} , i.e., $\mathbf{D}_{ii} = \sum_j S_{ij}$ is the diagonal matrix \mathbf{D} . The Laplacian matrix [23] is $\mathbf{L} = \mathbf{D} - \mathbf{S}$. Then, the SDA's objective function with regularization term $J(\mathbf{a})$ is:

$$\mathbf{a}_{opt} = \max_{\mathbf{a}} \frac{\mathbf{a}^T S_b \mathbf{a}}{\mathbf{a}^T (S_t + \alpha \mathbf{X} \mathbf{L} \mathbf{X}^T) \mathbf{a}} \quad (4)$$

We obtain the projective vector \mathbf{a} by maximizing the generalized eigenvalue problem where d is the weight matrix's rank for the labeled graph.

$$S_b \mathbf{a} = \lambda \left(S_t + \alpha \mathbf{X} \mathbf{L} \mathbf{X}^T \right) \mathbf{a} \quad (5)$$

2.3. Low-Rank Representation

The low-rank representation is first proposed to construct a graph (LR-graph) [23], which jointly receives the graph according to the low-rank constraint. It can efficiently catch the global structures of the data [39].

Suppose $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ is the samples set. Each column is an object sample. Therefore, we represent the feature matrix \mathbf{X} on a given dictionary \mathbf{A} [29] as:

$$\mathbf{X} = \mathbf{AZ} \quad (6)$$

where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n] \in \mathbb{R}^{l \times n}$ is the low-rank coefficient matrix and \mathbf{z}_i is the representation coefficient of \mathbf{x}_i . \mathbf{z}_i is a linear combination, which is called the representation coefficient. The so-called low-rank representation is a matrix for this coefficient.

For example, in HSIs, samples of the same category exhibit similar spectra, while samples from different groups do not present similar spectra. Due to the apparent intra-class similarity and inter-class dissimilarity, each sample can be represented well by others in the same class, while not by others in another class. Therefore, when being represented on the dictionary \mathbf{A} , samples from the i -th class \mathbf{x}_i will produce a significant representation on the \mathbf{A}_i component and small factors on other parts \mathbf{A}_j ($j \neq i$). The appropriate permutations on columns of \mathbf{X} , \mathbf{Z} will exhibit apparent diagonal-block structure, as shown in Figure 2.

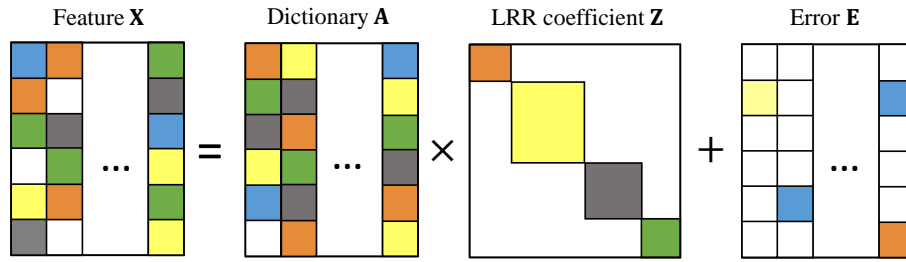


Figure 2. The representation with a diagonal-block structure.

However, the actual labels of samples are unknown in \mathbf{X} ; it is intractable to directly reveal the block-diagonal structure in \mathbf{Z} . Nevertheless, the underlying block-diagonal structure enables \mathbf{Z} to be low-rank [40]. Therefore, we exploit the low-rank property of \mathbf{Z} instead of the block-diagonal structure implicitly. Furthermore, each sample can be well represented due to the intra-class similarity. The following low-rank framework [29] searches the lowest rank solution:

$$\begin{aligned} \min_{\mathbf{Z}} \text{rank}(\mathbf{Z}) \\ \text{s.t. } \mathbf{X} = \mathbf{AZ} \end{aligned} \quad (7)$$

Due to the discrete nature of the rank function, the above optimization problem is onerous to sort out. It is evident from the literature that by matrix completion methods (e.g., [41]), the optimization problem is changed to a convex problem:

$$\begin{aligned} \min_{\mathbf{Z}} \|\mathbf{Z}\|_* \\ \text{s.t. } \mathbf{X} = \mathbf{AZ} \end{aligned} \quad (8)$$

where $\|\cdot\|_*$ is the nuclear norm (i.e., trace norm) [42], which is the matrix's singular values' summation. A more reasonable function is proposed by taking into consideration the noise or corruption in real-world situations, where the residual matrix is often sparse.

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} \quad & \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_l \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{AZ} + \mathbf{E} \end{aligned} \quad (9)$$

where $\|\cdot\|_l$ can be the $l_{2,1}$ -norm or l_1 -norm. Here, the $l_{2,1}$ -norm is chosen as the error term, which is $\|\mathbf{E}\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^n ([\mathbf{E}]_{ij})^2}$. λ is to balance the low-rank and the error term. Here, the inexact Augmented Lagrange Multipliers (ALM) method [43,44] is applied to obtain the optimization solution \mathbf{Z}^* .

2.4. Image Fusion and Recursive Filtering Feature

Initially, we extracted the spatial information of the HSI by using Image Fusion and Recursive Filtering (IFRF), which is one of the easiest approaches to image fusion [17]. The IFRF feature of an HSI contains the main information of the HSI, which eradicates noise and redundant information simultaneously. It plays a vital role in identifying objects and can be used to discriminate between different classes in the classification problem [17]. Hence, we preprocess the HSI first by IFRF to eliminate noise and redundant information.

Let us suppose $R = (r_1, r_2, \dots, r_D) \in \mathbb{R}^{M \times D}$ represents the original hyperspectral image, which has D -dimensional bands and M pixels in the i -th band of the whole image represented by r_i . We have attempted to spectrally partition the whole hyperspectral image (hyperspectral bands) into multiple subsets. Each subset is composed of several K continuous bands. Here, we define \mathcal{N} as the number of subsets. Further, we can get $\mathcal{N} = \lfloor D/K \rfloor$, where $\lfloor D/K \rfloor$ indicates the floor operation that computes the biggest integer no larger than D/K . The i -th ($i \in (1, 2, \dots, \mathcal{N})$) subset is shown as follows:

$$\mathbf{P}^i = \begin{cases} (r_i, \dots, r_{(i+K)}), & \text{if } (i+K) \leq D \\ (r_i, \dots, r_D), & \text{otherwise.} \end{cases} \quad (10)$$

Afterward, fuse the adjacent bands of each subset by image fusion method (the averaging method). For example, the i -th image fusion feature F_i , i.e., the i -th fusion band, is computed as follows,

$$Q^i = \frac{\sum_{n=1}^{N_i} \mathbf{P}_n^i}{N_i} \quad (11)$$

Here, \mathbf{P}_n^i refers to the n -th band in the hyperspectral image's i -th subset. N_i is the i -th subset's band number. After image fusion, we remove the noise pixels and redundant information for each subset. Then, we obtain the i -th feature by recursive filtering on the above fusion band Q^i .

$$O^i = \mathcal{RF}_{\delta_s, \delta_r}(Q^i) \quad (12)$$

Here, \mathcal{RF} indicates the recursive filtering transform method. δ_s is defined as the filter's spatial standard deviations, and δ_r is the range standard deviations [45]. After image fusion and recursive filtering, we obtain the feature image $\mathbf{O} = [O^1, \dots, O^{\mathcal{N}}] \in \mathbb{R}^{M \times \mathcal{N}}$. Let $\mathbf{X} = \mathbf{O}^T = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M] \in \mathbb{R}^{\mathcal{N} \times M}$ be the preprocessed features vector, where \mathbf{x}_i represents a pixel with \mathcal{N} band numbers (dimension) in the hyperspectral image.

3. Methodology

In the above section, we study the classic regularized discriminant analysis, semi-supervised discriminant analysis and the LRR, which can efficiently capture the global structures of the data. Furthermore, the labeled samples, as well as the unlabeled samples can be exploited simultaneously by depicting the underlying block-diagonal structure of \mathbf{Z} . Considering this, we have attempted to

decipher in our present work the robust regularized block low-rank discriminant analysis feature extraction method. We have comprehensively provided insight into the optimization method based on the inexact ALM algorithm. The framework of the HSI classification applying the proposed graph incorporation is found in the following part of this section.

3.1. Regularized Block Low-Rank Discriminant Analysis

The HSI features' vector $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M] \in \mathbb{R}^{\mathcal{N} \times M}$, \mathcal{N} is the channel number (data dimension), and M is the pixels number (sample or object number). Determining an appropriate subspace for classification is an important task. The low-rank representation feature graph is an onerous task; the operating time increases exponentially with a growing number of samples. Hence, we explore the low-rank structure via block theory, which divides the whole image into blocks for low-rank representation.

We choose the block size as \mathcal{S} pixels in each block partition. Let $\{g_1, g_2, \dots, g_m\}$ represent m blocks' set index of the image, where there are \mathcal{S} pixels in each g_i . To formulate the matrix \mathbf{X} , according to $\{g_1, g_2, \dots, g_m\}$, we can put the vectors belonging to the same blocks together in the form of $\mathbf{X} = \{\mathbf{X}_{g_1}, \mathbf{X}_{g_2}, \dots, \mathbf{X}_{g_m}\}$, $\mathbf{A} = \{\mathbf{A}_{g_1}, \mathbf{A}_{g_2}, \dots, \mathbf{A}_{g_m}\}$ and $\mathbf{E} = \{\mathbf{E}_{g_1}, \mathbf{E}_{g_2}, \dots, \mathbf{E}_{g_m}\}$. Further, the LRR optimization problem for each block is converted into the following form:

$$\begin{aligned} \min_{\mathbf{Z}_{g_i}, \mathbf{E}_{g_i}} \|\mathbf{Z}_{g_i}\|_* + \lambda \|\mathbf{E}_{g_i}\|_l \\ \text{s.t. } \mathbf{X}_{g_i} = \mathbf{A}_{g_i} \mathbf{Z}_{g_i} + \mathbf{E}_{g_i} \end{aligned} \tag{13}$$

where we choose the $l_{2,1}$ -norm as the error term $\|\cdot\|_l$, which is $\|\mathbf{E}_{g_i}\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^n ([\mathbf{E}_{g_i}]_{ij})^2}$. Here, we use the inexact ALM method [43,44] as the optimal solution $\mathbf{Z}_{g_i}^*$.

