Local Low-Rank Approximation With Superpixel-Guided Locality Preserving Graph for Hyperspectral Image Classification

Shujun Yang, *Member, IEEE*, Yu Zhang, *Member, IEEE*, Yuheng Jia, *Member, IEEE*, and Weijia Zhang, *Member, IEEE*

***Abstract*—Given the detrimental effect of spectral variations in a hyperspectral image (HSI), this article investigates to recover its discriminative representation to improve the classification performance. We propose a new method, namely local low-rank approximation with superpixel-guided locality preserving graph (LLRA-SLPG), which can reduce the spectral variations and preserve the local manifold structure of an HSI. Specifically, the LLRA-SLPG method first clusters pixels of an HSI into several groups (i.e., superpixels). By taking advantage of the local manifold structure, a Laplacian graph is constructed from the superpixels to ensure that a typical pixel should be similar to its neighbors within the same superpixel. The LLRA-SLPG model can increase the compactness of pixels belonging to the same class by reducing spectral variations while promoting local consistency via the Laplacian graph. The objective function of the LLRA-SLPG model can be solved efficiently in an iterative manner. Experimental results on four benchmark datasets validate the superiority of the LLRA-SLPG model over state-of-the-art methods, particularly in cases where only extremely few training pixels are available.**

***Index Terms*—Hyperspectral image classification, low-rank, superpixel segmentation, superpixel-guided locality preserving graph.**

I. Introduction

A HYPERSPECTRAL image (HSI) consists of multiple narrow spectral bands, which span from the visible light to the infrared spectrum [1]. Due to the rich spectral information, HSIs have wide applications in various areas, such as mineral detection [2], agriculture [3], [4], urban planning [5], [6], and environment monitoring [7], [8]. However, given the noise associated with complex environmental conditions, the acquired HSIs often suffer from spectral variations [9], i.e., pixels of the same material may change considerably and significantly impair the performance of HSI classification.

Many low-rank-based methods have been proposed to alleviate the impact of spectral variations and accordingly improve classification accuracy. Particularly, much attention has been paid to the low-rank approximation (LRA)-based methods [10], [11], which assume that the pixels of an HSI should be distributed in one or more low-dimensional space(s) [12], [13]. For example, some methods [14], [15] assume that the data come from a unified subspace and then employ robust LRA on a whole HSI. To better capture the data drawn from the union of multiple subspaces, a multi-subspace-based LRA [16] is used to model an entire HSI [17], [18], [19], [20], [21], [22], [23]. All the above methods assume the global low-rank property, i.e., applying robust LRA or multisubspace-based LRA to the entire HSI. However, they neglect a fact that pixels within a homogeneous local region are often from the same class, i.e., the local low-rank property. To this end, many local LRA-based methods have been proposed [24], [25], [26], [27], [28], which first divide an HSI into multiple rectangular patches and then process each patch individually via robust LRA. As shape-adaptive regions (i.e., superpixel) are better at capturing the complex local spatial structure than rectangular patches, the superpixel-based segmentation has also been used in some local LRA-based methods [29], [30], [31]. It is worth noting that all of the above studies conduct LRA on 2D matrices. As an HSI can be naturally represented as a 3-D tensor, some tensor LRA-based methods have been proposed [32], [33], [34], [35], [36]. Additionally, due to the powerful representation ability of deep neural networks [37], [38], [39], some low-rank-based deep learning methods [39], [40] have been proposed. Besides, for other advanced feature extraction methods [41], [42] irrelevant to low-rank-based methods, the graph preserving is utilized to enhance the representation of an HSI.

Although some of the existing local LRA-based methods use superpixels to characterize the complex local spatial structure; they fail to preserve the local manifold structure, which is

Manuscript received 20 May 2022; revised 8 July 2022; accepted 9 August 2022. Date of publication 18 August 2022; date of current version 16 September 2022. This work was supported in part by NSFC key under Grant 62136005, in part by NSFC General under Grant 62076118, and Shenzhen fundamental research program under Grant JCYJ20210324105000003. *(Corresponding author: Yu Zhang.)*

Shujun Yang is with the Guangdong Provincial Key Laboratory of Brain-Inspired Intelligent Computation, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: sjyang8-c@my.cityu.edu.hk).

Yu Zhang is with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China, and also with the Peng Cheng Laboratory, Shenzhen 518066, China (e-mail: yu.zhang.ust@gmail.com).

Yuheng Jia is with the School of Computer Science and Engineering, Southeast University, Nanjing 211189, China, and also with the Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, Nanjing 211189, China (e-mail: yuheng.jia@my.cityu.edu.hk).

Weijia Zhang is with the School of Computer Science and Engineering, Southeast University, Nanjing 211189, China (e-mail: weijia.zhang.xh@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2022.3199885

unfavorable to the discriminative ability of the representation. In other words, they cannot sufficiently capture the locality of an HSI, which limits the further improvement of classification performance.

