



Article

Latent Low-Rank Projection Learning with Graph Regularization for Feature Extraction of Hyperspectral Images

Lei Pan ¹, Hengchao Li ^{2,*}, Xiang Dai ¹, Ying Cui ¹, Xifeng Huang ¹ and Lican Dai ¹

¹ Southwest Institute of Electronic Technology, Chengdu 610036, China; panlei@my.swjtu.edu.cn (L.P.); wwy@cuit.edu.cn (X.D.); lizc@my.swjtu.edu.cn (Y.C.); wshu@my.swjtu.edu.cn (X.H.); licand@mail.ustc.edu.cn (L.D.)

² School of Information Science & Technology, Southwest Jiaotong University, Chengdu 610031, China

* Correspondence: hcli@home.swjtu.edu.cn; Tel.: +86-139-8212-3202

Abstract: Due to the great benefit of rich spectral information, hyperspectral images (HSIs) have been successfully applied in many fields. However, some problems of concern also limit their further applications, such as high dimension and expensive labeling. To address these issues, an unsupervised latent low-rank projection learning with graph regularization (LatLRPL) method is presented for feature extraction and classification of HSIs in this paper, in which discriminative features can be extracted from the view of latent space by decomposing the latent low-rank matrix into two different matrices, also benefiting from the preservation of intrinsic subspace structures by the graph regularization. Different from the graph embedding-based methods that need two phases to obtain the low-dimensional projections, one step is enough for LatLRPL by constructing the integrated projection learning model, reducing the complexity and simultaneously improving the robustness. To improve the performance, a simple but effective strategy is exploited by conducting the local weighted average on the pixels in a sliding window for HSIs. Experiments on the Indian Pines and Pavia University datasets demonstrate the superiority of the proposed LatLRPL method.



Citation: Pan, L.; Li, H.; Dai, X.; Cui, Y.; Huang, X.; Dai, L. Latent Low-Rank Projection Learning with Graph Regularization for Feature Extraction of Hyperspectral Images. *Remote Sens.* **2022**, *14*, 3078. <https://doi.org/10.3390/rs14133078>

Academic Editor: Fabio Del Frate

Received: 8 May 2022

Accepted: 14 June 2022

Published: 27 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: feature extraction; hyperspectral image; latent low-rank representation; graph regularization; projection learning

1. Introduction

Depending on hundreds of spectral bands captured by the hyperspectral imaging systems [1], hyperspectral images (HSIs) can better present the substance compositions of regions of interest from the view of pixel-level representations. Accordingly, lots of fields have made attempts to apply HSIs, such as agricultural monitoring, urban planning, and military reconnaissance, showing promising results. One problem that has to be faced directly when enjoying the advantages of HSIs is the high spectral dimensionality [2–4], which usually brings much more computational cost and degrades the performance to some extent. Feature extraction [5–8] can effectively solve this problem by reducing the dimensionality of HSIs and simultaneously preserving the significant information. Due to the great success in feature extraction and classification of natural images [9,10], the deep learning-based methods have attracted lots of attention [11–13]. Some researchers also attempted to utilize deep learning methods to process hyperspectral images, showing superiority in many tasks [14,15]. In this paper, the traditional machine learning-based methods are of concern.

Among all the feature extraction methods, principal component analysis (PCA) [16] is the most popular and also an effective unsupervised method to reduce the dimensionality of the data by maximizing the variance in low-dimensional space. Due to the lack of prior knowledge of samples, PCA may not perform very well. To overcome this problem, some supervised feature extraction methods have been proposed. For example, linear

discriminant analysis (LDA) [17] was developed by maximizing the trace ratio of between-class and within-class scatter matrices, which is improved by local Fisher's discriminant analysis (LFDA) [18].

To make full use of the intrinsic structure expressed implicitly in the data, some manifold learning-based methods [19,20] have been proposed. Locality preserving projection (LPP) [21] was designed to preserve the similar local relationships among samples from the original high-dimensional space to the low-dimensional space, which evolved from locally linear embedding (LLE) [22]. Neighborhood preserving embedding (NPE) [23] can capture the intrinsic subspace structures by minimizing the reconstructed error between one sample and its neighbors, which is extended from Laplacian eigenmaps (LE) [24].

A unified framework called graph embedding was proposed in [25] by constructing two graphs, which takes the aforementioned methods into account. In this framework, the graph weight matrix plays the most important role in the quality of the low-dimensional features. Traditional methods, such as k -nearest neighbor and ε -radius ball, are usually influenced by noise. Due to the good robustness and promising performance, sparse representation (SR) [26] has attracted increasing attention for image classification [27] and target detection [28] since it was proposed for the first time. Therefore, some researchers employ the SR coefficient to construct the weight matrix, generating many classical feature extraction methods for HSIs. For example, sparse graph embedding (SGE) [29] was proposed to construct the sparse graph for feature extraction by solving a l_1 optimization problem. In [30], sparsity preserving projection (SPP) was proposed by constructing the adjacency weight from a modified SR method. Subsequently, sparse graph-based discriminant analysis (SGDA) [31] was developed by exploiting the prior label information, which is further extended to block SGDA by constructing a within-class graph. Based on SGDA, He et al. [32] pointed out that the locality constraint could improve the discriminative ability of a sparse coefficient and constructed the weighted SGDA model. Due to the great computational burden brought about by the optimization of the l_1 norm, Ly et al. [33] replaced the l_1 norm with the l_2 regularization by borrowing the idea from [34], improving the computational efficiency.

SR-based graph embedding methods capture the intrinsic structure of the data from the perspective of locality, which, however, ignores the positive effect from the view of the global information. In [35], Liu et al. proposed a low-rank representation (LRR) model to explore the global structure information of the data and extended it to a low-rank graph embedding model for unsupervised feature extraction. Commonly, the data structure can be presented by exploiting both local and global information [36,37]. Accordingly, one method presented in [38] takes both sparse regularization and low-rank regularization into consideration to construct the low-rank and sparse graph for the extraction of discriminative low-dimensional features. Following SLGDA, some distinguished works are proposed. For instance, to deal with the nonlinear property of the hyperspectral data, one feature extraction method named classical kernel SLGDA [39] was proposed for HSIs to advance the computational efficiency by accelerating the process of kernel transformation with a virtual kernel technique. In [40], to fully exploit the spatial information of the hyperspectral data, the tensor theory was introduced into the work [38], resulting in an effective spatial-spectral feature extraction method that can learn the local and global structures of the hyperspectral data.

Graph embedding-based feature extraction methods can effectively extract low-dimensional features from the hyperspectral data; however, there are still some challenges that require conquering. First, the key point of these methods is the construction of adjacent graphs that represent the intrinsic subspace structure. The aforementioned methods addressed this problem using KNN, SR, and LRR, which cannot present satisfying results. Second, graph embedding-based methods acquire low-dimensional features through two disconnected steps, i.e., graph construction and projection optimization, which loses some valuable and discriminative information.

In the process of hyperspectral imaging, various kinds of noise will have a negative influence on the quality of HSIs [41–43]. In other words, some pixels can be grossly corrupted. In this case, both SR and LRR are not competent at preserving the intrinsic subspace structure of the hyperspectral data. To address this issue, Liu et al. [44] proposed a latent low-rank representation (LatLRR) model by considering the data structure from two views, i.e., column space and row space. When the data are corrupted in the column space, it can also be recovered from the row space. Depending on this advantage of LatLRR, Pan et al. [45] proposed two data recovery models from the spectral domain and spatial domain for the hyperspectral data. Another big difference between LRR and LatLRR is that LatLRR is able to extract salient features of the data without graph construction. This inborn advantage shows great significance in face recognition, which can characterize some salient and discriminative features of one's face, such as eyes and noses. However, the dimension of the extracted salient features is the same as that of the original data [46]. Strictly speaking, it is actually not a common feature extraction method. In addition, LatLRR still limits its view on the preservation of the global structure of the data.

To overcome these problems, a novel latent low-rank projection learning with graph regularization (LatLRPL) method is proposed for unsupervised feature extraction of HSIs in this paper. In LatLRPL, two different representation matrices with the same dimension are adopted to decompose the salient projection matrix in LatLRR, forming a discriminative projection matrix with a controllable feature dimension. To capture the local intrinsic structure of hyperspectral data, both original space and low-dimensional space are considered. On the one hand, a spectral constraint matrix from the original space is constructed to enhance the discriminative ability of corresponding subspaces. On the other hand, manifold learning on the learned low-dimensional space is exploited as the graph regularization to preserve the local subspace structures. Experiments on two common hyperspectral datasets show the superiority of the proposed LatLRPL method. The main contributions of this paper can be summarized as follows.

- (1) Compared with LatLRR, LatLRPL is able to learn discriminative features with the desired dimensionality for the hyperspectral data, accomplishing the dimensionality reduction in the real sense. In addition, local structure and global structure can be well preserved in the learned low-dimensional feature space.
- (2) Different from graph embedding-based methods that require two steps to obtain the projection, LatLRPL integrates representation learning and projection learning into one framework, thereby improving the robustness and computational efficiency.
- (3) To better preserve the local intrinsic subspace structures, regularization on both the original space and the learned low-dimensional space is integrated into LatLRPL by using spectral constraint and graph regularization, respectively.

The remainder of this paper is organized as follows. Section 2 introduces some related work. The framework and optimization of the proposed LatLRPL method are elaborated in Section 3. In Section 4, experimental results and analysis are provided. Section 5 provides the conclusions for this paper.