For properly optimizing problem (13), by presenting an auxiliary variable \mathbf{J}_i regarding $\mathbf{Z}_{g_i}^*$, we convert it to the similar issue as follows [29]:

$$\begin{aligned} \min_{\mathbf{Z}_{g_i}, \mathbf{E}_{g_i}, \mathbf{J}_i} \|\mathbf{J}_i\|_* + \lambda \|\mathbf{E}_{g_i}\|_{2,1} \\ \text{s.t. } \mathbf{X}_{g_i} = \mathbf{A}_{g_i} \mathbf{Z}_{g_i} + \mathbf{E}_{g_i}, \\ \mathbf{Z}_{g_i} = \mathbf{J}_i \end{aligned} \tag{14}$$

Incorporating an intermediate variable \mathbf{J}_i , we minimize the augmented Lagrange function as follows via the inexact ALM method [43]:

$$\begin{aligned} L(\mathbf{Z}_{g_i}, \mathbf{E}_{g_i}, \mathbf{J}_i, \mathbf{Y}_1, \mathbf{Y}_2) = \|\mathbf{J}_i\|_* + \lambda \|\mathbf{E}_{g_i}\|_{2,1} + \text{tr}[\mathbf{Y}_1, \mathbf{X}_{g_i} - \mathbf{A}_{g_i} \mathbf{Z}_{g_i} - \mathbf{E}_{g_i}] + \text{tr}[\mathbf{Y}_2, \mathbf{Z}_{g_i} - \mathbf{J}_i] \\ + \frac{\mu}{2} (\|\mathbf{X}_{g_i} - \mathbf{A}_{g_i} \mathbf{Z}_{g_i} - \mathbf{E}_{g_i}\|_F^2 + \|\mathbf{Z}_{g_i} - \mathbf{J}_i\|_F^2) \end{aligned} \tag{15}$$

where \mathbf{Y}_1 , \mathbf{Y}_2 and $\mu > 0$ are Lagrangian multipliers and the penalty parameter, respectively. To minimize the problem (15), we update the variables \mathbf{Z}_{g_i} , \mathbf{J}_i , \mathbf{E}_{g_i} , \mathbf{Y}_1 and \mathbf{Y}_2 alternately with others fixed. Meanwhile, the sub-problem is convex regarding all the variables; it can supply a relevant, unique solution. The inexact ALM is applied to update the variables \mathbf{J}_i , \mathbf{Z}_{g_i} and \mathbf{E}_{g_i} iteratively. The optimization process outline is given in the following Algorithm 1.

After the block low-rank representation, we get $\mathbf{Z}_{g_i} \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}$. Then, we combine these feature representation subsets with a whole feature graph $\mathbf{Z} = \{\mathbf{Z}_{g_1}, \mathbf{Z}_{g_2}, \dots, \mathbf{Z}_{g_m}\} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M] \in \mathbb{R}^{\mathcal{S} \times M}$.

The combined low-rank graph is composed of multiple blocks graph \mathbf{Z} , which is asymmetric. Traditionally, the symmetrization process is used to determine the mean value of its transposition by itself to satisfy the symmetric requirement of SDA. However, the graph adjacency matrix \mathbf{Z} is not a square matrix. Our previous work shows that the k -nearest neighbor is an effective symmetrization

process that can significantly improve the performance of SDA [38]. Hence, the k NN algorithm integrally handles the combined low-rank graph, which addresses two issues, such as addressing the symmetrization process and further maximally preserving the local information of the image by the k NN's property.

The samples \mathbf{z}_i and \mathbf{z}_j are considered as neighbors if \mathbf{z}_i is among the k -nearest neighbor of \mathbf{z}_j or \mathbf{z}_j is among the k -nearest neighbor of \mathbf{z}_i . Here, we use the Heat kernel weighting [13] method to assign weights for \mathbf{S} as follows:

$$\mathbf{S}_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{z}_i - \mathbf{z}_j\|^2}{2\sigma^2}\right), & \text{if } \mathbf{z}_i \in N_k(\mathbf{z}_j) \text{ or } \mathbf{z}_j \in N_k(\mathbf{z}_i) \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

where $N_k(\mathbf{z}_i)$ denotes k neighbors of \mathbf{z}_i in (16).

Therefore, the BLRDA method can achieve high-quality feature representations for classification.

Algorithm 1 Resolving function (15) by the inexact ALM.

Input: Mapped data graph \mathbf{X}_{g_i} , regularization parameter α for local affinity. $\mathbf{Z}_{g_i} = \mathbf{J}_i = \mathbf{0}$, $\mathbf{E}_{g_i} = \mathbf{0}$, $\mathbf{Y}_1 = \mathbf{0}$, $\mathbf{Y}_2 = \mathbf{0}$, $\mu = 10^{-6}$, $\mu_{\max} = 10^6$, $\rho = 1.1$ and $\varepsilon = 10^{-8}$.

1: **while** No convergence **do**

2: Fix other variables and update \mathbf{J}_i ;

$$\mathbf{J}_i = \arg \min \frac{1}{\mu} \|\mathbf{J}_i\|_* + \frac{1}{2} \|\mathbf{J}_i - (\mathbf{Z}_{g_i} + \mathbf{Y}_2 / \mu)\|_F^2$$

3: Fix other variables and update \mathbf{Z}_{g_i} ;

$$\mathbf{Z}_{g_i} = (\mathbf{I} + \mathbf{A}_{g_i}^T \mathbf{A}_{g_i})^{-1} (\mathbf{A}_{g_i}^T (\mathbf{X}_{g_i} - \mathbf{E}_{g_i}) + \mathbf{J}_i + ((\mathbf{A}_{g_i}^T \mathbf{Y}_1 - \mathbf{Y}_2) / \mu))$$

4: Fix the others and update \mathbf{E}_{g_i} ;

$$\mathbf{E}_{g_i} = \arg \min \frac{\lambda}{\mu} \|\mathbf{E}_{g_i}\|_{2,1} + \frac{1}{2} \|\mathbf{E}_{g_i} - (\mathbf{X}_{g_i} - \mathbf{A}_{g_i} \mathbf{Z}_{g_i} + \mathbf{Y}_1 / \mu)\|_F^2$$

5: Update the multipliers as follows:

$$\mathbf{Y}_1 = \mathbf{Y}_1 + \mu \mathbf{X}_{g_i} - \mathbf{A}_{g_i} \mathbf{Z}_{g_i} - \mathbf{E}_{g_i},$$

$$\mathbf{Y}_2 = \mathbf{Y}_2 + \mu (\mathbf{Z}_{g_i} - \mathbf{J}_i)$$

6: Update parameter μ by $\mu = \min(\rho\mu, \mu_{\max})$;

7: Examine convergence conditions

$$\|\mathbf{X}_{g_i} - \mathbf{A}_{g_i} \mathbf{Z}_{g_i} - \mathbf{E}_{g_i}\|_\infty < \varepsilon \text{ and } \|\mathbf{Z}_{g_i} - \mathbf{J}_i\|_\infty < \varepsilon$$

8: **end while**

Output: The Laplacian regularized low-rank representation graph \mathbf{Z}_{g_i} .

3.2. BLRDA Feature Extraction for HSI Classification

The semi-supervised discriminant analysis method can successfully solve the overfitting problem with few labeled samples. Given a set of samples $\{(\mathbf{x}_i, y_i)\}_{i=1}^l$ with c classes and unlabeled samples $\{\mathbf{x}_i\}_{i=l+1}^m$, the l_k is the samples of the k -th class. The algorithmic procedure of HSI classification by applying the regularized block low-rank discriminant analysis feature extraction method for both supervised and semi-supervised classification is stated below:

Step 1 Construct the adjacency graph: Construct the block low-rank and k NN graph \mathbf{S} in Formula (16) for the regularization term. In addition, calculate the graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{S}$.

Step 2 Construct the labeled graph: for the labeled graph, construct the matrix as:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{l \times l} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (17)$$

Define $\tilde{\mathbf{I}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where $\mathbf{I} \in l \times l$ is an identity matrix.

Step 3 Eigen-problem: Calculate the eigenvectors for the generalized eigenvector problem:

$$\mathbf{X}\mathbf{W}\mathbf{X}^T \mathbf{a} = \lambda \mathbf{X} (\tilde{\mathbf{I}} + \alpha \mathbf{L}) \mathbf{X}^T \mathbf{a} \quad (18)$$

where $[\mathbf{x}_1, \dots, \mathbf{x}_m, \mathbf{x}_{m+1}, \dots, \mathbf{x}_{m+l}]$. The \mathbf{W} is of rank d , and we will have d eigenvectors denoted as $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_d\}$.

Step 4 Regularized discriminant analysis embedding: Let the transformation matrix $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_d] \in \mathbb{R}^{M \times d}$. Finally, embedding the samples into a d -dimensional subspace,

$$\mathbf{x} \rightarrow \mathbf{z} = A^T \mathbf{x} \quad (19)$$

we can obtain the between-class scatter S_b and the total class scatter matrix S_t .

$$\mathbf{X}\mathbf{W}\mathbf{X}^T = \mathbf{X}_l \mathbf{W}_{l \times l} \mathbf{X}_l^T = S_b$$

$$\mathbf{X}\tilde{\mathbf{I}}\mathbf{X}^T = \mathbf{X}_l \mathbf{X}_l^T = S_t$$

Therefore, the Eigen-problem in Formula (18) is the same as the Eigen problem in Formula (5), and we can obtain the projective matrix \mathbf{A} :

$$S_b \mathbf{A} = \lambda (S_t + \alpha \mathbf{X} \mathbf{L} \mathbf{X}^T) \mathbf{A} \quad (20)$$

where \mathbf{S} is the graph after k NN, i.e., \mathbf{S}_{ij} is the graph of nearby data points' relationships. $\mathbf{L} = \mathbf{D} - \mathbf{S}$ is the Laplacian matrix [23]. $\mathbf{D}_{ii} = \sum_j \mathbf{S}_{ij}$ is the diagonal matrix \mathbf{D} .

Then, graph \mathbf{S} can be embedded into the d -dimensional subspace as the SDA embedding matrix Φ .

$$\Phi = A^T \mathbf{S} \quad (21)$$

Step 5 HSIs' Classification: Finally, perform the supervised and semi-supervised classifier nearest neighbor and SVM classifier for classification. The classifiers mentioned above are simple and ubiquitously-used classifiers.

4. Experiments and Analysis

To investigate the performance of the BLRDA feature extraction method, we have conducted extensive experiments on several real multi-class hyperspectral images. In this section, we have incorporated the hyperspectral images used in the present study and illustrated the performance of our proposed methods. We performed the experimental analysis on Intel Core CPU 2.60 GHz and 8 GB RAM machines.