Considering the above points, we propose a local low-rank approximation with superpixel-guided locality preserving graph (LLRA-SLPG) model, which can increase the within-class compactness by reducing the spectral variations for each superpixel and promote the local consistency via a proposed superpixel-guided locality preserving graph. Moreover, our proposed LLRA-SLPG method is the first to combine the local manifold structures with the local LRA into a unified model. Our targets are to use the local low-rank term with the ℓ_1 norm to separate the low-rank part and sparse noise part of an HSI, and use the superpixel-guided locality preserving graph to further preserve the local manifold structure of the low-rank part. Specifically, the proposed graph regularizer in the LLRA-SLPG method makes the pixel similar to its neighbors within the same superpixel. We formulate the objective function of the LLRA-SLPG method as a constrained optimization problem, which can be solved efficiently in an iterative manner. Extensive experiments validate the superiority of the LLRA-SLPG model over state-of-the-art methods, particularly for the cases with few training samples. The contributions of this article are given as follows.

1) Our proposed LLRA-SLPG method is the first to fuse the local LRA and the local manifold structures into a unified model simultaneously. Specifically, existing local LRA-based methods cannot adequately capture the complex local spatial structures of the original HSI, e.g., local manifolds (or local consistencies) are ignored, which further limits the classification performance. To address this issue, our proposed LLRA-SLPG method designs a superpixel-guided locality preserving graph to enhance preserving the local spatial structure of an HSI, thus further improving the discriminative ability of the representation.
2) The proposed LLRA-SLPG method can be solved efficiently in an iterative manner. The experiments demonstrate its advantage over state-of-the-art methods, especially when the training samples are extremely few.
3) We organize the remainder of this article as follows. Section II gives a brief review of existing LRA-based methods for HSI. Then, we present the proposed LLRA-SLPG method in Section III, followed by the comprehensive experimental results and analyses in Section IV. Finally, Section V concludes this article.

II. Related Work

This section reviews existing LRA-based methods for modeling HSIs.

Some methods assume that the data are distributed in a single low-rank subspace and simply apply LRA on a whole HSI to recover a discriminative representation. For example, Mei et al. [14] apply ℓ_1-based LRA to reduce the spectral variations and improve the HSI classification. Given the data drawn from a union of multiple subspaces, multisubspace-based LRA is introduced to represent the HSI in some methods [17], [18], [19], [20], [21], [22], [23]. Specifically, Lu et al. [17] proposed a graph regularized LRA method within multiple subspaces to remove striping noise of an HSI. Sumarsono and Du used a multisubspace-based LRA model [18] to preprocess the spectral feature, which is then used by both supervised and unsupervised learning methods. Wang et al. [19] incorporated the local geometric structure into a multisubspace-based LRA model to improve the classification performance. Wang et al. [20] and Mei et al. [21] used cluster-based regularizers to incorporate the superpixel information into the multisubspace-based LRA model, that is, they represent the feature of a superpixel as a cluster center and make all the pixels in the superpixel close to the cluster center. Xu et al. [22] integrated a hypergraph-based regularizer into a multisubspace-based LRA model for the unsupervised HSI classification, whose graph induces the spatial-spectral information based on a superpixel segmentation. Xing et al. [23] incorporated a classwise regularization into a multisubspace-based LRA model to capture the classwise block-diagonal structure, which maps pixels from one class into the same subspace.

All the aforementioned methods adopt the global low-rank assumption. As the pixels within a small local region are usually from the same class, some local LRA-based methods have been proposed. Zhang et al. [24] and Zhu et al. [25] divided an HSI into regular patches and apply robust LRA patch by patch for HSI restoration. As the noise level in different bands may change significantly, He et al. [26] proposed a noise-adjusted iteration framework that uses a patchwise LRA method for HSI denoising. Apart from applying the patchwise LRA in the spectral domain, Mei et al. [27], [28] explored the low-rank property from the spatial domain (i.e., applying LRA to a whole HSI on each spectral band). Specifically, Mei et al. [27] applied LRA to an HSI from the spectral and spatial domains in two distinct steps. Moreover, Mei et al. [28] also proposed a unified model to combine the spectral and spatial low-rank property.

Since the patchwise segmentation cannot thoroughly exploit the complicated local spatial structure of an HSI, some superpixel-induced local LRA-based methods have been proposed [29], [30], [31], in which each superpixel can be regarded as a shape-adaptive region. Specifically, Xu et al. [29] first perform LRA on all the superpixels to extract the low-rank data, followed by a Markov random field to define the local correlation. Fan et al. [30] first employed PCA to obtain the first component of an HSI, which is then processed by the superpixel segmentation method to get the homogeneous regions. Finally, they apply the LRA to each homogeneous region to remove the noise and outliers. Yang et al. [31] proposed a discriminative low-rank model that can increase the intraclass similarity by applying LRA on each superpixel while promoting the global separability between classes. Based on such a model, a superpixel-based classification framework is proposed to utilize the prediction of a typical classifier to improve the superpixel segmentation.

Due to an HSI can be inherently depicted as a 3-D tensor, some tensor-based LRA methods have been investigated [32], [33], [34], [35], [36]. An et al. [32] proposed a tensor-based

low-rank graph to perform the dimension reduction for an HSI, and this method characterizes the intraclass compactness and the interclass separability via a multimanifold regularizer. Deng et al. [33] developed a tensor low-rank discriminative embedding model, which utilizes the low-rank reconstruction to uncover the potential sample relationships and incorporates the label information to improve the discriminability of features. Deng et al. [34] proposed a patch tensor-based multigraph embedding framework, which builds three different types of subgraphs to capture the intrinsically geometrical structure of HSIs. Sun et al. [35] proposed a lateral-slice sparse tensor robust principal component analysis to remove noises or outliers in an HSI to improve the subsequent classification performance. Liu et al. [36] proposed a local–global balanced tensor LRA method, which can be viewed as an extension of [31] to the tensor case.