2. Related Work

In this section, some classical work about feature extraction, such as low-rank representation, latent low-rank representation, and locality preserving projection, has been briefly reviewed in regard to their motivations and mathematic theories, which is referential for the design of the proposed method. To formulate the related work clearly, definitions about some variables are first provided. Given an HSI $\mathbf{H} \in \mathbb{R}^{H \times W \times D}$ with D spectral bands, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$ is defined as a sample set containing N samples, in which $\mathbf{x}_i \in \mathbb{R}^{D \times 1}$ denotes the i th sample.

2.1. Low-Rank Representation for Feature Extraction

Sparse representation (SR) attempts to represent one sample by using a few labeled samples in the dictionary. Although it has achieved promising performance in many

fields, SR can not present the global data structure. As such, Liu et al. [35] first proposed the low-rank representation (LRR) by representing one sample set jointly, which can be formulated as

$$\min_{\mathbf{Z}, \mathbf{E}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} \quad s.t. \mathbf{X} = \mathbf{XZ} + \mathbf{E} \quad (1)$$

where $\|\cdot\|_*$ denotes the nuclear norm used to approximate the rank of one matrix, $\|\mathbf{E}\|_{2,1}$ is the $l_{2,1}$ norm used to represent sample-specific corruptions, defined as $\|\mathbf{E}\|_{2,1} = \sum_{j=1}^N \sqrt{\sum_{i=1}^d (\mathbf{E}_{ij})^2}$, $\mathbf{Z}$ denotes the lowest rank reconstructed coefficient matrix with the global structure information captured, λ is the regularization parameter. Similar to sparse graph embedding [29], the low-rank coefficient matrix can be used to construct the low-rank graph for feature extraction by employing the graph embedding framework.

2.2. Latent Low-Rank Representation for Feature Extraction

In practical applications, the observed data are usually contaminated by various kinds of noise. More seriously, some samples are corrupted badly. LRR shows good performance in the preservation of global data structure; however, it only considers this from the column space. When the data are deadly corrupted, LRR does not perform effectively. To overcome this shortcoming, Liu et al. [44] proposed a latent low-rank representation (LatLRR) model by taking both the column space and the row space of the data into account. Experimental results indicate that LatLRR is very effective when coping with corrupted data. The objective function of LatLRR is defined as

$$\min_{\mathbf{Z}, \mathbf{L}, \mathbf{E}} \|\mathbf{Z}\|_* + \|\mathbf{L}\|_* + \lambda \|\mathbf{E}\|_{2,1}, \quad (2)$$

$$s.t. \mathbf{X} = \mathbf{XZ} + \mathbf{LX} + \mathbf{E},$$

where $\mathbf{Z} \in \mathbb{R}^{N \times N}$ is the column coefficient matrix that reconstructs the data from the column space, and $\mathbf{L} \in \mathbb{R}^{D \times D}$ is the row coefficient matrix that reconstructs the data from the row space. This is the reason why LatLRR performs better than LRR for the corrupted data. Additionally, $\mathbf{L}$ can also be exploited as the projection matrix to extract the salient feature from the data, which has been deeply researched in face recognition. However, from the objective function (2), it can be seen that the salient feature $\mathbf{LX}$ has the same dimension as the original data $\mathbf{X}$. This is not preferred in the case of low-dimensional feature extraction.

2.3. Locality Preserving Projection for Feature Extraction

Locality preserving projection (LPP), as one popular manifold learning method, shows its effectiveness in capturing the local geometric data structure for feature extraction. The nature of LPP is to preserve the local relationship in the original space to the low-dimensional space. For this purpose, a neighbor graph is first constructed by searching the nearest neighbors of one sample. Specifically, the weight factors can be defined as follows.

$$\mathbf{W}_{ij} = \begin{cases} 1, & \text{if } \mathbf{x}_i \in N_k(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_k(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $N_k(\mathbf{x}_j)$ is a set of k nearest neighbors of $\mathbf{x}_i$. When $\mathbf{x}_j$ belongs to one of the nearest neighbors of $\mathbf{x}_i$, or $\mathbf{x}_i$ is in the nearest neighbor set of $\mathbf{x}_j$, the weight value between $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined as 1. Otherwise, 0 is given. The graph weight matrix represents the local

relationships of samples in the original space. Then, based on the idea of LPP, its objective function is designed as

$$\begin{aligned} & \min_{\mathbf{P}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{P}} \sum_{i,j=1}^N \|\mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \mathbf{x}_j\|_2^2 \mathbf{W}_{ij} \\ & = \min_{Tr(\mathbf{P}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{P}=1)} Tr(\mathbf{P}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{P}) \end{aligned} \quad (4)$$

where $\mathbf{D}$ denotes a diagonal matrix with its i th element defined as $\mathbf{D}_{ii} = \sum_{j=1}^N \mathbf{W}_{ij}$, $\mathbf{L}$ denotes the Laplacian matrix with $\mathbf{L} = \mathbf{D} - \mathbf{W}$, and $\mathbf{P}$ is the projection matrix that can preserve the local structure information.

3. Proposed Unsupervised Projection Learning Method

In order to better capture the local and global structure of the hyperspectral data simultaneously, as well as reduce the redundancy of graph embedding-based methods, a novel latent low-rank projection learning with graph regularization (LatLRPL) method is proposed for the feature extraction of HSIs in this section. The model construction and optimization of LatLRPL are formulated explicitly in the following sections.

3.1. Model Construction of LatLRPL

Advantages of LatLRR in processing corrupted data and extracting salient features make it very popular in face recognition, which, however, fails to reduce the dimensionality of the extracted salient feature. To solve this problem, two different matrixes are adopted to decompose the low-rank projection matrix, i.e., $\mathbf{L} = \mathbf{Q}\mathbf{P}^T$, with $\mathbf{Q} \in \mathbb{R}^{D \times d}$ and $\mathbf{P} \in \mathbb{R}^{D \times d}$, and d denotes the number of preserved features. In this paper, $\mathbf{P}$ is exploited as the projection matrix to obtain the low-dimensional salient feature. As such, the model of LatLRR can be transformed as follows.

$$\begin{aligned} & \min_{\mathbf{Z}, \mathbf{P}, \mathbf{Q}, \mathbf{E}} \|\mathbf{Z}\|_* + \|\mathbf{Q}\mathbf{P}^T\|_* + \lambda \|\mathbf{E}\|_{2,1}, \\ & s.t. \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{Q}\mathbf{P}^T\mathbf{X} + \mathbf{E}. \end{aligned} \quad (5)$$

From Equation (5), we can observe that it is awfully difficult to obtain the solutions of $\mathbf{Q}$ and $\mathbf{P}$. Therefore, the low-rank constraint on $\mathbf{Q}\mathbf{P}^T$ by the nuclear norm is relaxed through the Frobenius norm multiplied by a regularization parameter. Correspondingly, Equation (5) can be changed as

$$\begin{aligned} & \min_{\mathbf{Z}, \mathbf{P}, \mathbf{Q}, \mathbf{E}} \|\mathbf{Z}\|_* + \frac{\beta}{2} \|\mathbf{Q}\mathbf{P}^T\|_F^2 + \lambda \|\mathbf{E}\|_{2,1}, \\ & s.t. \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{Q}\mathbf{P}^T\mathbf{X} + \mathbf{E}, \mathbf{Q}^T\mathbf{Q} = \mathbf{I}, \end{aligned} \quad (6)$$

where β is the regularization parameter with a value much less than 1, which can constrain $\|\mathbf{Q}\mathbf{P}^T\|_F^2$ to be much more approximate to $\|\mathbf{Q}\mathbf{P}^T\|_*$ and simultaneously balance the effect of $\|\mathbf{Q}\mathbf{P}^T\|_F^2$ in the objective function. To avoid the trivial solution, an orthogonal constraint is imposed on $\mathbf{Q}$.

For simplification, $\|\mathbf{Q}\mathbf{P}^T\|_F^2$ can be further manipulated as the following.

$$\begin{aligned} \|\mathbf{Q}\mathbf{P}^T\|_F^2 &= Tr((\mathbf{Q}\mathbf{P}^T)^T(\mathbf{Q}\mathbf{P}^T)) \\ &= Tr(\mathbf{P}\mathbf{Q}^T\mathbf{Q}\mathbf{P}^T) \\ &= Tr(\mathbf{P}\mathbf{P}^T) \\ &= Tr(\mathbf{P}^T\mathbf{P}) \\ &= \|\mathbf{P}\|_F^2 \end{aligned} \quad (7)$$

Based on Equation (7), problem (6) is reformulated as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{P}, \mathbf{Q}, \mathbf{E}} \quad & \|\mathbf{Z}\|_* + \frac{\beta}{2} \|\mathbf{P}\|_F^2 + \lambda \|\mathbf{E}\|_1, \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{XZ} + \mathbf{QP}^T \mathbf{X} + \mathbf{E}, \mathbf{Q}^T \mathbf{Q} = \mathbf{I}, \end{aligned} \quad (8)$$

where $\|\mathbf{E}\|_1$ is used to represent sparse random noise by considering the characteristics of the hyperspectral data.

As mentioned above, the low-rank constraint is helpful for capturing the global intrinsic structure, while it fails to reveal the local structure of the data. According to our previous work [47], the locality constraint is applied to advance the representation of the local structure by exploiting the spectral similarity in the hyperspectral data. Based on this, problem (8) can be rewritten as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{P}, \mathbf{Q}, \mathbf{E}} \quad & \|\mathbf{C} \odot \mathbf{Z}\|_* + \frac{\beta}{2} \|\mathbf{P}\|_F^2 + \lambda \|\mathbf{E}\|_1, \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{XZ} + \mathbf{QP}^T \mathbf{X} + \mathbf{E}, \mathbf{Q}^T \mathbf{Q} = \mathbf{I}, \end{aligned} \quad (9)$$

where $\mathbf{C}$ is the locality constrained matrix, whose elements can be calculated as

$$\mathbf{C}_{ij} = 1 - \left(1 - \left(\frac{a}{\max_{v_i} \{a\}} \right)^2 \right)^2, \quad (10)$$

where $a = \text{dist}(\mathbf{x}_i, \mathbf{x}_j)$ denotes the spectral similarity of samples $\mathbf{x}_i$ and $\mathbf{x}_j$. The local constraint $\mathbf{C}$ can make similar samples represented by similar codes, thus improving the quality of the representation of the local structure.