4.1. Experimental Setup

4.1.1. Datasets

We evaluate the proposed method on three hyperspectral images, namely the Indian Pines image, the Pavia University scene and the Salinas image (http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).

- The Indian Pines image is for the agricultural Indian Pine test site in Northwestern Indiana, which was obtained by the sensor AVIRIS. There are 145×145 pixels and 224 spectral bands between the wavelength range 400 nm and 2500 nm. The image scene comprises two-thirds agriculture and one-third forest or other natural perennial vegetation. It includes two major dual-lane highways, a rail line, as well as some low-density housing, other built structures and smaller roads. There is minuscule coverage, approximately less than 5%, because of some of

the crops present; corn and soybeans are in the early stages of growth in June. The ground truth is divided into sixteen classes and is not all mutually exclusive. Twenty water absorption bands (Nos. 104–108, 150–163 and 220) were removed. The Indian Pines image's color composite and the ground truth image are demonstrated in Figure 3.

- The Pavia University scene was obtained from an urban area surrounding the University of Pavia, Italy on 8 July 2002. It was recorded by a Reflective Optics System Imaging Spectrometer, which has a 1.3-m spatial resolution and spectral coverage ranging between 0.43 and 0.86 μm . There are 115 bands of size 610×340 in the image. After removing 12 channels, 103 channels are left for testing. The Pavia University scene's color composite and the corresponding ground truth image are shown in Figure 4.
- The Salinas image was obtained from Salinas Valley, CA, USA, by the AVIRIS sensor. It has a high spatial resolution with 3.7-m pixels. The image includes 224 bands and comprises 512 lines with 217 samples. Similar to the Indian Pines image, 20 water absorption bands, including [108–112], [154–167] and 224, were discarded. This image is available only as at-sensor radiance data. The image contains vegetables, bare soil, vineyard fields, and so on, with 16 classes. Figure 5 shows the color composite and the corresponding ground truth of the Salinas image.

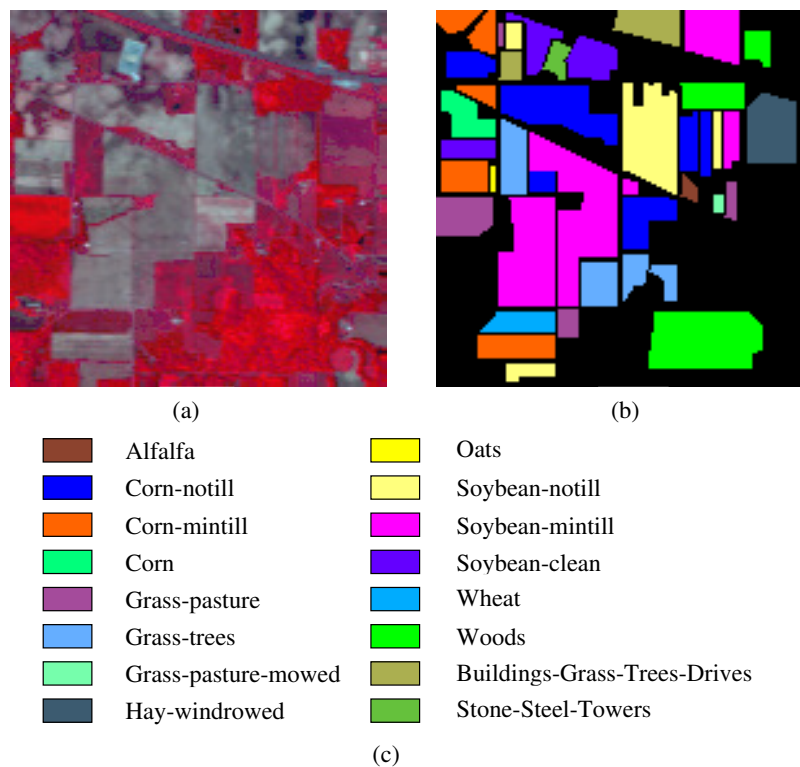


Figure 3. Indian Pines dataset. (a) Three-band color composite of the Indian Pines image. (b,c) Ground truth image and reference data.

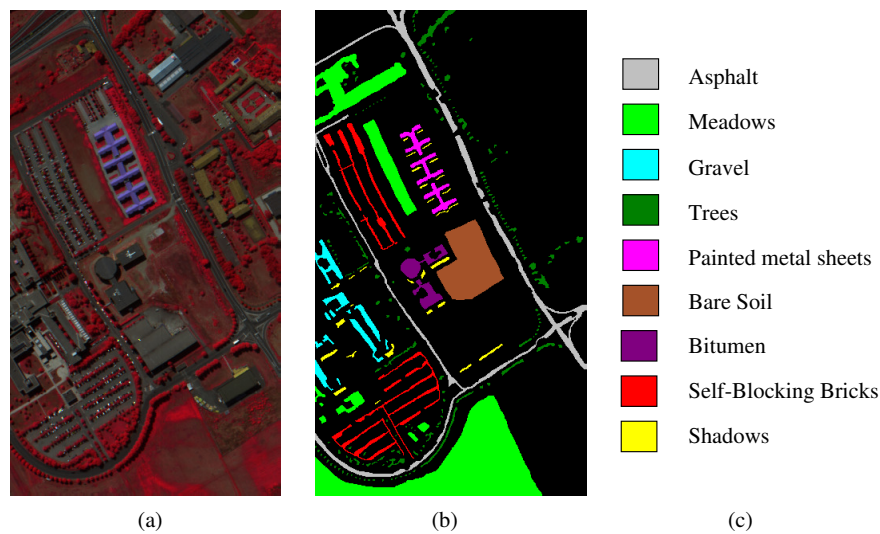


Figure 4. Pavia University scene dataset. (a) Three-band color composite of the Pavia University scene. (b,c) Ground truth image and reference data.

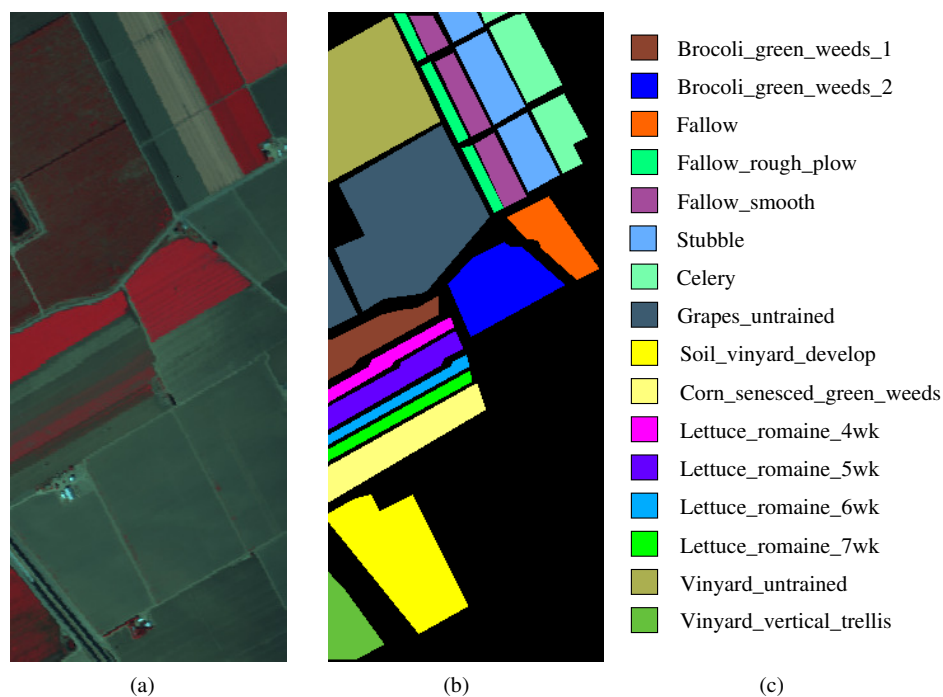


Figure 5. Salinas dataset. (a) Three-band color composite of the Salinas image. (b,c) Ground truth image and reference data.

4.1.2. Evaluation Criteria

To evaluate the proposed method of HSIs, we give some evaluation criteria as follows.

Classification Accuracy (CA) refers to the pixels in the image classification in each class. In the field of remote sensing classification, the confusion matrix [46] is frequently used, which is defined in the form: $M = [m_{ij}]_{n \times n}$, where m_{ij} denotes that the number of pixels labeled by j should belong to class i . n is the class number. The reliability of classification depends on the diagonal values of the confusion matrix. Higher values of the confusion matrix are suggested for favorable results.

The three primary indicators we used include the Overall Accuracy (OA), Average Accuracy (AA) and the kappa coefficient [47]. For the three metrics, OA refers to the percentage of pixels

correctly classified. AA measures the average correctly-classified pixels' percentage of each class. To make the measurements more objective, the kappa coefficient is used to estimate the percentage of correctly-classified pixels. Whereas OA and AA check how many of all pixels are classified correctly, assuming that the reference classification (ground truth) is true, here it is assumed that both classification and reference classification are independent class assignments of equal reliability; how well they agree is what is measured. The big advantage of the kappa coefficient over overall accuracy is that the kappa coefficient considers chance agreement and corrects for it. Chance agreement means here the probability that classification and reference classification agree by mere chance. Assuming statistical independence, we obtain for this probability the estimation [48,49].

4.1.3. Classifier Settings

We conduct a series of experiments to test and verify the proposed BLRDA feature extraction method under the supervised and semi-supervised classifiers.

- The Support Vector Machine (SVM) classifier is a commonly-used supervised learning model for classification and regression analysis.
- Nearest Neighbor (NN) is used as the semi-supervised classifier, which stores all available samples and classifies new samples based on similarity measures (e.g., distance functions).