Given the powerful representation learning capacity of deep neural networks [37], [38], [39], some deep learning models [39], [40] were proposed to combine the low-rank property with deep learning. Specifically, Wang et al. [39] proposed an unsupervised segmented stacked denoising autoencoder for extracting features, followed by a low-rank classifier for the HSI classification. Zhang et al. [40] proposed an end-to-end low-rank spatial-spectral network for removing noises in HSIs. By integrating the low-rank property into a deep convolutional neural network (DCNN), this method benefits from the powerful feature representation ability of DCNN and the good generalization ability of the low-rank property.

III. Methodology

$\mathbf{Y} \in \mathbb{R}^{b\times n}$ specifies an HIS, which contains n pixels, and each pixel is represented with b dimensional spectral bands. To suppress the spectral variations, a straightforward approach is to adopt the global low-rank approximation of an HSI by solving the following objective function as

$$\min_{[\mathbf{Z},\mathbf{N}]} \|\mathbf{Z}\|_* + \gamma\|\mathbf{N}\|_1, \text{ s.t. } \mathbf{Y} = \mathbf{Z} + \mathbf{N} \tag{1}$$

where $\|\cdot\|_1$ and $\|\cdot\|_*$ are the ℓ_1 norm and nuclear norm operations, respectively, γ is a nonnegative regularization parameter, $\mathbf{Z} \in \mathbb{R}^{b\times n}$ is the low-rank part of $\mathbf{Y}$, and $\mathbf{N} \in \mathbb{R}^{b\times n}$ is the sparse noise part (or called spectral variations). A drawback for this approach defined in problem (1) is that the local spatial structure cannot be exploited in such a global manner [24], [29], as pixels within a homogeneous region often belong to the same category.

To address these problems, we follow the superpixel-induced local LRA-based methods, i.e., applying the LRA on each superpixel. Note that such a way can enhance the data compactness for each homogeneous region (i.e., superpixel), which increases the intraclass similarity. However, existing methods cannot preserve complex local spatial structures of the original HSI, such as local manifolds (or local consistencies). To address this issue, we propose a superpixel-guided locality preserving graph-based regularizer to maintain the local consistency by forcing each pixel to have a similar spectral feature representation to its neighbors (i.e., the neighboring pixels located within a squared window of a homogeneous region). In this manner, the local manifold structure of the original HSI can be adequately exploited to improve the discriminative ability of the representation for the low-rank part $\mathbf{Z}$.

A. Superpixel-Guided Locality Preserving Graph

This section describes how to design a superpixel-guided locality-preserving graph in detail.

We first adopt the entropy rate superpixel method [43] to generate superpixels. To define such a graph, we first refer to (p_i, q_i) as the pixel location of the ith pixel (i.e., $\mathbf{y}_i$). The neighboring pixels of $\mathbf{y}_i$ are defined as the pixels within a squared window with the radius r centered on $\mathbf{y}_i$, and these neighboring pixels that include $\mathbf{y}_i$ constitute the neighbor set $\mathrm{NB_i}$, that is, the pixel locations in $\mathrm{NB_i}$ should satisfy $\{(p_j, q_j) \mid |p_i - p_j| \le r, |q_i - q_j| \le r\}$. To preserve the local consistency of the original HSI $\mathbf{Y}$, i.e., to make the discriminative representation of the ith pixel similar to that of its neighbors within the same superpixel, the weight $\mathbf{W}_{i,j}$ between the ith and jth pixels is defined as

$$\mathbf{W}_{i,j} = \begin{cases} \exp\left(-\gamma \|\mathbf{y}_i - \mathbf{y}_j\|_2^2\right), & \mathbf{y}_j \in \mathrm{NB_i} \wedge \mathrm{I_i} = \mathrm{I_j} \\ 0, & \mathbf{y}_j \notin \mathrm{NB_i} \vee \mathrm{I_i} \neq \mathrm{I_j} \end{cases} \tag{2}$$

where I_i and I_j represent the indices of superpixels that the ith and jth pixel belong to, respectively.

Finally, the superpixel-guided Laplacian graph regularizer can then be established based upon $\mathbf{W}$ as

$$\begin{aligned} &\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{z}_i - \mathbf{z}_j\|_2^2 \mathbf{W}_{i,j} \\ &= \mathrm{Tr}\left(\mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathrm{T}}\right) - \mathrm{Tr}\left(\mathbf{Z}\mathbf{W}\mathbf{Z}^{\mathrm{T}}\right) \\ &= \mathrm{Tr}\left(\mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}}\right) \end{aligned} \tag{3}$$

where $\mathbf{z}_i$ denotes the ith column of $\mathbf{Z}$, $\mathrm{Tr}(\cdot)$ computes the trace of the corresponding matrix, $\mathbf{D}$ is a diagonal matrix with the ith diagonal element defined as $\mathbf{D}_{i,i} = \sum_{j=1}^{n} \mathbf{W}_{i,j}$, and $\mathbf{G} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix.