Different from the spectral constraint, manifold learning is qualified to explore the intrinsic structure of the low-dimensional space. In this paper, the locality preserving projection is utilized as the graph regularization to learn the local structure of the hyperspectral data. As such, the objective function of the proposed LatLRPL method can be finally constructed as follows.

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{P}, \mathbf{Q}, \mathbf{E}} \quad & \|\mathbf{C} \odot \mathbf{Z}\|_* + \frac{\beta}{2} \|\mathbf{P}\|_F^2 + \lambda \|\mathbf{E}\|_1 + \gamma \text{Tr}(\mathbf{P}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{P}), \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{XZ} + \mathbf{QP}^T \mathbf{X} + \mathbf{E}, \mathbf{Q}^T \mathbf{Q} = \mathbf{I}, \end{aligned} \quad (11)$$

where γ is one parameter that can control the effect of the graph regularization.

From Equation (11), on the one hand, projection learning is integrated into a representation learning model, which can effectively improve the efficiency and robustness of projection computation with the desired dimensionality. On the other hand, regularization of the original data space and the low-dimensional feature space is able to capture the local intrinsic structure of the hyperspectral data. Theoretically, discriminative features can indeed be extracted by LatLRPL from the noisy and high-dimensional hyperspectral data. The framework of the proposed LatLRPL method is presented in Figure 1.

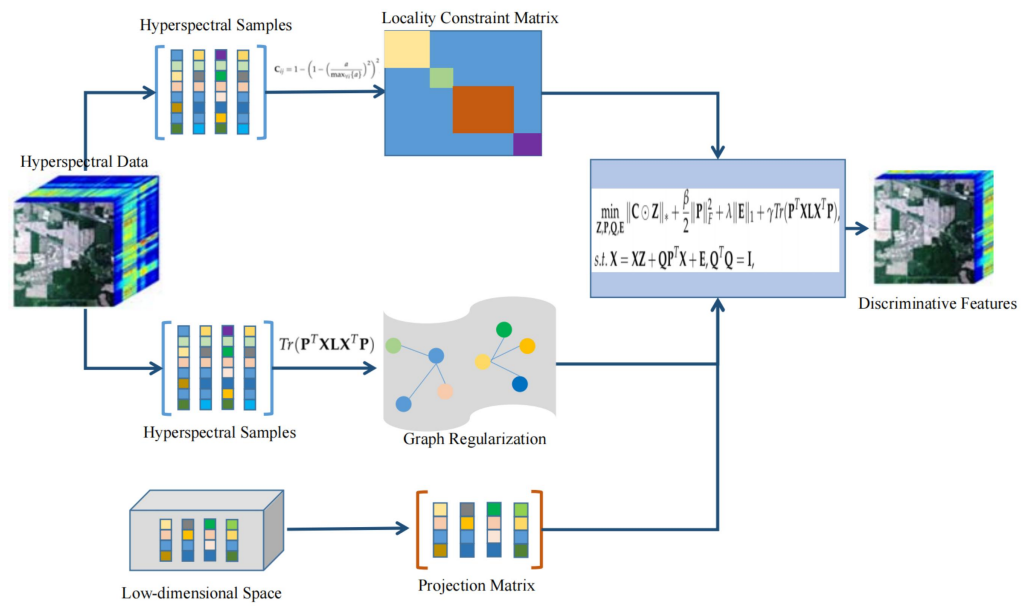


Figure 1. The framework of the proposed LatLRPL method.

3.2. Optimization of LatLRPL

In order to obtain the projection matrix, the alternating direction method of multipliers (ADMM) [48] is used to effectively optimize the objective function of the LatLRPL method. To effectively solve problem (11), we introduce two variables **A** and **B** to make the problem separable, generating the following function.

$$\begin{aligned} \min_{Z, A, P, B, Q, E} & \|C \odot A\|_* + \frac{\beta}{2} \|P\|_F^2 + \lambda \|E\|_1 + \gamma \text{Tr}(B^T X L X^T B), \\ \text{s.t. } & X = XZ + QP^T X + E, Z = A, P = B, Q^T Q = I. \end{aligned} \quad (12)$$

For problem (12), the augmented Lagrangian function is formulated as

$$\begin{aligned} L(Z, A, P, B, Q, E) &= \|C \odot A\|_* + \frac{\beta}{2} \|P\|_F^2 + \lambda \|E\|_1 + \gamma \text{Tr}(B^T X L X^T B) + \langle Y_1, X - XZ - QP^T X - E \rangle \\ &+ \langle Y_2, Z - A \rangle + \langle Y_3, P - B \rangle + \frac{\mu}{2} (\|X - XZ - QP^T X - E\|_F^2 + \|Z - A\|_F^2 + \|P - B\|_F^2), \end{aligned} \quad (13)$$

where Y_1 , Y_2 , and Y_3 are Lagrangian multipliers, μ is a penalty factor.

After some operations, Equation (13) can be further written as

$$\begin{aligned} L(Z, A, P, B, Q, E) &= \|C \odot A\|_* + \frac{\beta}{2} \|P\|_F^2 + \lambda \|E\|_1 + \gamma \text{Tr}(B^T X L X^T B) + \frac{\mu}{2} (\|X - XZ - QP^T X - E + \frac{Y_1}{\mu}\|_F^2 \\ &+ \|Z - A + \frac{Y_2}{\mu}\|_F^2 + \|P - B + \frac{Y_3}{\mu}\|_F^2) \end{aligned} \quad (14)$$

ADMM is an iterative optimization method in which variables are updated one by one. Therefore, the $(t + 1)$ th iterative schemes are shown in the following.

$$A_{t+1} = \min_A \|C \odot A\|_* + \frac{\mu_t}{2} \|Z_t - A + \frac{Y_{2,t}}{\mu_t}\|_F^2 \quad (15)$$

$$\begin{aligned} Z_{t+1} &= \frac{\mu_t}{2} (\|X - XZ - Q_t P_t^T X - E_t + \frac{Y_{1,t}}{\mu_t}\|_F^2 + \|Z - A_{t+1} + \frac{Y_{2,t}}{\mu_t}\|_F^2) \\ &= (X^T X + I)^{-1} (X^T S_1 + A_{t+1} - \frac{Y_{2,t}}{\mu_t}) \end{aligned} \quad (16)$$

$$\begin{aligned}\mathbf{B}_{t+1} &= \gamma \text{Tr}(\mathbf{B}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{B}) + \frac{\mu_t}{2} \|\mathbf{P}_t - \mathbf{B}\|_F^2 + \frac{\mathbf{Y}_{3,t}}{\mu_t} \|\mathbf{F}\|_F^2 \\ &= (\mu_t \mathbf{I} + \gamma(\mathbf{S}_2^T + \mathbf{S}_2))^{-1} (\mu_t \mathbf{P}_t + \mathbf{Y}_{3,t})\end{aligned}\quad (17)$$

$$\begin{aligned}\mathbf{P}_{t+1} &= \frac{\beta}{2} \|\mathbf{P}\|_F^2 + \frac{\mu_t}{2} (\|\mathbf{P} - \mathbf{B}_{t+1} + \frac{\mathbf{Y}_{3,t}}{\mu_t}\|_F^2 + \|\mathbf{X} - \mathbf{XZ}_{t+1} - \mathbf{Q}_t \mathbf{P}^T \mathbf{X} - \mathbf{E}_t + \frac{\mathbf{Y}_{1,t}}{\mu_t}\|_F^2) \\ &= ((\beta + \mu_t) \mathbf{I} + \mu_t \mathbf{X} \mathbf{X}^T)^{-1} (\mu_t \mathbf{X} \mathbf{S}_3^T \mathbf{Q}_t - \mu_t \mathbf{S}_4)\end{aligned}\quad (18)$$

$$\begin{aligned}\mathbf{Q}_{t+1} &= \frac{\mu_t}{2} \|\mathbf{X} - \mathbf{XZ}_{t+1} - \mathbf{Q} \mathbf{P}_{t+1}^T \mathbf{X} - \mathbf{E}_t + \frac{\mathbf{Y}_{1,t}}{\mu_t}\|_F^2 \\ &= \frac{\mu_t}{2} \|\mathbf{S}_3 - \mathbf{Q} \mathbf{P}_{t+1}^T \mathbf{X}\|_F^2, \text{ s.t. } \mathbf{Q}^T \mathbf{Q} = \mathbf{I}\end{aligned}\quad (19)$$

$$\begin{aligned}\mathbf{E}_{t+1} &= \min_{\mathbf{E}} \lambda \|\mathbf{E}\|_1 + \frac{\mu_t}{2} \|\mathbf{X} - \mathbf{XZ}_{t+1} - \mathbf{Q}_{t+1} \mathbf{P}_{t+1}^T \mathbf{X} - \mathbf{E} + \frac{\mathbf{Y}_{1,t}}{\mu_t}\|_F^2 \\ &= \mathcal{S}_{\frac{\lambda}{\mu_t}}(\mathbf{X} - \mathbf{XZ}_{t+1} - \mathbf{Q}_{t+1} \mathbf{P}_{t+1}^T \mathbf{X} + \frac{\mathbf{Y}_{1,t}}{\mu_t})\end{aligned}\quad (20)$$

where $\mathbf{S}_1 = \mathbf{X} - \mathbf{Q}_t \mathbf{P}_t^T \mathbf{X} - \mathbf{E}_t + \mathbf{Y}_{1,t}/\mu_t$, $\mathbf{S}_2 = \mathbf{X} \mathbf{L} \mathbf{X}^T$, $\mathbf{S}_3 = \mathbf{X} - \mathbf{XZ}_{t+1} - \mathbf{E}_t + \mathbf{Y}_{1,t}/\mu_t$, $\mathbf{S}_4 = \mathbf{Y}_{3,t}/\mu_t - \mathbf{B}_{t+1}$, $\mathcal{S}_{\frac{\lambda}{\mu}}(\cdot)$ denotes the soft threshold value method [49] with λ/μ as the threshold value.