4.1.4. Comparative Algorithms

To explain the significant improvement achieved by using the regularized block low-rank discriminant analysis feature extraction method in the hyperspectral images' classification, several comparative methods are shown in the paper. For fairness, these comparative graphs incorporate the regularized discriminant analysis algorithm, which is shown below,

- BSDA (Block Sparse Representation Discriminant Analysis) method [43]: The Sparse Representation (SR)-graph considers the sparse representation by resolving the problem: $\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_1$. The weight of the graph is $\mathbf{W}_{ij} = \left| a_j^i \right|$.
- BKDA (Block k -Nearest Neighbor Discriminant Analysis) method [43]: Euclidean distance is employed as our similarity measure, and the Gaussian kernel is adopted for the k -nearest neighbor and k NN feature graph. The nearest neighbor number is set as five here.
- BLEDA (Block Locally Linear Embedding Discriminant Analysis) method [11]: Neighbors in Locally Linear Embedding (LLE) consider reconstructing a sample from its neighbor points and then minimizes the l_2 reconstruction error.

$$\begin{aligned} \min \sum_i \left\| \mathbf{x}_i - \sum_j \mathbf{W}_{ij} \mathbf{x}_j \right\|^2 \\ \text{s.t. } \sum_j \mathbf{W}_{ij} = 1 \end{aligned} \quad (22)$$

$\mathbf{W}_{ij} = 0$ if \mathbf{x}_j does not belong to the neighbors of \mathbf{x}_i . The nearest neighbor number k is set as five.

- Image Fusion and Recursive Filtering (IFRF) method mentioned in Section 2.4.

4.2. Supervised and Semi-Supervised HSI Classification Results

To examine the performance of the combined BLRDA feature extraction method, we perform experiments on the three HSIs. Ten independent runs of each algorithm are evaluated by resampling the training samples in each run. We have chosen the mean values as the results. Unlike most of the existing supervised and semi-supervised HSI experiments, we test the performance of all comparative methods using only a small part of the label samples. In a practical scenario, the labeled samples are difficult to capture, and the unlabeled ones are usually present in substantial numbers. Table 1

provides the training and testing data for all classes. For the Indian Pines image, the Pavia University scene and the Salinas image, the training set is approximately 6%, 4% and 0.4%, which are minimal sets compared to the entire dataset. Subsequently, the training sets are chosen randomly. Considering the classes with a meager number of samples, we have incorporated a minimum threshold of training samples. Here, we set the minimum threshold of training samples for each class as five, which can eradicate the difference between the classes with a low number of samples. In our experiments, the filter's spatial standard deviations and range standard deviations are δ_s and δ_r with 200 and 0.3, respectively. The parameter σ in k -nearest neighbor $S_{ij} = \exp\left(-\frac{\|z_i - z_j\|^2}{2\sigma^2}\right)$ is 0.1, which is provided randomly. Hence, the following results are not under the best parameters, which shows the robustness and superiority of our method.

Initially, we utilize different graph construction methods to get the regularized graph of the given hyperspectral images. Then, the SDA algorithm is implemented for feature extraction. Furthermore, the NN method and SVM classifier are applied for final classification in the derived low-dimensional feature subspace. Tables 2–4 show the detailed classification results for CA, OA and AA, as well as the kappa coefficient obtained from various methods, where the bold numbers suggest the best results for different graph algorithms. Figures 6–8 show the classification results (randomly selected from our above experiments) acquired by several methods for the three hyperspectral images associated with the corresponding OA values. It has been observed that the results in the figures are random for each method and do not have comparability. These results reveal the supervised and semi-supervised classification results of different feature extraction methods. From these results, we can see the following.

In most cases, our proposed BLRDA feature extraction method brings about the highest classification accuracy. Therefore, it significantly improves the classification performance of hyperspectral images, which indicates that the BLRDA method is a superior HSI feature extraction method for both NN and SVM classifiers. Consequently, our feature extraction method is robust to both supervised and semi-supervised classification.

Table 1. Training and testing samples for the three hyperspectral images.

Class	Indian Pines Image			Pavia University Scene			Salinas Image		
	Train	Test	Sample No.	Train	Test	Sample No.	Train	Test	Sample No.
C1	8	38	46	271	6360	6631	14	1995	2009
C2	91	1337	1428	752	17,897	18,649	20	3706	3726
C3	55	775	830	90	2009	2099	13	1963	1976
C4	20	217	237	128	2936	3064	11	1383	1394
C5	34	449	483	59	1286	1345	16	2662	2678
C6	49	681	730	207	4822	5029	21	3938	3959
C7	7	21	28	59	1271	1330	20	3559	3579
C8	34	444	478	153	3529	3682	51	11,220	11,271
C9	7	13	20	43	904	947	30	6173	6203
C10	64	908	972				19	3259	3278
C11	153	2302	2455				10	1058	1068
C12	41	552	593				13	1914	1927
C13	18	187	205				9	907	916
C14	81	1184	1265				10	1060	1070
C15	29	357	386				35	7233	7268
C16	11	82	93				13	1794	1807
Total	702	9547	10,249	1762	41,014	42,776	305	53,824	54,129

From Tables 2 and 4, we reach the same conclusion that the proposed approaches BLRDA + NN and BLRDA + SVM are preferable to the other methods, particularly BLRDA + SVM. For example, with our methods, the Corn-notill, Corn-mintill, Grass-pasture, Grass-pasture-mowed

Oats Soybean-notill and Buildings-Grass-Trees-Drives classes are significantly improved. For the Pavia University scene, BLRDA + SVM gives better results. The traditional graph construction methods (such as k NN-graph and LLE-graph) may perform well in some classes, but they are not as stable as our algorithm.

Table 2. Supervised and semi-supervised classification results for the Indian Pines image. BLRDA, Block Low-Rank Discriminant Analysis. BSDA, Block Sparse Representation Discriminant Analysis. BKDA, Block k -Nearest Neighbor Discriminant Analysis. BLEDA, Block Locally Linear Embedding Discriminant Analysis. IFRF, Image Fusion and Recursive Filtering.

Method	BLRDA	BSDA	BKDA	BLEDA	IFRF	BLRDA	BSDA	BKDA	BLEDA	IFRF
Classifier	NN					SVM				
1	1	0.9868	0.9423	0.9351	0.9181	1	0.9675	0.9893	0.9919	0.9925
2	0.9646	0.9348	0.924	0.8979	0.8492	0.9561	0.9339	0.928	0.9133	0.9615
3	0.9621	0.9251	0.9321	0.9333	0.8658	0.9629	0.9588	0.9383	0.9148	0.9674
4	0.9237	0.9126	0.9334	0.9498	0.8186	0.9879	0.8822	0.983	0.9689	0.9499
5	0.9736	0.9787	0.9748	0.9639	0.9719	0.9912	0.9737	0.948	0.9629	0.9880
6	0.9916	0.987	0.9808	0.9807	0.9488	0.9963	0.993	0.9968	0.9990	0.9824
7	0.9886	0.8523	0.8115	0.9181	0.5556	1	0.858	0.9569	0.9788	0.7186
8	1	1	1	1	1	1	1	1	1	1
9	0.8884	0.8567	0.8629	0.8881	0.7147	1	0.9643	0.9714	0.9857	0.8865
10	0.9414	0.9569	0.9447	0.9548	0.8705	0.9824	0.9353	0.943	0.9579	0.9534
11	0.9884	0.9757	0.9767	0.9743	0.9239	0.9596	0.9496	0.949	0.9562	0.9670
12	0.9825	0.9578	0.9234	0.9677	0.8124	0.9782	0.9672	0.967	0.9315	0.9543
13	1	1	1	1	0.9815	1	1	0.9989	1	1
14	0.9951	0.9973	0.9954	0.9969	0.9962	0.9996	0.9994	0.9975	0.9963	0.9944
15	0.9731	0.9753	0.9867	0.974	0.9653	0.9929	0.9719	0.967	0.9518	0.9672
16	0.9878	0.9878	0.9878	0.9878	0.9816	0.9877	0.9836	0.9863	0.9781	0.9755
OA	0.9979	0.9662	0.9567	0.9571	0.9112	0.9713	0.9572	0.9568	0.9544	0.9703
AA	0.9967	0.956	0.9453	0.9563	0.8859	0.9966	0.9513	0.9633	0.9645	0.9537
Kappa	0.9976	0.9615	0.9506	0.951	0.8986	0.9967	0.9512	0.9506	0.9480	0.9661

Table 3. Supervised and semi-supervised classification results for the Pavia University scene. BLRDA, Block Low-Rank Discriminant Analysis. BSDA, Block Sparse Representation Discriminant Analysis. BKDA, Block k -Nearest Neighbor Discriminant Analysis. BLEDA, Block Locally Linear Embedding Discriminant Analysis. IFRF, Image Fusion and Recursive Filtering.

Method	BLRDA	BSDA	BKDA	BLEDA	IFRF	BLRDA	BSDA	BKDA	BLEDA	IFRF
Classifier	NN					SVM				
1	0.9826	0.972	0.9711	0.966	0.9039	0.9655	0.9423	0.9504	0.9566	0.9756
2	0.9968	0.9929	0.9925	0.9911	0.9833	0.9951	0.9431	0.9937	0.9941	0.9967
3	0.9724	0.9615	0.9345	0.9575	0.8753	0.9641	0.9635	0.9424	0.9439	0.9451
4	0.9906	0.9967	0.9889	0.9939	0.9634	0.9951	0.9946	0.9704	0.9861	0.9817
5	1	1	0.9996	0.9984	0.9966	0.9979	0.9586	0.8561	0.9241	0.94
6	0.9982	0.9973	0.9959	0.9961	0.9795	0.9973	0.9783	0.9989	0.9979	0.9988
7	0.9722	0.9539	0.9502	0.9772	0.8787	0.9731	0.897	0.9659	0.9737	0.966
8	0.9651	0.9436	0.9447	0.9472	0.8510	0.9600	0.9991	0.943	0.9488	0.9569
9	0.9501	0.9661	0.9145	0.927	0.8811	0.9557	0.9352	0.9202	0.9251	0.9573
OA	0.9887	0.9831	0.9801	0.9815	0.9465	0.9847	0.9797	0.9755	0.9771	0.9828
AA	0.9809	0.9762	0.967	0.9731	0.9236	0.9782	0.9636	0.9543	0.9611	0.9687
Kappa	0.9850	0.978	0.9736	0.9754	0.9289	0.9796	0.9731	0.9674	0.9696	0.9772

Table 4. Supervised and semi-supervised classification results for the Salinas image. BLRDA, Block Low-Rank Discriminant Analysis. BSDA, Block Sparse Representation Discriminant Analysis. BKDA, Block k -Nearest Neighbor Discriminant Analysis. BLEDA, Block Locally Linear Embedding Discriminant Analysis. IFRF, Image Fusion and Recursive Filtering.