B. The Model

The objective function of the proposed LLRA-SLPG model is formulated as

$$\begin{aligned} &\min_{[\mathbf{Z},\mathbf{N}]} \sum_{i=1}^{S} \|\mathbf{Z}_i\|_* + \lambda\|\mathbf{N}\|_1 + \beta\mathrm{Tr}\left(\mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}}\right) \\ &\text{s.t. } \mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_S],\ \mathbf{N} = [\mathbf{N}_1, \mathbf{N}_2, \ldots, \mathbf{N}_S] \\ &\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_S],\ \mathbf{Y} = \mathbf{Z} + \mathbf{N} \end{aligned} \tag{4}$$

where S denotes the number of superpixels, $\mathbf{Y}_i \in \mathbb{R}^{b\times n_i}$ denotes the ith superpixel which contains n_i pixels, $\mathbf{Z}_i \in \mathbb{R}^{b\times n_i}$ and $\mathbf{N}_i \in \mathbb{R}^{b\times n_i}$ denote decomposed parts whose sum is $\mathbf{Y}_i$, and λ, β are nonnegative regularization parameters.

In the objective of problem (4), the first term (i.e., $\sum_{i=1}^{S} \|\mathbf{Z}_i\|_*$) focuses on the data compactness for each $\mathbf{Z}_i$ via employing LRA on each superpixel. The second term (i.e., $\lambda\|\mathbf{N}\|_1$) uses ℓ_1 norm to make the spectral variations $\mathbf{N}$ sparse.

The third term makes the local manifold of $\mathbf{Y}$ consistent with that of the low-rank representation $\mathbf{Z}$.

C. Optimization

To solve problem (4), we first introduce an auxiliary variable $\mathbf{Q} \in \mathbb{R}^{b\times n}$ and reformulate it as

$$\begin{aligned}\min_{[\mathbf{Z},\mathbf{N},\mathbf{Q}]} &\sum_{i=1}^{S} \|\mathbf{Z}_i\|_* + \lambda\|\mathbf{N}\|_1 + \beta \mathrm{Tr}\left(\mathbf{Q}\mathbf{G}\mathbf{Q}^{\mathrm{T}}\right) \\ s.t.\ &\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_S], \mathbf{N} = [\mathbf{N}_1, \mathbf{N}_2, \ldots, \mathbf{N}_S] \\ &\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_S] \\ &\mathbf{Y} = \mathbf{Z} + \mathbf{N},\ \mathbf{Q} = \mathbf{Z}. \end{aligned} \tag{5}$$

The inexact augmented Lagrangian multiplier (IALM) [44] is employed to solve problem (5) alternatively. The augmented Lagrangian of problem (5) is formulated as

$$\begin{aligned}\min_{[\mathbf{Z},\mathbf{N},\mathbf{Q}]} &\sum_{i=1}^{S} \|\mathbf{Z}_i\|_* + \lambda\|\mathbf{N}\|_1 + \beta \mathrm{Tr}\left(\mathbf{Q}\mathbf{G}\mathbf{Q}^{\mathrm{T}}\right) \\ &+ \frac{\rho}{2}\left\|\mathbf{Y} - \mathbf{Z} - \mathbf{N} + \frac{\mathbf{\Gamma}_1}{\rho}\right\|_F^2 + \frac{\rho}{2}\left\|\mathbf{Q} - \mathbf{Z} + \frac{\mathbf{\Gamma}_2}{\rho}\right\|_F^2 \end{aligned} \tag{6}$$

where $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2 \in \mathbb{R}^{b\times n}$ are two Lagrange multipliers, ρ is a nonnegative penalty parameter, and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. We alternatively update the three variables (i.e., $\mathbf{Z}$, $\mathbf{N}$, and $\mathbf{Q}$) by solving the three subproblems in each iteration until convergence. In the following, we give the details of the three subproblems as well as the update of Lagrange multipliers.

a): By fixing other variables, the subproblem to update $\{\mathbf{Z}_i\}_{i=1}^S$ is formulated as

$$\begin{aligned}\min_{[\mathbf{Z}_1,\mathbf{Z}_2,\ldots,\mathbf{Z}_S]} &\sum_{i=1}^{S} \|\mathbf{Z}_i\|_* + \frac{\rho}{2}\left\|\mathbf{Y} - \mathbf{Z} - \mathbf{N} + \frac{\mathbf{\Gamma}_1}{\rho}\right\|_F^2 \\ &+ \frac{\rho}{2}\left\|\mathbf{Q} - \mathbf{Z} + \frac{\mathbf{\Gamma}_2}{\rho}\right\|_F^2. \end{aligned} \tag{7}$$