For problem (15), the variable $\mathbf{A}$ has a closed-form solution according to [47], that is,

$$\mathbf{A}_{t+1} = (\mu_t \mathbf{Z}_t + \mathbf{Y}_{2,t}) \oslash (2(\mathbf{C} \odot \mathbf{C}) + \mu_t \mathbf{1}), \quad (21)$$

where $\oslash$ denotes the elementwise division, $\odot$ denotes the elementwise multiplication, $\mathbf{1}$ represents the all-ones matrix.

For problem (19), it looks like an orthogonal ProCrustes problem [50], which can be solved as

$$\mathbf{U} \Sigma \mathbf{V}^T = \mathcal{SVD}(\mathbf{S}_3 \mathbf{X}^T \mathbf{P}_{t+1}). \quad (22)$$

Finally, we can obtain $\mathbf{Q} = \mathbf{U} \mathbf{V}^T$.

The Lagrangian multipliers are updated as follows.

$$\begin{aligned}\mathbf{Y}_{1,t+1} &= \mathbf{Y}_{1,t} + \mu_t (\mathbf{X} - \mathbf{XZ}_{t+1} - \mathbf{Q}_{t+1} \mathbf{P}_{t+1}^T \mathbf{X} - \mathbf{E}_{t+1}), \\ \mathbf{Y}_{2,t+1} &= \mathbf{Y}_{2,t} + \mu_t (\mathbf{Z}_{t+1} - \mathbf{A}_{t+1}), \\ \mathbf{Y}_{3,t+1} &= \mathbf{Y}_{3,t} + \mu_t (\mathbf{P}_{t+1} - \mathbf{B}_{t+1}).\end{aligned}\quad (23)$$

After obtaining the projection matrix $\mathbf{P}$, the low-dimensional features can be computed by $\hat{\mathbf{X}} = \mathbf{P}^T \mathbf{X}$. The complete optimization process of LatLRPL is summarized in Algorithm 1.

Algorithm 1: ADMM for solving LatLRPL.

Input: N unlabeled samples $\mathbf{X} \in \mathbb{R}^{D \times N}$, regularization parameters β , λ , and γ , number of preserved features d .

Initialize: $\mathbf{Z}_0 = \mathbf{A}_0 = \mathbf{0}$, $\mathbf{P}_0 = \mathbf{B}_0 = \mathbf{0}$, $\mathbf{E}_0 = \mathbf{0}$, $\mathbf{Y}_{1,0} = \mathbf{0}$, $\mathbf{Y}_{2,0} = \mathbf{0}$, $\mathbf{Y}_{3,0} = \mathbf{0}$, $\mu_0 = 10^{-2}$, $\mu_{\max} = 10^6$, $\rho_0 = 1.1$, $\varepsilon = 10^{-5}$, $\text{maxIter} = 200$, $t = 0$.

1. Compute decomposition matrix $\mathbf{Q}$ by $\mathbf{Q} = \mathcal{PCA}(\mathbf{X})$.
2. Construct the spectral constraint $\mathbf{C}$ according to (10).
3. Construct the neighborhood graph $\mathbf{W}$ according to (3).
4. **Repeat:**
5. Compute $\mathbf{Q}_{t+1}$, $\mathbf{P}_{t+1}$, $\mathbf{B}_{t+1}$, $\mathbf{Z}_{t+1}$, $\mathbf{A}_{t+1}$, $\mathbf{E}_{t+1}$ according to (16)–(18), (20)–(22).
6. Update the Lagrangian multipliers according to (23).
7. Update μ : $\mu_{t+1} = \min(\rho_0 \mu_t, \mu_{\max})$.
8. Check convergence conditions:
 $\|\mathbf{X} - \mathbf{XZ}_{t+1} - \mathbf{Q}_{t+1}\mathbf{P}_{t+1}^T\mathbf{X} - \mathbf{E}_{t+1}\|_\infty < \varepsilon$, $\|\mathbf{Z}_{t+1} - \mathbf{A}_{t+1}\|_\infty < \varepsilon$,
 $\|\mathbf{P}_{t+1} - \mathbf{B}_{t+1}\|_\infty < \varepsilon$.
9. $t \leftarrow t + 1$.
10. **Until** convergence conditions are satisfied or $t > \text{maxIter}$.

Output: Discriminative projection matrix $\mathbf{P} \in \mathbb{R}^{D \times d}$.

3.3. Spatial Extension of LatLRPL

In HSIs, neighboring pixels usually represent the same land covers, which shows good advantages of feature extraction and classification. To improve the performance of the proposed LatLRPL method, spatial information is also taken into consideration in a simple but effective way in this paper. Specifically, a sliding window is exploited to extract pixels from the original hyperspectral data, followed by the weighted average operation on the pixels in this squared window. Then, the weighted average value will be taken as the value of the central pixel in the window. Finally, new hyperspectral data with each pixel fusing spatial information can be obtained, whose discriminative low-dimensional features can be extracted by the proposed LatLRPL method. For simplicity, SpaLatLRPL is used to denote this extended method.

4. Experiments and Discussions

In this section, some spectral-based feature extraction methods, e.g., LPP, NPE, SGE, LRGE, and spatial-spectral-based methods, e.g., MPCA, and TLPP, are adopted for comparison with the proposed method through the experiments on two hyperspectral datasets. To effectively evaluate the performance, the support vector machine (SVM) classifier is used to classify the low-dimensional features. The class-specific accuracy, overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ) are employed for quantitative assessment after ten repetitions of the experiment.

4.1. Hyperspectral Datasets

The first dataset is called Indian Pines, which was acquired in June 1992 by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. The AVIRIS sensor can acquire the wavelength range of 0.4–2.45 μm , which covers 220 spectral bands. After removing 20 water-absorption bands, 200 bands are reserved to conduct the experiments. This dataset presents a rural scenario with a size of 145×145 pixels. A total of 10,366 samples from 16 different land-cover classes are distinguished.

The second dataset is called Pavia University, which was collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor in Italy. After removing 12 noisy bands, 103 bands are preserved, which cover a spectral range from 0.43 to 0.86 μm . This dataset presents nine ground-truth classes distributed in a region of 610×340 pixels. The Indian

Pines and Pavia University datasets are available from http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes, accessed on 15 March 2022.

4.2. Analysis of the Regularization Parameters

From the proposed LatLRPL model, it can be directly seen that three regularization parameters λ , β , and γ play an important role in balancing the contributions of different regularization terms. As such, these three parameters are tuned meticulously, in which the search range is set as $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$. Figures 2 and 3 show the variation of the performance with different λ , β , and γ for the Indian Pines and the Pavia University datasets, respectively. To make it clear for observation, the best performance is marked in each figure. After comparison, $[10^{-3}, 10^{-3}, 10^{-1}]$ is confirmed for the Indian Pines dataset, and $[10^{-5}, 10^{-3}, 10^{-3}]$ is suitable for the Pavia University dataset.