Method	BLRDA	BSDA	BKDA	BLEDA	IFRF	BLRDA	BSDA	BKDA	BLEDA	IFRF
Classifier	NN					SVM				
1	1	1	1	0.9999	0.9929	1	1	1	0.9983	1
2	1	0.9992	0.9956	0.9999	0.9842	1	0.9948	0.9919	0.998	1
3	0.9992	0.988	0.9976	0.9965	0.9839	0.9997	0.9957	0.9957	0.9871	0.9979
4	0.9772	0.9567	0.9589	0.9455	0.9002	0.9834	0.9387	0.9455	0.9609	0.9449
5	0.9914	0.9996	0.9994	0.9969	0.9732	1	1	0.9983	0.9993	0.9918
6	0.9999	0.9992	0.9996	1	1	1	0.997	0.9992	0.9937	1
7	0.9988	0.9993	0.9993	0.9966	0.9864	0.9971	0.9979	0.9978	0.9946	0.9953
8	0.9927	0.9859	0.9831	0.9878	0.9547	0.9875	0.98	0.9918	0.9882	0.9826
9	0.9981	0.9967	0.9949	0.9924	0.9967	0.9999	0.9969	0.9948	0.9988	0.9993
10	0.9935	0.9969	0.9936	0.9951	0.9850	0.9962	0.987	0.9918	0.9592	0.9933
11	0.9998	0.9967	0.9976	0.9863	0.9222	1	0.9976	0.9944	0.9951	0.9939
12	0.9737	0.9644	0.971	0.9741	0.9308	0.9979	0.9751	0.9868	0.9675	0.9961
13	0.9825	0.9842	0.9907	0.9678	0.8912	1	0.9983	0.9893	0.9886	0.9733
14	0.9627	0.9766	0.9879	0.9845	0.9789	0.9895	0.9862	0.9707	0.9622	0.9697
15	0.9617	0.9677	0.9802	0.9567	0.8273	0.9963	0.91	0.9646	0.9475	0.9869
16	0.9994	1	1	0.9999	0.9970	1	0.9948	0.9994	1	0.9998
OA	0.99	0.9883	0.99	0.9866	0.9500	0.9931	0.9795	0.9893	0.9871	0.9905
AA	0.989	0.9885	0.9911	0.9869	0.9565	0.9946	0.9861	0.9903	0.9887	0.9891
Kappa	0.9888	0.9869	0.9888	0.9851	0.9443	0.9923	0.9771	0.9881	0.9856	0.9894

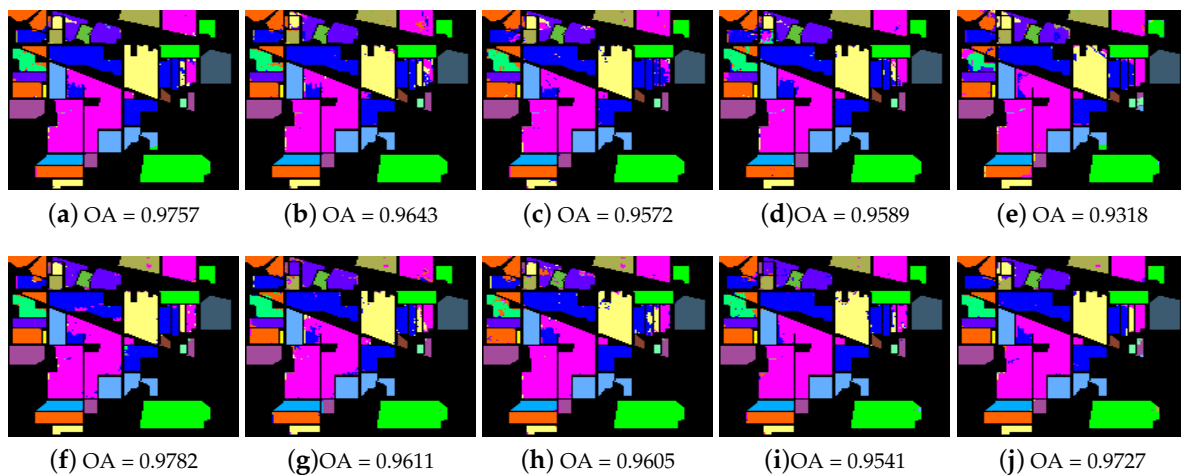


Figure 6. Supervised and semi-supervised classification results (Indian Pine image) obtained by: (a) BLRDA + NN; (b) BSDA + NN; (c) BKDA + NN; (d) BLEDA + NN; (e) IFRF + NN; (f) BLRDA + SVM; (g) BSDA + SVM; (h) BKDA + SVM; (i) BLEDA + SVM; (j) IFRF + SVM.

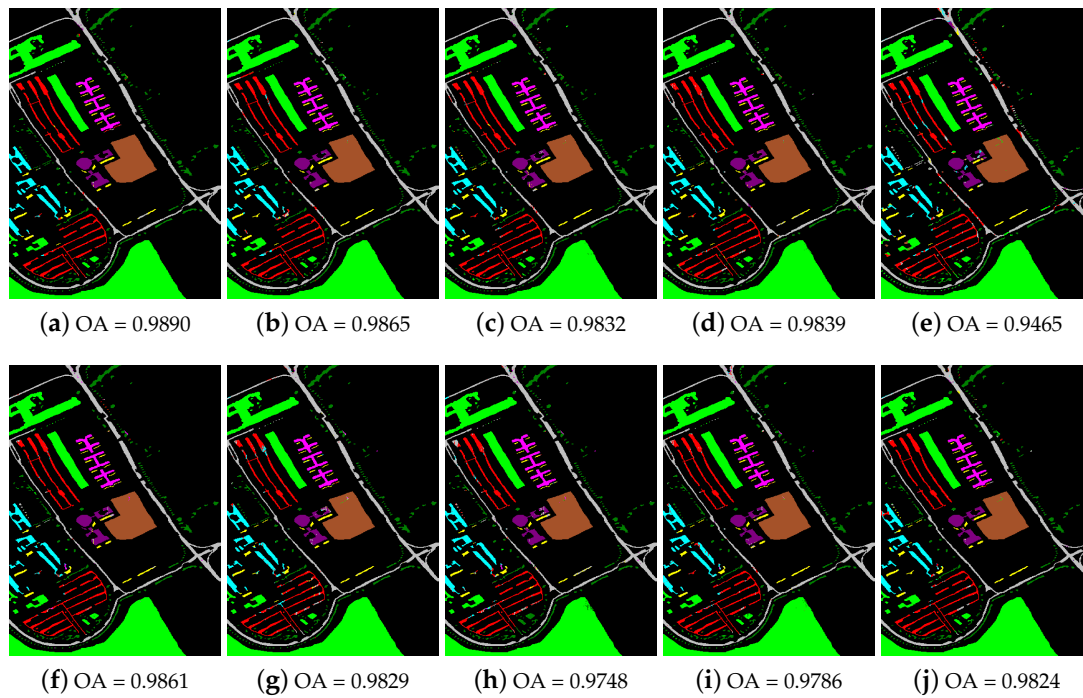


Figure 7. Supervised and semi-supervised classification results (Pavia University scene) obtained by: (a) BLRDA + NN; (b) BSDA + NN; (c) BKDA + NN; (d) BLEDA + NN; (e) IFRF + NN; (f) BLRDA + SVM; (g) BSDA + SVM; (h) BKDA + SVM; (i) BLEDA + SVM; (j) IFRF + SVM.

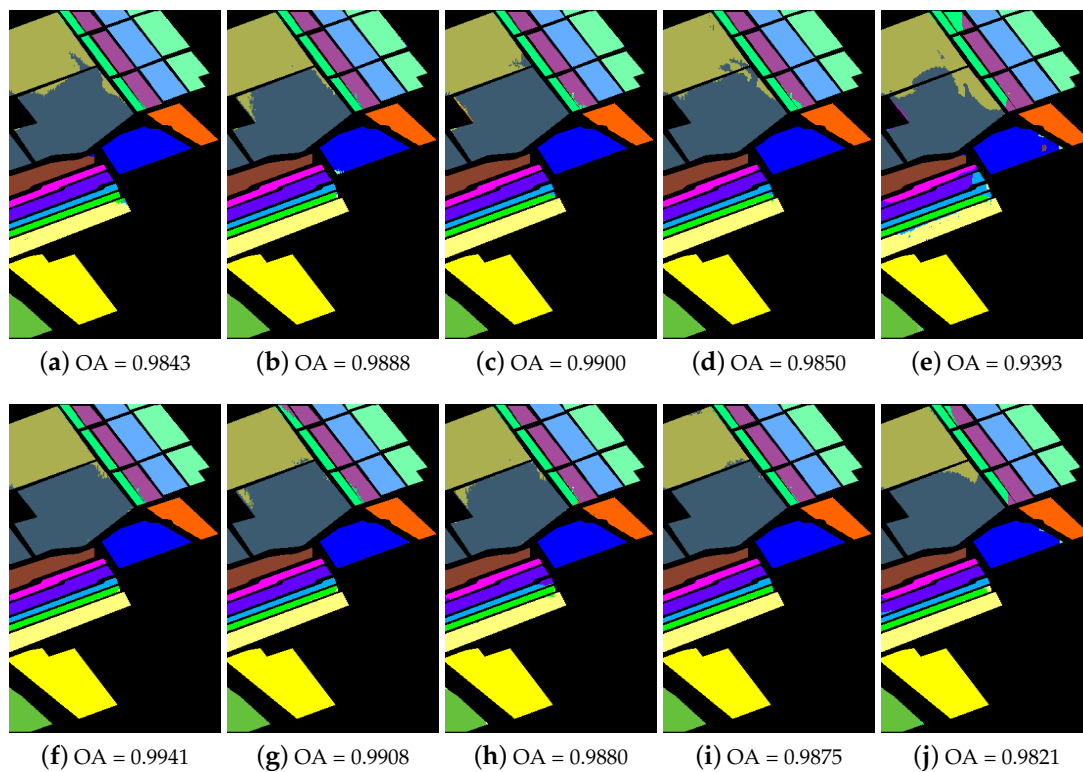


Figure 8. Supervised and semi-supervised classification results (Salinas image) obtained by: (a) BLRDA + NN; (b) BSDA + NN; (c) BKDA + NN; (d) BLEDA + NN; (e) IFRF + NN; (f) BLRDA + SVM; (g) BSDA + SVM; (h) BKDA + SVM; (i) BLEDA + SVM; (j) IFRF + SVM.

4.3. Running Time

We analyzed the running times of different models on the Indian Pines scene, Pavia University scene and Salinas scene images. We used 10 separate runs to calculate the total time. We give the mean running time. As shown in Table 5, the execution time of the BLRDA method was slightly longer compared to the others. Although our algorithm is slower than the traditional k NN algorithms, the performance is much better than these baseline algorithms at an acceptable running time.