It is easy to see that each $\mathbf{Z}_i$ can be solved independently through the following problem as

$$\begin{aligned}\min_{\mathbf{Z}_i}\ &\|\mathbf{Z}_i\|_* + \frac{\rho}{2}\left\|\mathbf{Y}_i - \mathbf{Z}_i - \mathbf{N}_i + \frac{\mathbf{\Gamma}_{1,i}}{\rho}\right\|_F^2 \\ &+ \frac{\rho}{2}\left\|\mathbf{Q}_i - \mathbf{Z}_i + \frac{\mathbf{\Gamma}_{2,i}}{\rho}\right\|_F^2 \end{aligned} \tag{8}$$

where $\mathbf{\Gamma}_{1,i}$, $\mathbf{\Gamma}_{2,i}$, and $\mathbf{Q}_i$ denote the ith component of $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, and $\mathbf{Q}$, respectively. According to [45], problem (8) has a closed-form solution as

$$\mathbf{Z}_i = \mathbf{U}\mathrm{diag}\left(\mathcal{S}_{(2\rho)^{-1}}\left(\mathrm{diag}\left(\mathbf{\Sigma}\right)\right)\right)\mathbf{V}^{\mathsf{T}} \tag{9}$$

where $\mathbf{U}$, $\mathbf{\Sigma}$, $\mathbf{V}$ are derived from the singular value decomposition (SVD): $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}} = \frac{1}{2}(\mathbf{Y}_i - \mathbf{N}_i + \mathbf{Q}_i + \frac{1}{\rho}(\mathbf{\Gamma}_{1,i} + \mathbf{\Gamma}_{2,i}))$, $\mathrm{diag}(\cdot)$ transforms a diagonal matrix to a vector or transforms a vector to a diagonal matrix, and $\mathcal{S}_\varepsilon(x)$ is a soft thresholding operator, i.e., $\mathcal{S}_\varepsilon(x) = 0$ if $|x| \le \varepsilon$ and $\mathcal{S}_\varepsilon(x) = (1 - \varepsilon/|x|)x$ otherwise.

b): By fixing other variables, the subproblem to update $\mathbf{N}$ is formulated as

$$\min_{\mathbf{N}} \lambda\|\mathbf{N}\|_1 + \frac{\rho}{2}\left\|\mathbf{Y} - \mathbf{Z} - \mathbf{N} + \frac{1}{\rho}\mathbf{\Gamma}_1\right\|_F^2. \tag{10}$$

It is easy to see that problem (10) has a closed-form solution [46], [47] as

$$\mathbf{N} = \mathcal{S}_{\lambda/\rho}\left(\mathbf{Y} - \mathbf{Z} + \frac{\mathbf{\Gamma}_1}{\rho}\right). \tag{11}$$

c): By fixing other variables, the subproblem to update $\mathbf{Q}$ is formulated as

$$\min_{\mathbf{Q}} \beta \mathrm{Tr}\left(\mathbf{Q}\mathbf{G}\mathbf{Q}^{\mathrm{T}}\right) + \frac{\rho}{2}\left\|\mathbf{Q} - \mathbf{Z} + \frac{\mathbf{\Gamma}_2}{\rho}\right\|_{\mathrm{F}}^2. \tag{12}$$

By setting its gradient with respect to variable $\mathbf{Q}$ to zero, problem (12) has an analytical solution as

$$\mathbf{Q} = (\rho\mathbf{Z} - \mathbf{\Gamma}_2)[\beta(\mathbf{G}^T + \mathbf{G}) + \rho\mathbf{I}]^{-1} \tag{13}$$

where $\mathbf{I}$ denotes an identity matrix with appropriate size.

d): The Lagrange multipliers $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$ and the penalty parameter ρ are updated as

$$\begin{aligned}\mathbf{\Gamma}_1^{\mathrm{iter}+1} &= \mathbf{\Gamma}_1^{\mathrm{iter}} + \rho^{\mathrm{iter}}(\mathbf{Y} - \mathbf{Z} - \mathbf{N}) \\ \mathbf{\Gamma}_2^{\mathrm{iter}+1} &= \mathbf{\Gamma}_2^{\mathrm{iter}} + \rho^{\mathrm{iter}}(\mathbf{Q} - \mathbf{Z}) \\ \rho^{\mathrm{iter}+1} &= \min\left(\rho_{\max}, \mu\rho^{\mathrm{iter}}\right) \end{aligned} \tag{14}$$

where iter denotes the index of the iteration, and the parameter $\mu > 1$ improves the convergence rate.

The proposed LLRA-SLPG method is shown in Algorithm 1 of the Appendix, which shows the initial values of the variables and the convergence condition. After obtaining $\mathbf{Z}$, a classifier (e.g., SVM) receives it to conduct the classification.

D. Complexity and Convergence of Algorithm 1

The computational complexity of Algorithm 1 in the Appendix is mainly determined by the steps of updating $\{\mathbf{Z}_i\}_{i=1}^S$ and $\mathbf{Q}$. Specifically, the step of updating $\{\mathbf{Z}_i\}_{i=1}^S$ needs to solve the SVD whose complexity is $\mathcal{O}\left(\sum_{i=1}^S bn_i \times \min(b, n_i)\right)$. Moreover, the step of updating $\mathbf{Q}$ needs to take the inverse of a matrix with the size of $n \times n$, and calculate the matrix product between two matrices whose sizes are $b \times n$ and $n \times n$, respectively. Therefore, the computational complexity of updating $\mathbf{Q}$ is $\mathcal{O}(n^3 + bn^2)$. In summary, the total computational complexity of Algorithm 1 is $\mathcal{O}(n^3 + bn^2 + \sum_{i=1}^S bn_i \times \min(b, n_i))$ in each iteration.

Remark: It should be emphasized that as the matrix of $\beta(\mathbf{G}^T + \mathbf{G}) + \rho\mathbf{I}$ in problem (13) is very sparse, its inverse can be calculated fast in practice.