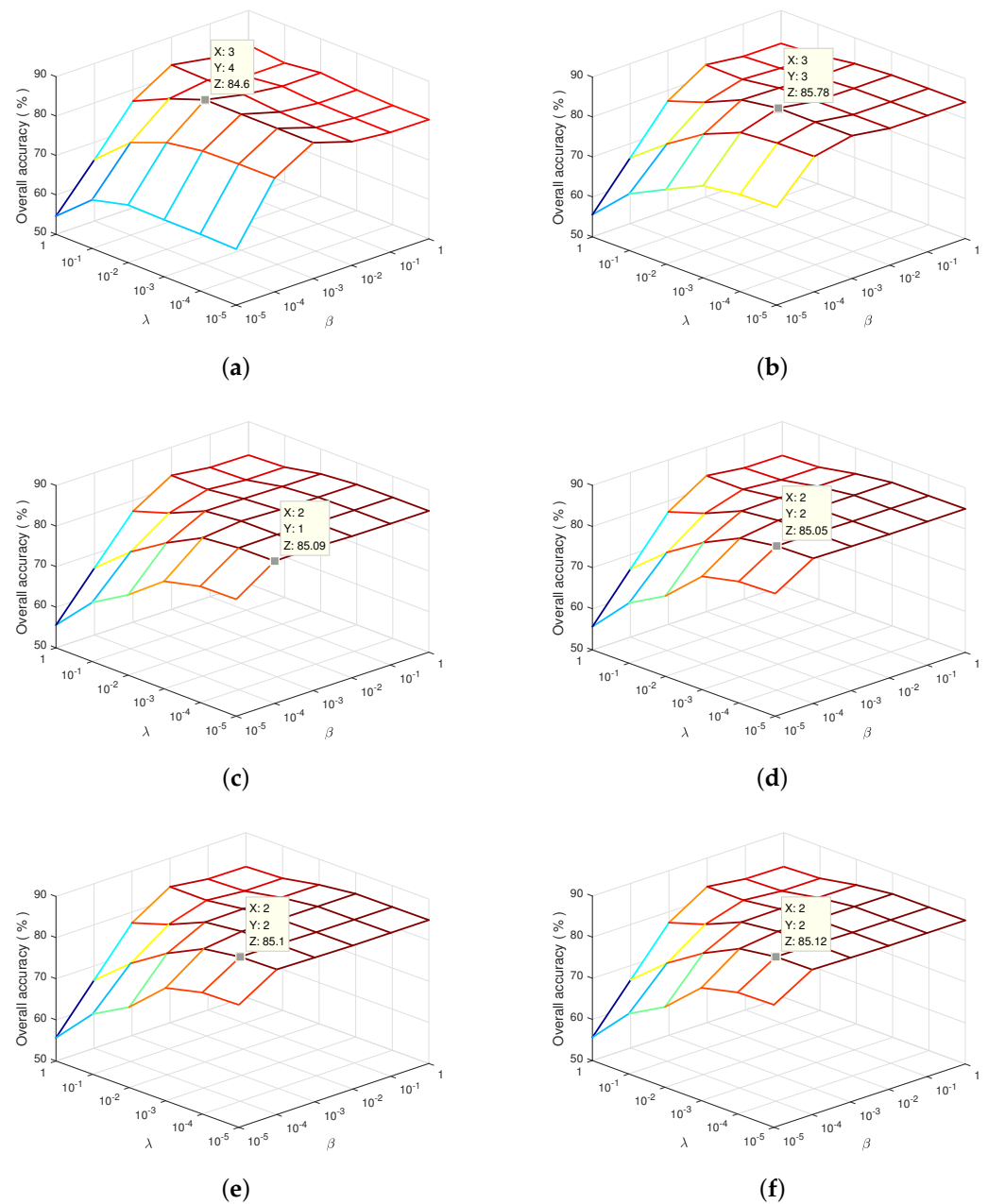


Figure 2. Parameter analysis of β , λ with different γ for the Indian Pines data: (a) $\gamma = 1$, (b) $\gamma = 0.1$, (c) $\gamma = 0.01$, (d) $\gamma = 0.001$, (e) $\gamma = 0.0001$, (f) $\gamma = 0.00001$.

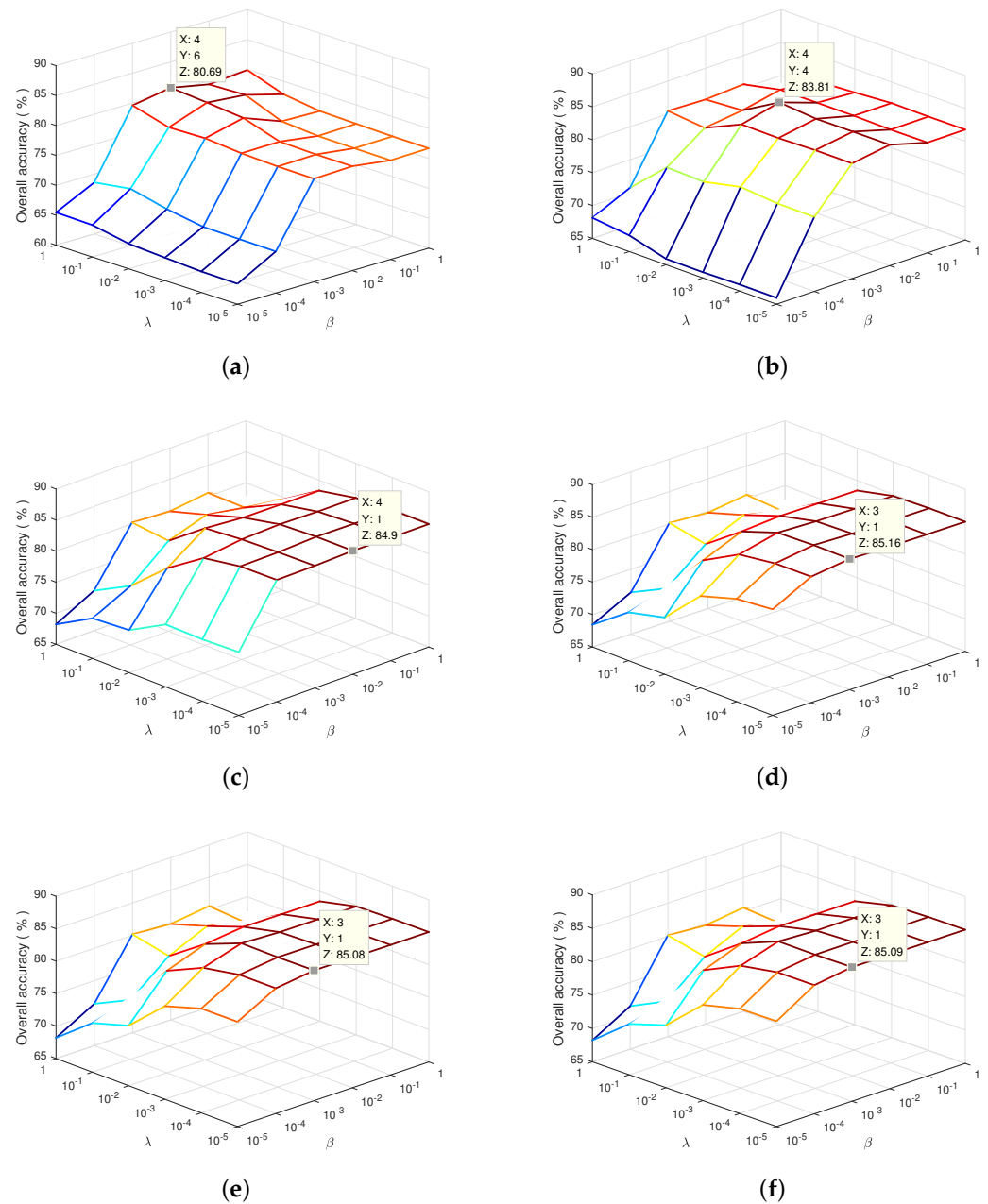


Figure 3. Parameter analysis of β , λ with different γ for the Pavia University data: (a) $\gamma = 1$, (b) $\gamma = 0.1$, (c) $\gamma = 0.01$, (d) $\gamma = 0.001$, (e) $\gamma = 0.0001$, (f) $\gamma = 0.00001$.

4.3. Analysis of the Sliding Window Size

For the proposed SpaLatLRPL method, the quality of data preprocessing is determined by the size of the sliding window. According to previous researches, the neighborhood size is searched from the range of $\{3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9, 11 \times 11, 13 \times 13\}$, whose results are shown in Figure 4 for the Indian Pines and the Pavia University datasets. Obviously, SpaLatLRPL can obtain the best performance when the window size is fixed at 9×9 and 11×11 for the two HSI datasets, respectively. Specifically, due to a much larger scene and significantly fewer categories presented in the Pavia University dataset than those in the Indian Pines dataset, a little larger neighborhood size is more suitable.

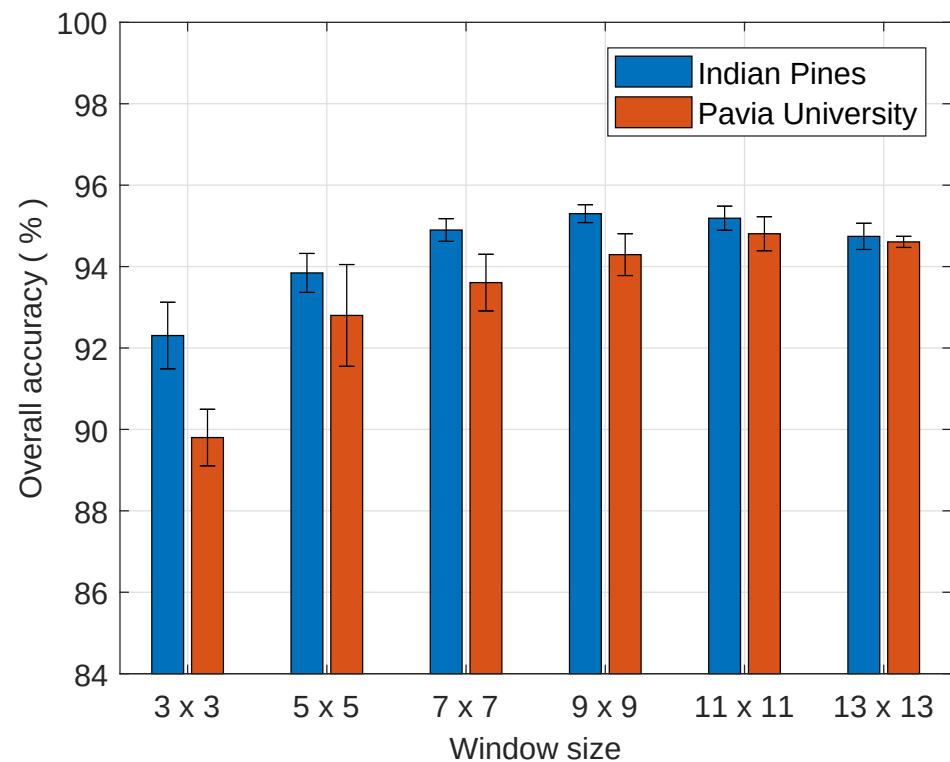


Figure 4. Parameter analysis of window size for two datasets.