Table 5. Run time of different methods on real-word HSIs (units).

Method	BLRDA	BSDA	BKDA	BLEDA	IFRF	BLRDA	BSDA	BKDA	BLEDA	IFRF
Classifier	NN					SVM				
Indian Pines	32.05	69.36	19.60	68.03	12.24	7.54	68.78	8.16	67.92	73.87
Pavia U	238.11	385.92	176.01	292.81	17.72	125.41	215.38	140.19	233.12	62.92
Salinas	307.85	297.81	243.32	230.61	8.58	163.50	163.43	125.95	128.94	9.70

4.4. Robustness of the BLRDA Algorithm

In the above subsection, we evaluated the performance of the proposed BLRDA method. Fully excavating the superiority of the proposed method, we analyze the robustness of the BLRDA in this subsection. Taking into consideration practical situations, we analyzed the robustness based on the labeled samples' size and noise.

4.4.1. Robustness to the Size of Labeled Samples

We analyze the impact of different sizes of training and testing sets in this subsection. We perform experiments on these three images, namely the Indian Pines image, the Pavia University scene and the Salinas image. The test was carried out with 10 independent runs for each algorithm. We have calculated the mean value of the results. Figure 9 shows the supervised and semi-supervised classification accuracy of HSIs with different feature extraction methods. The figure compares the overall classification results with different training size samples in each class. The percentage of training samples (in percentage form) grow from 2–14% for the Indian Pines image, 2–8% for the Pavia University scene and 0.2–0.8% for the Salinas image.

In most cases, our proposed BLRDA method consistently achieves the best results, which is robust to the label percentage variations. With the size of the training sample increasing, OA generally increases in all methods that show a similar trend. When the training sample is fixed, the effectiveness of the BLRDA method is usually superior to others, in terms of both the NN and SVM classifier. Similarly, the three classification criteria increase with the number of training samples simultaneously.

It is noteworthy that the method we proposed achieves higher classification accuracy even at very low label rates, while some other compared algorithms are not as robust as our BLRDA algorithm, especially when the label rate is low. Due to the high cost and difficulty of labeling data, our proposed graph for the SDA algorithm is much more robust and suitable for real-world HSIs.

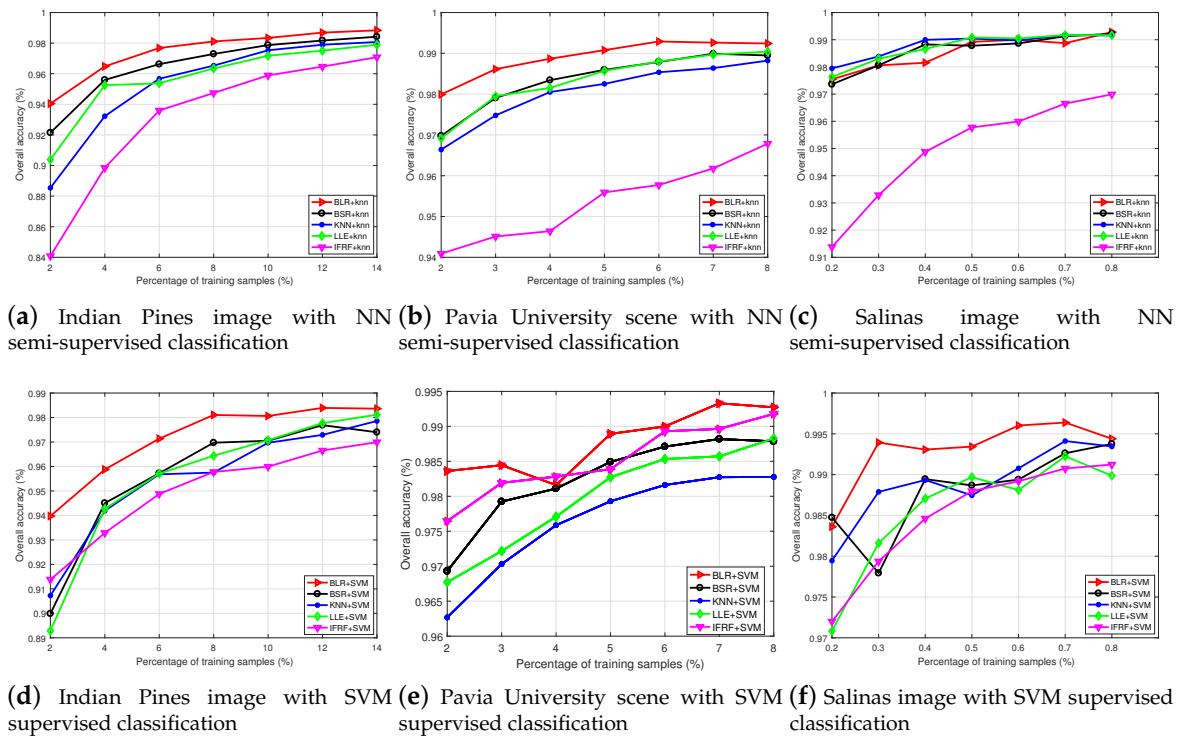


Figure 9. Supervised and semi-supervised classification accuracy of HSIs with different percentages of training samples. LLE, Locally Linear Embedding.

4.4.2. Robustness on Simulated Noisy Hyperspectral Images

In the simulated experiment, we evaluated the noise robustness of the BLRDA method on the three hyperspectral images. Zero-mean Gaussian noise with different variances was added to all bands. For the Indian Pines image and Salinas image, the variance value ranges from 50–250. In the image of the Pavia University scene, the variance of noise varies from 100–500. For different bands, the noise intensity is equal. Figure 10 gives an example of noisy image samples in the Indian Pines image where the band is randomly selected, which is similar to the other two hyperspectral images.

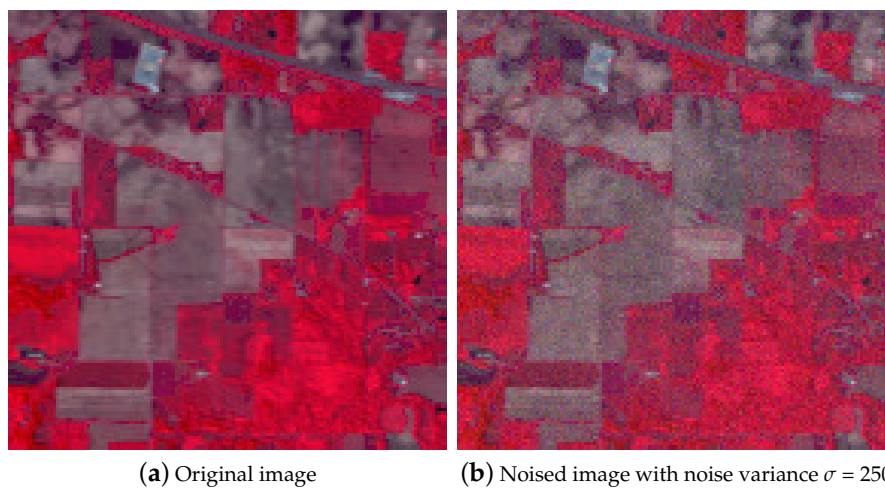


Figure 10. Noised hyperspectral image example: Three-band color composite of noised Indian Pines image.

Compare the performance of different methods in these noisy environments. Ten independent runs of each algorithm have been evaluated, where the training samples in each run are resampled. We have calculated the mean value of the results. The labeled sample rates are 6%, 2% and 0.2%, respectively, as shown in Table 6. Figures 11–13 show the classification results (randomly selected from our above experiments) for the three noisy hyperspectral images associated with the corresponding OA values. It has been observed that the results in the figures are random for each method and do not have comparability. These results reveal the NN classification results of different feature extraction algorithms with some noise variance σ shown in the caption of these figures. We can see that the results of our method are robust to noise and labeled sample size. With few labeled samples, our method BLRDA is more powerful than other methods, which benefits from the robustness of the low-rank representation to noise and the global property of the LR-graph. As the noise gradually increases, the performance of some methods drops substantially. Our method is robust to noise and suffers little performance degradation. Above all, the BLRDA performs very well with respect to all three experimental hyperspectral images.



Figure 11. NN classification results on noisy Indian Pines image (noise variance $\sigma = 250$) obtained by: (a) BLRDA; (b) BSDA; (c) BKDA; (d) BLEDA; (e) IFRF.

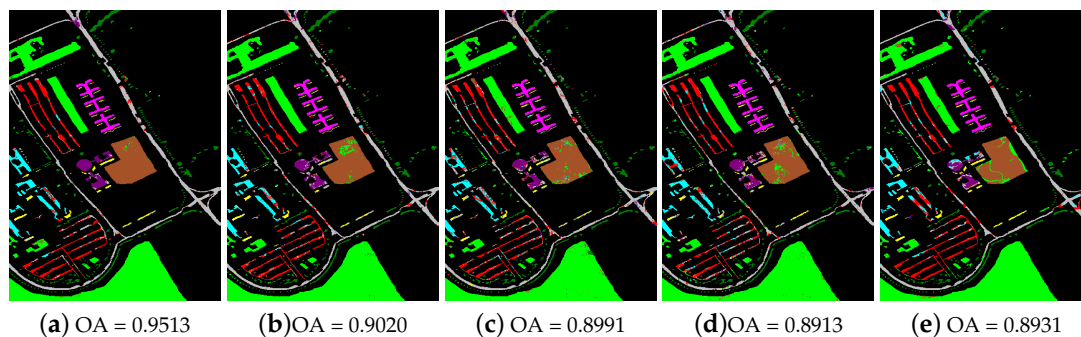


Figure 12. NN classification results on noisy Pavia University scene (noise variance $\sigma = 500$) obtained by: (a) BLRDA; (b) BSDA; (c) BKDA; (d) BLEDA; (e) IFRF.