Moreover, in Algorithm 1, we use IALM to solve the objective function in an iterative manner, which is similar to the classic Expectation Maximization (EM) [48] algorithm. Moreover, it is worth pointing out that IALM has been widely used in many LRA-based methods [28], [31], [35]. The global convergence of

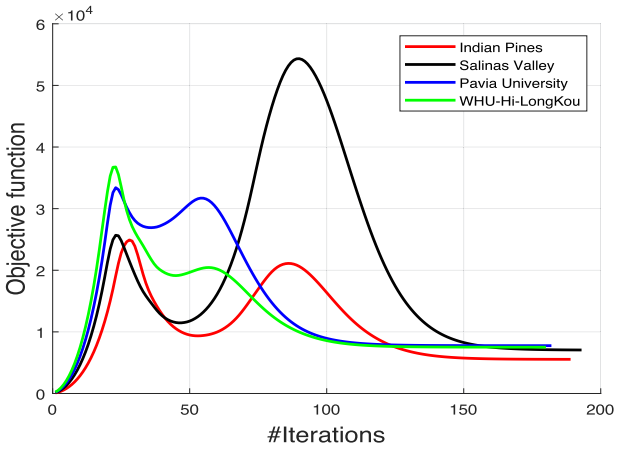

Fig. 1. Convergence of Algorithm 1 in the Appendix on all the four benchmark datasets.

TABLE I
STATISTICS OF THE *INDIAN PINES* DATASET

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

IALM has been theoretically proved when the convex problem has at most two blocks [44], [49]. For the proposed LLRA-SLPG method, where there are three blocks, to the best of our knowledge, it is still unsolved to theoretically prove the global convergence of IALM with three or more blocks [49]. Fortunately, we find that LLRA-SLPG empirically converges well on all the benchmark datasets (see Fig. 1).

IV. EXPERIMENTS

In this section, we empirically evaluate the proposed method.

A. Datasets and Experiment Settings

We use four widely used benchmark datasets[1] to evaluate the effectiveness of the proposed method. The details of these datasets are described as follows.

1) Indian Pines Dataset: This scene is acquired by Airborne Visible and InfraRed Imaging Spectrometer (AVIRIS) sensor over Indian Pines test site, consisting of 145×145 pixels with 200 spectral bands after band removal. The ground truth map contains 10249 pixels that belongs to 16 classes.

2) Salinas Valley Dataset: The AVIRIS sensor collects this scene over Salinas Valley, consisting of 512×217 pixels with 204 spectral bands after band removal. The ground truth map contains 54129 pixels that belongs to 16 classes.

3) Pavia University Dataset: The ROSIS sensor acquires this scene over Pavia, Italy, which consists of 610×340 pixels with 103 spectral bands. The ground truth map contains 42776 pixels that belongs to 9 classes.

4) WHU-Hi-LongKou Dataset: The imaging sensor acquires this scene over Longkou Town, Hubei province, China, on July 17, 2018, which consists of 550×400 pixels with 270 bands. A rectangular part (from 151 to 350 rows and 1 to 400 columns) suffering from noise heavily is used for testing. The ground truth map contains 74 474 pixels that belongs to nine classes.

[1]http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes http://rsidea.whu.edu.cn/resource_WHUHi_sharing.htm

TABLE II
STATISTICS OF THE *SALINAS VALLEY* DATASET

Class	Name	Samples
1	Brocoli_green_weeds_1	2009
2	Brocoli_green_weeds_2	3726
3	Fallow	1976
4	Fallow_rough_plow	1394
5	Fallow_smooth	2678
6	Stubble	3959
7	Celery	3579
8	Grapes_untrained	11271
9	Soil_vinyard_develop	6203
10	Corn_senesced_green_weeds	3278
11	Lettuce_romaine_4wk	1068
12	Lettuce_romaine_5wk	1927
13	Lettuce_romaine_6wk	916
14	Lettuce_romaine_7wk	1070
15	Vinyard_untrained	7268
16	Vinyard_vertical_trellis	1807

TABLE III
STATISTICS OF THE *PAVIA UNIVERSITY* DATASET

Class	Name	Samples
1	Asphalt	6631
2	Meadows	18649
3	Gravel	2099
4	Trees	3064
5	Painted metal sheets	1345
6	Bare Soil	5029
7	Bitumen	1330
8	Self-Blocking Bricks	3682
9	Shadows	947

TABLE IV
STATISTICS OF THE *WHU-HI-LONGKOU* DATASET

Class	Name	Samples
1	Corn	7583
2	Cotton	8374
3	Sesame	2497
4	Broad-leaf soybean	19988
5	Narrow-leaf soybean	4151
6	Rice	2977
7	Water	24321
8	Roads and houses	2838
9	Mixed weed	1745

TABLE V
OA(%) OF THE LLRA-SLPG METHOD WITH DIFFERENT VALUES OF λ AND β ON THE *INDIAN PINES* DATASET

		β													
		0	0.001	0.01	0.1	1	2	5	10	20	50	70	100	1000	10000
λ	0.01	91.38	91.35	91.72	92.84	93.90	93.77	93.41	92.90	91.47	90.08	90.03	90.06	90.14	90.17
	0.05	90.83	90.97	91.96	94.52	**96.58**	**96.91**	**96.96**	**96.66**	**96.07**	94.92	93.90	92.63	90.02	90.01
	0.1	79.02	79.10	80.32	89.58	**96.42**	**96.80**	**97.13**	**97.06**	**96.85**	**96.09**	95.65	95.12	90.37	90.01

Here, the percentage of training pixels per class is 5%. Values higher than 96% are shown in bold and with underline.