4.4. Analysis of the Number of Features

For feature extraction methods, the number of preserved features is still a problem of concern. Figure 5 shows the performance variation with different numbers of features for all compared methods, from which we can observe that the performance improves quickly with the number of features increasing and remains almost stable when the number of features achieves a certain value. To conduct a better comparison of the performance, the number of features is set as 30 for all considered methods in both datasets.

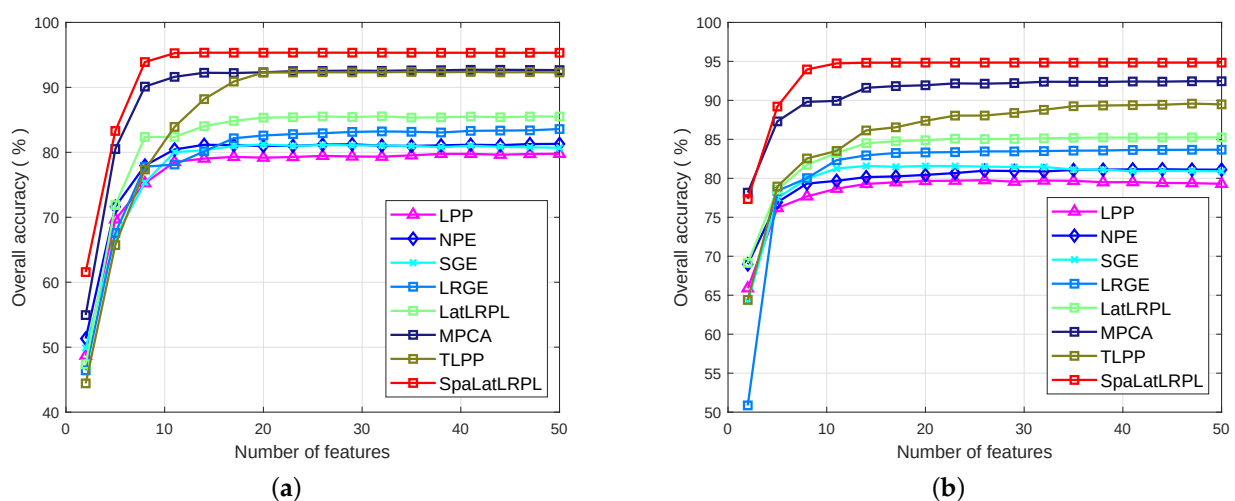


Figure 5. Parameter analysis of the number of preserved features: (a) Indian Pines, (b) Pavia University.

4.5. Comparison of the Classification Performance

Tables 1 and 2 present the class-specific accuracy, OA, AA, and κ coefficient for all compared methods by conducting the experiments on the two HSI datasets. Specifically, 10% samples randomly selected from the total samples are used to train the SVM classifier,

and the rest of the samples are utilized for testing in the Indian Pines dataset, while the percentage of training samples in the Pavia University dataset utilized for training is 1%. From Table 1, the proposed unsupervised LatLRPL method performs best among all the spectral-based feature extraction methods, presenting at least two percent gain when compared with that of LPP [21], NPE [23], SGE [29], and LRGE [35]. Specifically, LPP, NPE, SGE, and LRGE construct the neighborhood graph depending on the nearest distance or the largest representative coefficients, and the inaccuracy of the graph may lead to the error propagation to low-dimensional projection, which can degrade the discriminant ability of the extracted features. To be different, the proposed LatLRPL method learns the projection matrix in an integrated latent low-rank representation model, which is available to capture the intrinsic and discriminative features from the high-dimensional hyperspectral data.

Table 1. The comparison of performance for all methods on the Indian Pines dataset.

Class	LPP	NPE	SGE	LRGE	LatLRPL	MPCA	TLPP	SpaLatLRPL
1	39.59 ± 16.2	37.55 ± 12.6	42.45 ± 13.1	33.06 ± 4.87	29.80 ± 16.2	59.18 ± 5.40	90.48 ± 9.43	66.67 ± 8.25
2	75.83 ± 1.82	79.95 ± 2.24	79.36 ± 1.56	82.08 ± 0.99	83.41 ± 2.06	93.21 ± 1.39	89.88 ± 1.93	94.45 ± 0.74
3	64.61 ± 3.00	67.06 ± 2.82	66.79 ± 4.51	70.39 ± 2.15	77.20 ± 2.81	93.96 ± 0.27	83.89 ± 5.86	95.83 ± 0.70
4	52.70 ± 9.30	57.35 ± 6.72	53.93 ± 6.93	53.93 ± 8.71	50.71 ± 9.33	84.68 ± 7.29	89.89 ± 6.00	86.73 ± 9.25
5	92.98 ± 3.21	92.53 ± 2.76	93.42 ± 3.22	92.66 ± 3.68	91.81 ± 4.33	94.71 ± 1.37	95.82 ± 3.48	97.39 ± 1.06
6	95.39 ± 1.55	96.49 ± 1.27	96.19 ± 1.48	95.62 ± 1.20	94.88 ± 1.22	97.67 ± 1.08	98.46 ± 0.56	97.57 ± 0.73
7	20.00 ± 22.1	13.04 ± 13.4	31.30 ± 29.1	6.08 ± 8.48	38.26 ± 24.9	0.00 ± 0.00	55.07 ± 6.64	0.00 ± 0.00
8	99.27 ± 0.41	99.36 ± 0.19	99.45 ± 0.20	99.23 ± 0.59	98.27 ± 1.77	97.35 ± 1.74	99.92 ± 0.13	98.33 ± 0.65
9	3.33 ± 7.45	0.00 ± 0.00	4.44 ± 6.09	0.00 ± 0.00	8.88 ± 12.2	1.85 ± 3.21	33.33 ± 24.2	11.11 ± 19.2
10	58.05 ± 3.02	62.80 ± 5.24	60.32 ± 6.38	63.51 ± 4.03	77.57 ± 2.60	88.94 ± 1.35	89.55 ± 1.92	93.61 ± 0.43
11	80.87 ± 2.78	82.97 ± 0.60	82.35 ± 1.28	87.28 ± 1.62	89.30 ± 1.52	95.02 ± 0.61	92.65 ± 1.50	97.63 ± 0.60
12	77.07 ± 2.89	79.46 ± 1.68	80.83 ± 4.72	85.06 ± 2.28	85.75 ± 2.94	87.22 ± 5.27	87.10 ± 4.76	92.77 ± 2.20
13	98.32 ± 0.77	97.91 ± 1.39	98.43 ± 0.97	98.12 ± 1.32	94.87 ± 0.86	96.16 ± 2.69	97.21 ± 1.98	95.29 ± 2.40
14	96.21 ± 2.47	95.93 ± 3.05	96.64 ± 2.33	97.65 ± 0.90	96.07 ± 1.17	96.31 ± 1.12	99.43 ± 0.84	98.43 ± 0.13
15	57.66 ± 4.27	57.43 ± 5.50	60.76 ± 4.92	54.74 ± 0.95	57.43 ± 6.42	85.19 ± 4.69	91.91 ± 2.21	92.11 ± 5.87
16	82.59 ± 8.74	80.47 ± 9.58	83.29 ± 7.64	83.06 ± 6.31	87.53 ± 8.71	80.39 ± 13.2	84.31 ± 5.43	87.45 ± 8.76
OA	79.29 ± 0.65	81.20 ± 0.72	81.07 ± 0.73	83.15 ± 0.51	85.46 ± 0.58	92.79 ± 0.51	92.41 ± 1.01	95.34 ± 0.22
AA	68.40 ± 1.18	68.77 ± 0.47	70.62 ± 2.06	68.90 ± 0.46	72.61 ± 2.01	78.24 ± 0.85	86.18 ± 0.59	81.58 ± 0.74
κ	76.25 ± 0.68	78.46 ± 0.83	78.30 ± 0.89	80.65 ± 0.60	83.36 ± 0.67	91.77 ± 0.58	91.35 ± 1.16	94.68 ± 0.33

For the comparison of spatial-spectral-based methods, the proposed SpaLatLRPL method also shows a great improvement in the performance, obtaining an almost three-point gain when compared with that of MPCA and TLPP. Specifically, SpaLatLRPL presents good results in similar classes, such as class 10, class 11, and class 12. However, SpaLatLRPL performs not well in small classes, such as class 7 and class 9, which may be the reason that the neighborhood size adopted in the weighted average strategy is not proper for small classes. To overcome this problem, an adaptive neighborhood size may work well. By comparing the performance of the spectral-based methods and the spatial-spectral-based methods, we can observe that the improvement is prominent, indicating that the spatial

information is very beneficial for the feature extraction and classification task of HSIs. From Table 2, it can also be seen that the spatial-spectral-based methods perform much better than spectral-based methods in the Pavia University dataset, illustrating the importance of the spatial information. Overall, our proposed methods (LatLRPL and SpaLatLRPL) show great effectiveness in the feature extraction of HSIs, which also demonstrates that the spectral constraint and the graph regularization imposed on the original space and the low-dimensional space can improve the ability of the proposed methods in capturing local intrinsic structure, and in addition, the integrated projection learning framework is robust to noise.

Table 2. The comparison of performance for all methods on the Pavia University dataset.