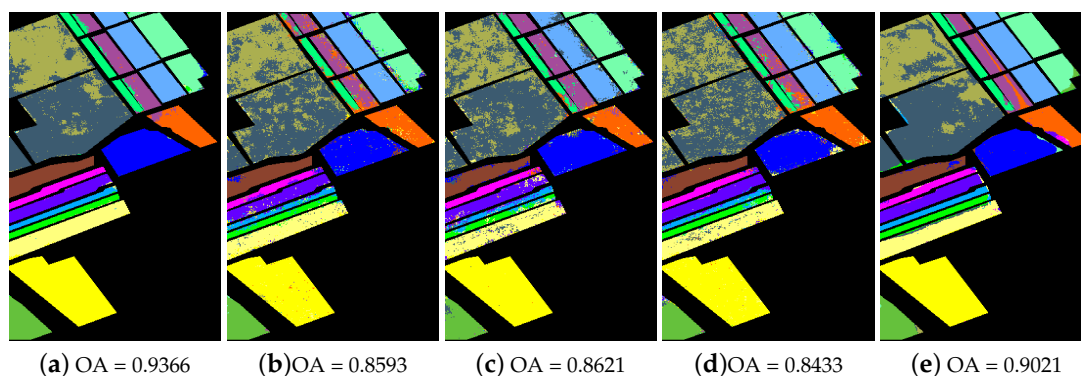


Figure 13. NN classification results on noisy Salinas image (noise variance $\sigma = 500$) obtained by: (a) BLRDA; (b) BSDA; (c) BKDA; (d) BLEDA; (e) IFRF.

Table 6. Overall accuracy with varying variance Gaussian noise on three HSIs.

Hyperspectral Images	σ	BLRDA	BSDA	BKDA	BLEDA	IFRF
Indian Pines image (6%)	50	0.9765	0.9485	0.9412	0.9396	0.9061
	100	0.9672	0.9300	0.9224	0.9192	0.8847
	150	0.9629	0.9005	0.9042	0.8934	0.8667
	200	0.9601	0.8843	0.8904	0.8779	0.8387
	250	0.9481	0.8627	0.8760	0.8462	0.8163
Pavia University scene (2%)	100	0.9734	0.9607	0.9534	0.9668	0.9099
	200	0.9725	0.9626	0.9336	0.9431	0.9018
	300	0.9627	0.9525	0.9261	0.9271	0.9085
	400	0.9583	0.9361	0.9123	0.9130	0.8913
	500	0.9529	0.9279	0.9052	0.9017	0.8952
Salinas image (0.2%)	50	0.9753	0.9642	0.9518	0.9453	0.9238
	100	0.9678	0.9280	0.9198	0.9280	0.9147
	150	0.9495	0.8996	0.8900	0.8970	0.8958
	200	0.9482	0.8726	0.8619	0.8742	0.8899
	250	0.9321	0.8679	0.8503	0.8353	0.8720

4.5. Parameters of the BLRDA Graph Effect

We evaluate the parameters for the BLRDA method. We conduct 10 independent runs per algorithm. We have calculated the mean value of the results. We show the performance of different block sizes and reduction dimensions in Tables 7 and 8. From Table 7, we can see that the block size increases for (25, 50, 75 and 100), and the training set is about 6%, 4% and 0.4% for the Indian Pines image, the Pavia University scene and the Salinas image.

In general, the classification results increase slightly with the growing block size and the reduced dimension. It is observed that the increase in block size simultaneously accelerates the running time considerably. Therefore, we could use the small block size in actual situations for the purpose of efficiency with minimal classification accuracy loss.

From Table 7, we can conclude that the reduced dimension is robust, which will work on a small block size. Overall, our proposed BLRDA method is much more robust and excellent for supervised and semi-supervised classification of HSIs.

Table 7. Classification accuracy of the BLRDA method with different block sizes.

Images	Classifier	NN				SVM			
	Block Size	OA	AA	Kappa	Time	OA	AA	Kappa	Time
Indian Pines image	25	0.9784	0.9673	0.9754	28.06	0.9721	0.9831	0.9681	64.28
	50	0.9757	0.9699	0.9712	34.01	0.9728	0.9836	0.9713	69.96
	75	0.9762	0.9676	0.9728	40.88	0.9754	0.9791	0.9719	76.70
	100	0.9776	0.9725	0.9744	46.39	0.9782	0.9819	0.9751	82.40
University of Pavia image	25	0.9884	0.9814	0.9846	212.05	0.9833	0.9699	0.9779	295.10
	50	0.9887	0.9809	0.9850	238.11	0.9828	0.9687	0.9772	315.92
	75	0.9877	0.9817	0.9837	258.57	0.9847	0.9736	0.9796	339.07
	100	0.9863	0.9777	0.9818	277.80	0.9854	0.9734	0.9807	378.05
Salinas image	25	0.9963	0.9958	0.9960	315.90	0.9917	0.9922	0.9907	283.85
	50	0.9979	0.9967	0.9976	340.80	0.9930	0.9924	0.9923	315.72
	75	0.9981	0.9973	0.9979	383.07	0.9930	0.9917	0.9922	353.74
	100	0.9985	0.9978	0.9984	412.47	0.9941	0.9938	0.9934	383.43

Table 8. Classification accuracy of the BLRDA method with different reduced dimensions.

Images	Classifier	NN			SVM		
	Dimension	OA	AA	Kappa	OA	AA	Kappa
Indian Pines image	12	0.9976	0.9965	0.9974	0.9720	0.9743	0.9680
	14	0.9977	0.9966	0.9974	0.9702	0.9750	0.9660
	16	0.9980	0.9968	0.9978	0.9711	0.9772	0.9670
	18	0.9980	0.9968	0.9978	0.9688	0.9750	0.9644
	20	0.9980	0.9968	0.9978	0.9705	0.9755	0.9663
University of Pavia image	12	0.9883	0.9800	0.9845	0.9827	0.9690	0.9771
	14	0.9885	0.9808	0.9848	0.9823	0.9678	0.9765
	16	0.9887	0.9810	0.9850	0.9824	0.9686	0.9767
	18	0.9888	0.9812	0.9851	0.9832	0.9698	0.9777
	20	0.9890	0.9816	0.9854	0.9835	0.9683	0.9781
Salinas image	12	0.9981	0.9967	0.9979	0.9921	0.9923	0.9909
	14	0.9981	0.9969	0.9979	0.9931	0.9918	0.9920
	16	0.9983	0.9971	0.9981	0.9919	0.9925	0.9907
	18	0.9983	0.9970	0.9981	0.9954	0.9950	0.9946
	20	0.9983	0.9970	0.9982	0.9960	0.9954	0.9953

5. Discussion

Classification of HSIs plays a pivotal role in understanding the HSIs. In the present work, we propose a novel approach for HSI feature extraction, Regularized Block Low-rank Discriminant Analysis (BLRDA). Our goal is to enhance the classification accuracy of HSIs by the useful feature extraction method. Experimental results on the three images show that the BLRDA is a competitive feature extraction method with other comparative methods for HSI classification.

From the supervised and semi-supervised experiments illustrated in Tables 2–4, we observe that the BLRDA method is an effective feature extraction method, which achieves the highest classification accuracy compared to the other methods. The performance was remarkable even with simple supervised and semi-supervised classifiers (nearest neighbor and SVM) and randomly given parameters. In some case, traditional construct graph methods (such as k NN-graph and LLE-graph) may perform well in some classes, but they are not as stable as our proposed algorithm. The LR-graph captures the global data structures better and obtains the representation of all samples under global low-rank constraints, while the most common k -Nearest Neighbor (k NN) and Locally Linear Embedding (LLE) neighbors use fixed global parameters to determine the weights of the graph with Euclidean distances whose graphical structures are associated with noise instability and are sensitive to noise. Additionally, the SR-graph lacks global constraints, which greatly degrades performance when the data are grossly corrupted. However, the LR-graph addresses these drawbacks, which jointly receives the graph of the whole data and is demonstrated to be robust to the noise. Further, the k -nearest neighbor is used to integrally handle the combined low-rank graph, which performs two functions: preserving the local information of the image and further satisfying the algorithmic requirements for the subsequent dimension reduction procedure. Consequently, it significantly improves the classification performance of hyperspectral images, which indicates that the BLRDA method is a superior HSI feature graph both for supervised classification and for semi-supervised classification.

We analyzed robustness based on the labeled samples' size effect and noise effect. In Figure 9, we discover that the BLRDA method is usually superior to the others in terms of both the NN and SVM classifier, which is robust to the label percentage variations. Figure 9 also shows that the proposed method achieves higher classification accuracy even at meager label rates. Some of the other compared algorithms are not as robust as our BLRDA algorithm, in particular when the label rate is low. Due to the lack of labeled samples, our proposed method is much more robust and suitable for real-world HSI classification. As we can see in Table 6 and Figures 11–13, the results of our method are robust to noise and labeled sample size. As the noise increases, the performance of BLRDA drops very slowly,

from 0.9765–0.9481 in the case of the India Pines image, which has little effect in practical situations. However, for the comparative methods BSDA, BKDA, BLEDA and IFRF, overall accuracy drops quickly with the increase in noise. In addition to the India Pines image, we can see the same results from these three HSI results. This benefits from the robustness of the low-rank representation to noise, while k -Nearest Neighbor (k NN) and Locally Linear Embedding (LLE) neighbors use fixed global parameters with Euclidean distances, which are sensitive to noise. The LR-graph better captures the global property of the data. Therefore, the BLRDA method is demonstrated to be robust to the labeled samples' size and noise, which indicates that it is highly competitive in practical situations.

We evaluate the performance of parameters with different block sizes and reduction dimensions for the BLRDA method, as shown in Tables 7 and 8. From the tables, we can see that the classification results increase slightly with the growing block size and reduced dimension. Therefore, we could use small block sizes in actual situations for the purpose of efficiency, with classification accuracy loss so minimal that it could be ignored. Moreover, from Table 8, we can conclude that the reduced dimension is robust in a medium-sized block.

The proposed BLRDA method performs quite well and is competitive with respect to supervised and semi-supervised classification of HSIs. In future work, we plan to explore further real-world applications and improve the BLRDA method, thereby making it more efficient. Inspired by one reviewer, in the future, we could divide the image into blocks by a proper community detection algorithm [50–53] instead of dividing the image by the regular pixels' position, which is very great advice. The way of dividing blocks may impact the subsequent low-rank representation.

6. Conclusions

In this paper, we presented a novel graph for HSI feature extraction, which is referred to as Regularized Block Low-rank Discriminant Analysis (BLRDA). To reduce computational complexity, the entire image is divided into blocks, and the low-rank representation for each group is implemented separately. The global structure of the hyperspectral image can be captured by low-rank representation. Additionally, the local geometrical structure is preserved by the k -nearest neighbor algorithm. Therefore, the performance of image classification has been enhanced by the BLRDA method. Experiments on several real multi-class hyperspectral images indicate that our proposed BLRDA method is an efficient and robust feature extraction method for both supervised and semi-supervised classifiers.