TABLE VI
OA(%) OF THE LLRA-SLPG METHOD WITH DIFFERENT VALUES OF λ AND β ON THE *SALINAS VALLEY* DATASET

		β													
		0	0.001	0.01	0.1	1	2	5	10	20	50	70	100	1000	10000
λ	0.01	93.07	93.25	94.29	96.59	96.90	**97.32**	**97.08**	96.76	96.27	95.42	94.40	93.85	93.27	93.27
	0.05	88.18	88.18	88.19	89.95	95.82	96.64	**97.53**	**97.91**	**98.10**	**97.64**	**97.40**	**97.14**	93.51	93.38
	0.1	88.11	88.11	88.11	88.49	93.82	95.36	96.56	**97.38**	**97.93**	**98.09**	**97.95**	**97.75**	94.18	93.38

Here the percentage of training pixels per class is 0.5%. Values higher than 97% are shown in bold and with underline.

TABLE VII
OA(%) OF THE LLRA-SLPG METHOD WITH DIFFERENT VALUES OF λ AND β ON THE *PAVIA UNIVERSITY* DATASET

		β													
		0	0.001	0.01	0.1	1	2	5	10	20	50	70	100	1000	10000
λ	0.01	83.70	84.02	85.68	**90.79**	85.24	83.62	83.18	82.61	82.13	82.38	82.00	81.25	78.82	78.34
	0.05	79.56	79.56	79.58	83.33	**88.79**	**87.93**	84.48	83.43	83.17	82.67	81.94	81.74	80.74	78.69
	0.1	79.57	79.57	79.57	80.13	**88.42**	**88.32**	85.84	84.14	83.34	83.13	83.05	82.76	81.78	78.98

Here the percentage of training pixels per class is 0.2%. Values higher than 87% are shown in bold and with underline.

TABLE VIII
OA(%) OF THE LLRA-SLPG METHOD WITH DIFFERENT VALUES OF λ AND β ON THE *WHU-HI-LONGKOU* DATASET

		β													
		0	0.001	0.01	0.1	1	2	5	10	20	50	70	100	1000	10000
λ	0.01	94.87	94.93	96.04	**98.11**	**98.31**	**98.42**	**98.39**	**98.14**	97.76	97.57	97.51	97.44	97.40	97.38
	0.05	93.73	93.73	93.75	94.93	**98.24**	**98.06**	**98.03**	**98.20**	**98.29**	**98.25**	**98.16**	**98.17**	97.39	97.38
	0.1	93.64	93.64	93.64	93.34	97.72	**98.23**	**98.06**	**98.07**	**98.15**	**98.30**	**98.29**	**98.24**	97.48	97.38

Here the percentage of training pixels per class is 1%. Values higher than 98% are shown in bold and with underline.

TABLE IX
CLASSIFICATION RESULTS OF THE LLRA-SLPG METHOD WITH DIFFERENT VALUES OF r ON ALL THE FOUR DATASETS

Indian Pines					*Salinas Valley*				
r	1	2	3	4	r	1	2	3	4
OA (%)	**97.18**	97.13	97.04	96.87	OA (%)	97.93	98.10	**98.14**	98.13
AA (%)	96.52	96.74	**96.78**	96.75	AA (%)	97.43	97.43	97.46	**97.51**
κ (%)	**96.79**	96.73	96.62	96.43	κ (%)	97.69	97.88	**97.93**	97.92
Pavia University					*WHU-Hi-LongKou*				
r	1	2	3	4	r	1	2	3	4
OA (%)	89.07	**90.79**	89.02	89.90	OA (%)	98.26	98.42	98.53	**98.57**
AA (%)	81.06	85.63	80.92	**87.01**	AA (%)	95.46	95.82	96.02	**96.15**
κ (%)	85.37	**87.69**	85.31	86.57	κ (%)	97.79	98.00	98.14	**98.19**

Here the percentage of training pixels per class is 5%, 0.5%, 0.2%, and 1% for the *Indian Pines*, *Salinas Valley, Pavia University*, and *WHU-Hi-LongKou* datasets, respectively. Best results are shown in bold.

TABLE X
CLASSIFICATION RESULTS FOR ABLATION STUDY ON ALL THE FOUR DATASETS

Indian Pines			*Salinas Valley*			*Pavia University*			*WHU-Hi-LongKou*		
	M1	Proposed		M1	Proposed		M1	Proposed		M1	Proposed
OA (%)	91.38	**97.18**	OA (%)	93.07	**98.14**	OA (%)	83.70	**90.79**	OA (%)	94.87	**98.57**
AA (%)	87.57	**96.52**	AA (%)	93.90	**97.46**	AA (%)	80.60	**85.63**	AA (%)	89.54	**96.15**
κ (%)	90.17	**96.79**	κ (%)	92.28	**97.93**	κ (%)	78.26	**87.69**	κ (%)	93.50	**98.19**

Here the percentage of training pixels per class is 5%, 0.5%, 0.2%, and 1% for the *Indian Pines, Salinas Valley, Pavia University*, and *WHU-Hi-LongKou* datasets, respectively. Best results are shown in bold.