Class	LPP	NPE	SGE	LRGE	LatLRPL	MPCA	TLPP	SpaLatLRPL
1	81.75 ± 4.30	83.528 ± 2.9	83.13 ± 3.46	84.05 ± 2.95	85.99 ± 4.33	92.75 ± 1.06	93.80 ± 0.55	95.54 ± 3.63
2	94.21 ± 1.34	95.05 ± 0.54	94.52 ± 1.06	95.26 ± 1.82	96.47 ± 1.23	97.86 ± 0.60	95.65 ± 1.14	99.40 ± 0.14
3	37.39 ± 12.2	39.73 ± 19.9	53.90 ± 6.78	48.79 ± 7.13	49.72 ± 11.1	75.22 ± 3.16	75.38 ± 4.00	77.90 ± 4.93
4	82.34 ± 3.54	77.74 ± 3.84	79.79 ± 5.43	77.42 ± 2.76	81.64 ± 2.41	87.58 ± 2.67	91.84 ± 2.41	91.92 ± 1.84
5	99.22 ± 0.54	99.14 ± 0.50	99.20 ± 0.69	98.02 ± 1.72	99.07 ± 0.62	99.13 ± 0.81	99.98 ± 0.03	99.27 ± 0.75
6	51.97 ± 4.58	55.59 ± 3.78	55.79 ± 2.87	58.05 ± 3.46	60.92 ± 5.71	77.55 ± 3.78	72.93 ± 3.61	93.39 ± 1.73
7	36.28 ± 8.19	40.35 ± 9.21	43.95 ± 6.51	61.47 ± 3.87	57.48 ± 9.14	81.96 ± 7.08	67.49 ± 9.61	92.28 ± 1.60
8	70.95 ± 7.86	73.48 ± 10.8	72.34 ± 2.81	80.71 ± 3.01	84.47 ± 3.91	87.19 ± 2.22	80.52 ± 3.01	84.57 ± 3.34
9	73.39 ± 7.60	80.26 ± 11.4	63.77 ± 34.4	98.76 ± 2.20	93.26 ± 4.23	97.89 ± 1.14	88.93 ± 4.59	92.00 ± 1.29
OA	79.57 ± 0.94	80.91 ± 0.99	81.14 ± 0.76	83.45 ± 0.90	85.07 ± 0.80	91.46 ± 0.56	89.23 ± 0.97	94.84 ± 0.34
AA	69.72 ± 1.68	71.65 ± 2.41	71.82 ± 4.09	78.06 ± 0.31	78.78 ± 1.59	88.57 ± 0.60	85.17 ± 0.87	91.81 ± 0.61
κ	72.32 ± 1.22	74.08 ± 1.39	74.46 ± 1.13	77.63 ± 1.12	79.81 ± 1.06	88.56 ± 0.76	85.61 ± 1.29	93.13 ± 0.66

4.6. Presentation of Classification Maps

To present the performance of different feature extraction methods more directly, the classification maps are shown in Figures 6 and 7, corresponding to the results given in Tables 1 and 2 for the two datasets, respectively. Taking a close look at these figures, we can see that spatial-spectral-based methods obtain much smoother classification results, among which SpaLatLRPL shows the best classification performance, especially for the regions of “Soybean-notill”, “Soybean-mintill”, and “Soybean-clean” in the Indian Pines dataset and “Grave” and “Bare soil” in the Pavia University dataset. In addition, for spectral-based feature extraction methods, the proposed unsupervised LatLRPL method performs better than other considered methods. The classification maps illustrate the superiority of the proposed latent low-rank projection learning methods.

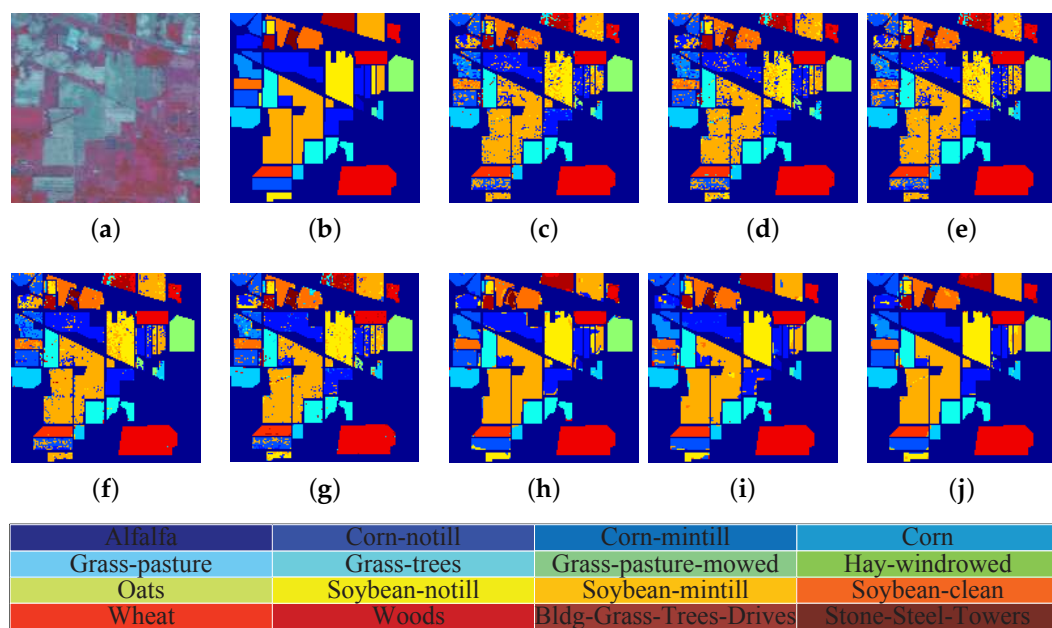


Figure 6. Classification maps of different methods for the Indian Pines dataset: (a) Pseudocolor image, (b) ground truth, (c) LPP, (d) NPE, (e) SGE, (f) LRGE, (g) LatLRPL, (h) MPCA, (i) TLPP, (j) SpaLatLRPL.

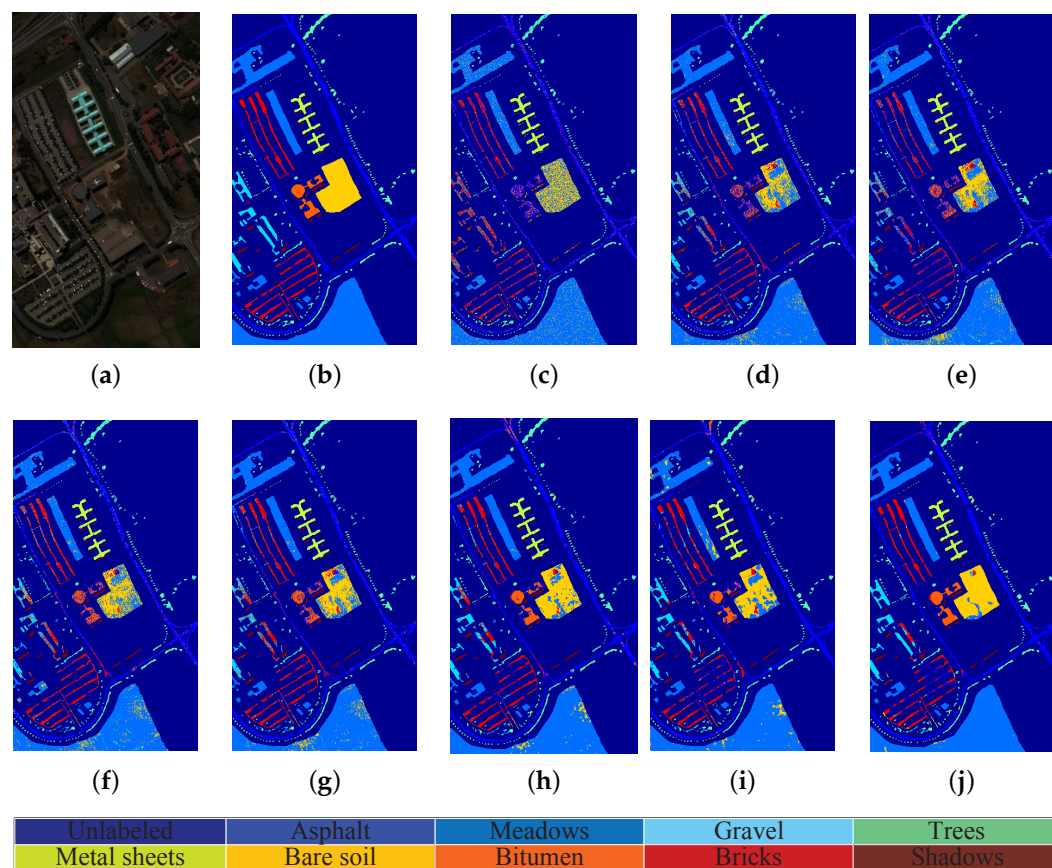


Figure 7. Classification maps of different methods for the Pavia University dataset: (a) Pseudocolor image, (b) ground truth, (c) LPP, (d) NPE, (e) SGE, (f) LRGE, (g) LatLRPL, (h) MPCA, (i) TLPP, (j) SpaLatLRPL.

5. Conclusions

In this paper, a novel unsupervised feature extraction method, i.e., latent low-rank projection learning with graph regularization (LatLRPL), has been proposed for the feature extraction and classification of HSIs. To obtain the discriminative features with arbitrary dimensionality, the latent low-rank matrix mentioned in the latent low-rank representation is first decomposed into two different matrices, one of which can be exploited as the projection matrix. To better preserve the local intrinsic structure, regularization on the original space and the learned low-dimensional space is integrated into LatLRPL by making use of spectral constraints and graph regularization, thus forming a more discriminative feature extraction model. Different from the graph embedding-based methods that require two steps to obtain the projection matrix, the proposed LatLRPL method can deal with this in one model. In order to take advantage of spatial information in HSIs, a local weighted average strategy is adopted by computing the value of a central pixel with its neighboring pixels in a sliding window, followed by conducting LatLRPL on this new data. Experiments on two HSI datasets have demonstrated that the proposed LatLRPL method and its extension, SpaLatLRPL, perform better than all compared feature extraction methods.