Author Contributions: B.Z. conceived of and designed the experiments and wrote the paper. W.D. performed some experiments and analyzed the data. K.X. gave some comments and idea for the work. Y.L. and A.A. contributed to the review and revision. S.C. and A.A. carefully edit the language to improve the writing quality.

Funding: This research was funded by [Hebei Province Natural Science Foundation] grant number [E2016202341] and [Research Project of Science and Technology for Hebei Province Higher Education Institutions] grant number [BJ2014013].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xu, L.; Zhang, H.; Zhao, M.; Chu, D.; Li, Y. Integrating Spectral and Spatial Features for Hyperspectral Image Classification Using Low-Rank Representation. In Proceedings of the IEEE International Conference on Industrial Technology (ICIT), Toronto, ON, Canada, 22–25 March 2017; pp. 1024–1029.
2. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352. [[CrossRef](#)]
3. Dorigo, W.A.; Zurita-Milla, R.; de Wit, A.J.; Brazile, J.; Singh, R.; Schaepman, M.E. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *Int. J. Appl. Earth Obs. Geoinform.* **2007**, *9*, 165–193. [[CrossRef](#)]

4. Dalponte, M.; Bruzzone, L.; Vescovo, L.; Gianelle, D. The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sens. Environ.* **2009**, *113*, 2345–2355. [[CrossRef](#)]
5. Ustin, S.L.; Roberts, D.A.; Gamon, J.A.; Asner, G.P.; Green, R.O. Using imaging spectroscopy to study ecosystem processes and properties. *AIBS Bull.* **2004**, *54*, 523–534. [[CrossRef](#)]
6. Heinz, D.C. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 529–545. [[CrossRef](#)]
7. Kruse, F.A.; Boardman, J.W.; Huntington, J.F. Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1388–1400. [[CrossRef](#)]
8. Cocks, T.; Jenssen, R.; Stewart, A.; Wilson, I.; Shields, T. The HyMap™ Airborne Hyperspectral Sensor: The System, Calibration and Performance. In Proceedings of the 1st EARSeL workshop on Imaging Spectroscopy, Zurich, Switzerland, 6–8 October 1998; pp. 37–42.
9. De Morsier, F.; Borgeaud, M.; Gass, V.; Thiran, J.P.; Tuia, D. Kernel low-rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3410–3420. [[CrossRef](#)]
10. Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [[CrossRef](#)]
11. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [[CrossRef](#)] [[PubMed](#)]
12. Tenenbaum, J.B.; De Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323. [[CrossRef](#)] [[PubMed](#)]
13. Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **2003**, *15*, 1373–1396. [[CrossRef](#)]
14. Ul Haq, Q.S.; Tao, L.; Sun, F.; Yang, S. A fast and robust sparse approach for hyperspectral data classification using a few labeled samples. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2287–2302. [[CrossRef](#)]
15. Kuo, B.C.; Chang, K.Y. Feature extractions for small sample size classification problem. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 756–764. [[CrossRef](#)]
16. Gu, Y.; Wang, Q.; Wang, H.; You, D.; Zhang, Y. Multiple kernel learning via low-rank nonnegative matrix factorization for classification of hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2739–2751. [[CrossRef](#)]
17. Kang, X.; Li, S.; Benediktsson, J.A. Feature extraction of hyperspectral images with image fusion and recursive filtering. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3742–3752. [[CrossRef](#)]
18. Zhou, D.; Bousquet, O.; Lal, T.N.; Weston, J.; Schölkopf, B. Learning with local and global consistency. In *Advances in Neural Information Processing Systems*; MIT Press: Vancouver, BC, Canada; 2004; pp.321–328.
19. Guo, Z.; Zhang, Z.; Xing, E.P.; Faloutsos, C. Semi-Supervised Learning based on Semiparametric Regularization. In Proceedings of the SIAM International Conference on Data Mining, Atlanta, Georgia, 24–26 April 2008; pp. 132–142.
20. Subramanya, A.; Talukdar, P.P. Graph-based semi-supervised learning. *Synth. Lect. Artif. Intell. Mach. Learn.* **2014**, *8*, 1–125. [[CrossRef](#)]
21. Cai, D.; He, X.; Han, J. Semi-Supervised Discriminant Analysis. In Proceedings of the IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–20 October 2007; pp. 1–7.
22. Jebara, T.; Wang, J.; Chang, S.F. Graph Construction and b-Matching for Semi-Supervised Learning. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 441–448.
23. Yan, S.; Wang, H. Semi-supervised Learning by Sparse Representation. In Proceedings of the SIAM International Conference on Data Mining, Sparks, NV, USA, 30 April–2 May 2009; pp. 792–801.
24. Zhu, X.; Lafferty, J.; Rosenfeld, R. Semi-Supervised Learning with Graphs. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2005.
25. Zhu, X.; Ghahramani, Z.; Lafferty, J.D. Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003; pp. 912–919.
26. Belkin, M.; Matveeva, I.; Niyogi, P. Regularization and semi-supervised learning on large graphs. In *International Conference on Computational Learning Theory*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 624–638.

27. Zhuang, L.; Gao, H.; Lin, Z.; Ma, Y.; Zhang, X.; Yu, N. Non-Negative Low Rank and Sparse Graph for Semi-Supervised Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 2328–2335.
28. Zhuang, L.; Gao, S.; Tang, J.; Wang, J.; Lin, Z.; Ma, Y.; Yu, N. Constructing a nonnegative low-rank and sparse graph with data-adaptive features. *IEEE Trans. Image Proc.* **2015**, *24*, 3717–3728. [[CrossRef](#)] [[PubMed](#)]
29. Liu, G.; Lin, Z.; Yu, Y. Robust Subspace Segmentation by Low-Rank Representation. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–25 June 2010; pp. 663–670.
30. Li, W.; Liu, J.; Du, Q. Sparse and Low-Rank Graph for Discriminant Analysis of Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4094–4105. [[CrossRef](#)]
31. Rohban, M.H.; Rabiee, H.R. Supervised neighborhood graph construction for semi-supervised classification. *Patt. Recognit.* **2012**, *45*, 1363–1372. [[CrossRef](#)]
32. Zhang, H.; He, W.; Zhang, L.; Shen, H.; Yuan, Q. Hyperspectral image restoration using low-rank matrix recovery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4729–4743. [[CrossRef](#)]
33. Veganzones, M.A.; Simoes, M.; Licciardi, G.; Yokoya, N.; Bioucas-Dias, J.M.; Chanussot, J. Hyperspectral super-resolution of locally low-rank images from complementary multisource data. *IEEE Trans. Image Proc.* **2016**, *25*, 274–288. [[CrossRef](#)] [[PubMed](#)]
34. Niu, Y.; Wang, B. Hyperspectral Anomaly Detection Based on Low-Rank Representation and Learned Dictionary. *Remote Sens.* **2016**, *8*, 289. [[CrossRef](#)]
35. Pan, L.; Li, H.C.; Deng, Y.J.; Zhang, F.; Chen, X.D.; Du, Q. Hyperspectral Dimensionality Reduction by Tensor Sparse and Low-Rank Graph-Based Discriminant Analysis. *Remote Sens.* **2017**, *9*, 452. [[CrossRef](#)]
36. He, W.; Zhang, H.; Zhang, L.; Shen, H. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 178–188. [[CrossRef](#)]
37. Qi, K.; Liu, W.; Yang, C.; Guan, Q.; Wu, H. Multi-Task Joint Sparse and Low-Rank Representation for the Scene Classification of High-Resolution Remote Sensing Image. *Remote Sens.* **2016**, *9*, 10. [[CrossRef](#)]
38. Zu, B.; Xia, K.; Pan, Y.; Niu, W. A Novel Graph Constructor for Semisupervised Discriminant Analysis: Combined Low-Rank and-Nearest Neighbor Graph. *Comput. Intell. Neurosci.* **2017**, *2017*. [[CrossRef](#)] [[PubMed](#)]
39. Cortes, C.; Mohri, M. On transductive regression. *Adv. Neural Inf. Proc. Syst.* **2007**, *19*, 305.
40. Zhang, L.; Wei, W.; Zhang, Y.; Shen, C.; van den Hengel, A.; Shi, Q. Cluster sparsity field for hyperspectral imagery denoising. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 631–647.
41. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM (JACM)* **2011**, *58*, 11. [[CrossRef](#)]
42. Cai, J.F.; Candès, E.J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Opt.* **2010**, *20*, 1956–1982. [[CrossRef](#)]
43. Lin, Z.; Chen, M.; Ma, Y. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv* **2010**, arXiv:1009.5055. [[CrossRef](#)]
44. Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; Ma, Y. Robust recovery of subspace structures by low-rank representation. *Patt. Anal. Mach. Intell. IEEE Trans.* **2013**, *35*, 171–184. [[CrossRef](#)] [[PubMed](#)]
45. Gastal, E.S.; Oliveira, M.M. Domain transform for edge-aware image and video processing. In *ACM Transactions on Graphics (ToG)*; ACM: New York, NY, USA, 2011; Volume 30, p. 69.
46. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [[CrossRef](#)]
47. Richards, J.A.; Richards, J. *Remote Sensing Digital Image Analysis*; Springer: Berlin/Heidelberg, Germany, 1999; Volume 3.
48. Thompson, W.D.; Walter, S.D. A reappraisal of the kappa coefficient. *J. Clin. Epidemiol.* **1988**, *41*, 949–958. [[CrossRef](#)]
49. Gwet, K. Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Stat. Methods Int. Reliab. Assess.* **2002**, *1*, 1–6.
50. Javed, M.A.; Younis, M.S.; Latif, S.; Qadir, J.; Baig, A. Community detection in networks: A multidisciplinary review. *J. Netw. Comput. Appl.* **2018**. [[CrossRef](#)]

51. Agreste, S.; De Meo, P.; Fiumara, G.; Piccione, G.; Piccolo, S.; Rosaci, D.; Sarné, G.M.; Vasilakos, A.V. An empirical comparison of algorithms to find communities in directed graphs and their application in web data analytics. *IEEE Trans. Big Data* **2017**, *3*, 289–306. [[CrossRef](#)]
52. He, K.; Li, Y.; Soundarajan, S.; Hopcroft, J.E. Hidden community detection in social networks. *Inf. Sci. Int. J.* **2017**, *425*, 92–106. [[CrossRef](#)]
53. Nathan, E.; Zakrzewska, A.; Riedy, J.; Bader, D.A. Local Community Detection in Dynamic Graphs Using Personalized Centrality. *Algorithms* **2017**, *10*, 102. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).