Author Contributions: All the authors made significant contributions to the work. L.P. designed the research model, analyzed the results and wrote the paper. H.L. contributed to the editing and review of the manuscript. X.D., X.H., Y.C. and L.D. reviewed the manuscript and provided some useful suggestions. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China under Grant 62001437 and Grant 61871335.

Data Availability Statement: All datasets used in this study are available from http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes, accessed on 15 March 2022.

Acknowledgments: The authors would like to thank the editors and the reviewers for their valuable suggestions.

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

References

1. Ghamisi, P.; Yokoya, N.; Li, J.; Liao, W.; Liu, S.; Plaza, J.; Rasti, B.; Plaza, A. Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 37–78. [CrossRef]
2. Wen, J.; Fowler, J.E.; He, M.; Zhao, Y.Q.; Deng, C.; Menon, V. Orthogonal nonnegative matrix factorization combining multiple features for spectral-spatial dimensionality reduction of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4272–4286. [CrossRef]
3. An, J.L.; Zhang, X.R.; Zhou, H.Y.; Jiao, L.C. Tensor-based low-rank graph with multimanifold regularization for dimensionality reduction of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4731–4746. [CrossRef]
4. Sun, W.W.; Du, Q. Hyperspectral band selection: A review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 118–139. [CrossRef]
5. Feng, J.; Feng, X.L.; Chen, J.T.; Cao, X.H.; Zhang, X.R.; Jiao, L.C.; Yu, T. Generative adversarial networks based on collaborative learning and attention mechanism for hyperspectral image classification. *Remote Sens.* **2020**, *12*, 1149. [CrossRef]
6. Feng, J.; Wu, X.D.; Shang, R.H.; Sui, C.H.; Li, J.; Jiao, L.C. Attention multibranch convolutional neural network for hyperspectral image classification based on adaptive region search. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5054–5070. [CrossRef]
7. Yang, L.X.; Zhang, R.; Yang, S.Y.; Jiao, L.C. Hyperspectral Image Classification via Slice Sparse Coding Tensor Based Classifier With Compressive Dimensionality Reduction. *IEEE Access* **2020**, *8*, 145207–145215. [CrossRef]
8. Hang, R.L.; Liu, Q.S.; Sun, Y.B.; Yuan, X.T.; Pei, H.C.; Plaza, J.; Plaza, A. Robust matrix discriminative analysis for feature extraction from hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2017**, *10*, 2002–2011. [CrossRef]
9. Varga, D. Multi-pooled inception features for no-reference image quality assessment. *Appl. Sci.* **2020**, *10*, 2186. [CrossRef]
10. Varga, D. No-reference video quality assessment using multi-pooled, saliency weighted deep features and decision fusion. *Sensors* **2022**, *22*, 2209. [CrossRef]
11. Lin, C.Z.; Lu, J.W.; Wang, G.; Zhou, J. Graininess-aware deep feature learning for pedestrian detection. *IEEE Trans. Image Process.* **2020**, *29*, 3820–3834. [CrossRef] [PubMed]
12. Sethy, P.K.; Barpanda, N.K.; Rath, A.K.; Behera, S.K. Deep feature based rice leaf disease identification using support vector machine. *Comput. Electron. Agric.* **2020**, *175*, 105527. [CrossRef]
13. Zhang, W.B.; Zhang, L.M.; Pfoser, D.; Zhao, L. Disentangled dynamic graph deep generation. *arXiv* **2020**, arXiv:2010.07276.

14. Jia, S.; Liao, J.H.; Xu, M.; Li, Y.; Zhu, J.S.; Sun, W.W.; Jia, X.P.; Li, Q.Q. 3-D gabor convolutional neural network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5509216. [\[CrossRef\]](#)
15. Wan, S.; Pan, S.R.; Zhong, P.; Chang, X.J.; Yang, J.; Gong, C. Dual interactive graph convolutional networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5510214. [\[CrossRef\]](#)
16. Jolliffe, I.T. *Principal Component Analysis*; Springer: New York, NY, USA, 2002.
17. Bandos, T.V.; Bruzzone, L.; Camps-Valls, G. Classification of hyperspectral images with regularized linear discriminant analysis. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 862–873. [\[CrossRef\]](#)
18. Sugiyama, M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *J. Mach. Learn. Res.* **2007**, *8*, 1027–1061.
19. Ma, L.; Crawford, M.M.; Yang, X.Q.; Guo, Y. Local-mainfold-learning-based graph construction for semisupervised hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2832–2844. [\[CrossRef\]](#)
20. Liao, D.P.; Qian, Y.T.; Tang, Y.Y. Constrained manifold learning for hyperspectral imagery visualization. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2017**, *11*, 1213–1226. [\[CrossRef\]](#)
21. He, X.F.; Niyogi, P. Locality preserving projections. In Proceedings of the 17th Annual Conference on Neural Information Processing Systems (NIPS 2003), Vancouver, BC, Canada, 8–13 December 2003; pp. 234–241.
22. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [\[CrossRef\]](#)
23. He, X.F.; Cai, D.; Yan, S.C.; Zhang, H.J. Neighborhood preserving embedding. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 17–21 October 2005; pp. 1208–1213.
24. Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **2003**, *15*, 1373–1396. [\[CrossRef\]](#)
25. Yan, S.C.; Xu, D.; Zhang, B.Y.; Zhang, H.-J.; Yang, Q.; Lin, S. Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 40–51. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Peng, J.T.; Li, L.Q.; Tang, Y.Y. Maximum likelihood estimation based joint sparse representation for the classification of hyperspectral remote sensing images. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 1790–1802. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Gu, Y.F.; Wang, Y.T.; Zheng, H.; Hu, Y. Hyperspectral target detection via exploiting spatial-spectral joint sparsity. *Neurocomputing* **2015**, *169*, 5–12. [\[CrossRef\]](#)
29. Cheng, B.; Yang, J.C.; Yan, S.C.; Fu, Y.; Huang, T.S. Learning with l_1 -graph for image analysis. *IEEE Trans. Image Process.* **2015**, *23*, 2241–2253.
30. Qiao, L.S.; Chen, S.C.; Tan, X.Y. Sparsity preserving projections with applications to face recognition. *Pattern Recognit.* **2010**, *43*, 331–341. [\[CrossRef\]](#)
31. Ly, N.H.; Du, Q.; Fowler, J.E. Sparse graph-based discriminant analysis for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3872–3884.
32. He, W.; Zhang, H.Y.; Zhang, L.P.; Philips, W.; Liao, W.Z. Weighted sparse graph based dimensionality reduction for hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 686–690. [\[CrossRef\]](#)
33. Ly, N.H.; Du, Q.; Fowler, J.E. Collaborative graph-based discriminant analysis for hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2014**, *7*, 2688–2696. [\[CrossRef\]](#)
34. Zhang, L.; Yang, M.; Feng, X.H. Sparse representation or collaborative representation: Which helps face recognition? In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011.
35. Liu, G.C.; Lin, Z.C.; Yan, S.C.; Sun, J.; Yu, Y.; Ma, Y. Robust recovery of subspace structure by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 171–184. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Zhuang, L.S.; Gao, S.H.; Tang, J.H.; Wang, J.J.; Lin, Z.C.; Ma, Y.; Yu, N.H. Constructing a nonnegative low-rank and sparse graph with data-adaptive features. *IEEE Trans. Image Process.* **2015**, *24*, 3717–3728. [\[CrossRef\]](#) [\[PubMed\]](#)
37. 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. [\[CrossRef\]](#)
38. Li, W.; Liu, J.B.; Du, Q. Sparse and low-rank graph for discriminant analysis of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4094–4105. [\[CrossRef\]](#)
39. Pan, L.; Li, H.-C.; Li, W.; Chen, X.-D.; Wu, G.-N.; Du, Q. Discriminant analysis of hyperspectral imagery using fast kernel sparse and low-rank graph. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6085–6098. [\[CrossRef\]](#)
40. 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. [\[CrossRef\]](#)
41. Mei, S.H.; Bi, Q.Q.; Ji, J.Y.; Hou, J.H.; Du, Q. Hyperspectral image classification by exploring low-rank property in spectral or/and spatial domain. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2017**, *10*, 2910–2921. [\[CrossRef\]](#)
42. He, W.; Zhang, H.Y.; Zhang, L.P.; Shen, H.F. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 178–188. [\[CrossRef\]](#)
43. Xue, J.Z.; Zhao, Y.Q.; Liao, W.Z.; Kong, S.G. Joint Spatial and Spectral Low-Rank Regularization for Hyperspectral Image Denoising. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 1940–1958. [\[CrossRef\]](#)

-
44. Liu, G.C.; Yan, S.C. Latent low-rank representation for subspace segmentation and feature extraction. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011.
 45. Pan, L.; Li, H.-C.; Sun, Y.-J.; Du, Q. Hyperspectral image reconstruction by latent low-rank representation for classification. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1422–1426. [[CrossRef](#)]
 46. Fang, X.Z.; Han, N.; Wu, J.G.; Xu, Y. Approximate low-rank projection learning for feature extraction. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 5228–5241. [[CrossRef](#)] [[PubMed](#)]
 47. Pan, L.; Li, H.-C.; Meng, H.; Li, W.; Du, Q.; Emery, W.J. Hyperspectral image classification via low-rank and sparse representation with spectral consistency constraint. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2117–2121. [[CrossRef](#)]
 48. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [[CrossRef](#)]
 49. Cai, J.-F.; Candès, E.J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **2010**, *24*, 1956–1982. [[CrossRef](#)]
 50. Zou, H.; Hastie, T.; Tibshirani, R. Sparse principle component analysis. *J. Comput. Graph. Stat.* **2006**, *15*, 265–286. [[CrossRef](#)]