TABLE XI
COMPARISON AMONG ALL THE METHODS IN TERMS OF THE CLASSIFICATION ACCURACY ON THE *INDIAN PINES* DATASET, IN WHICH THE TRAINING PERCENTAGE PER CLASS EQUALS 5%

Class	Training	Testing	Original	RPCA[51]	LLRSSTV[52]	S^3LRR[28]	SS-LRR[30]	LSSTRPCA[35]	LPGTRPCA[53]	OLRT[54]	Proposed
1	3	43	53.02	70.93	78.83	77.91	93.72	82.33	72.79	6.05	**100.00**
2	72	1356	72.41	80.30	80.47	94.62	86.67	82.91	94.20	75.11	**95.12**
3	42	788	62.78	75.60	71.05	89.64	88.81	83.95	92.31	79.82	**95.98**
4	12	225	43.29	54.00	55.68	85.86	66.44	67.20	**92.49**	44.13	92.36
5	25	458	88.95	90.98	84.45	91.11	93.69	93.10	91.07	82.75	**97.25**
6	37	693	94.04	94.66	93.10	97.94	98.31	**99.25**	96.81	91.96	98.27
7	2	26	78.85	84.62	70.38	61.15	93.85	88.85	93.46	45.77	**96.15**
8	24	454	94.56	96.43	97.00	97.29	99.41	98.39	98.92	94.01	**100.00**
9	1	19	16.84	51.05	18.42	51.05	77.89	60.53	55.26	33.16	**98.42**
10	49	923	66.13	74.96	74.63	87.46	88.72	79.91	88.98	72.51	**95.77**
11	123	2332	74.66	82.49	88.95	96.55	92.27	86.90	94.87	86.30	**98.58**
12	30	563	57.67	75.03	61.49	83.02	79.86	75.06	89.73	57.69	**93.30**
13	11	194	94.43	96.39	90.05	90.88	99.12	98.09	98.09	96.19	**99.48**
14	64	1201	90.10	91.66	98.00	96.97	97.57	96.21	97.65	92.91	**99.92**
15	20	366	43.14	46.64	81.03	**98.39**	76.31	78.39	92.90	74.73	97.40
16	5	88	87.73	88.86	89.31	81.93	90.11	89.66	**97.73**	42.61	86.25
OA	-	-	74.88	81.75	83.68	93.32	90.54	87.02	93.99	80.49	**97.18**
AA	-	-	69.91	78.41	77.06	86.36	88.92	85.05	90.45	67.23	**96.52**
κ	-	-	71.34	79.19	81.40	92.39	89.21	85.20	93.15	77.73	**96.79**

The best result and the second-best result are shown in bold and with underline, respectively. The "original" item represents the result of applying the SVM classifier to original spectral features of an HSI directly.

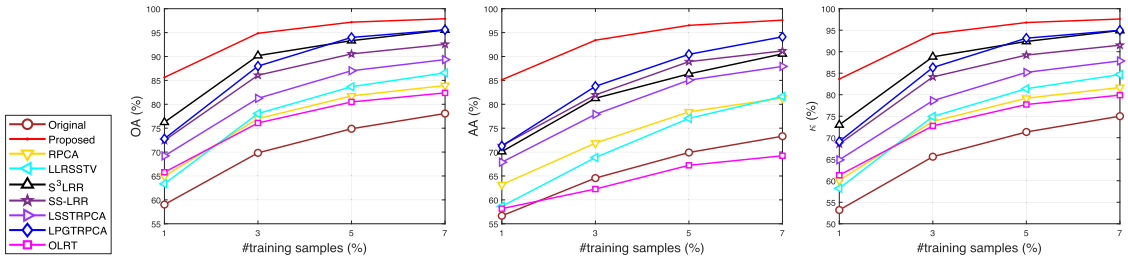

Fig. 2. Comparison among all the methods in terms of the classification accuracy under different percentages of training samples on the *Indian Pines* dataset.

More details for all the four benchmark datasets are given in Tables I–IV. The numbers of superpixels are empirically set as the suggested values in [31] for the four benchmark datasets, i.e., 64 (*Indian Pines*), 50 (*Salinas Valley*), 50 (*Pavia University*), and 64 (*WHU-Hi-LongKou*).

Three widely used evaluation criteria, i.e., overall accuracy (OA), average accuracy (AA), and Kappa coefficient (κ), are used to measure the classification results. For all the LRA-based methods, after learning the recovered representation, the SVM classifier equipped with a RBF kernel is used as the classifier. Specifically, our proposed LLRA-SLPG method uses the max normalization method [50] as a preprocessing.

B. Parameter and Convergence Analysis

In this section, we investigate how the three hyperparameters of the proposed LLRA-SLPG method, i.e., λ, β, and r, affect the classification performance. Note that λ depends on the severity of the noises in an HSI, β controls to preserve the local manifold structure, and r reflects the number of neighbors for the locality preserving graph. Following [31], candidate values of λ are in $\{0.01, 0.05, 0.1\}$. β and r are selected from $\{0, 0.001, 0.01, 0.1, 1, 2, 5, 10, 20, 50, 70, 100, 1000, 10000\}$ and $\{1, 2, 3, 4\}$, respectively.

1) Tuning of Regularization Parameters λ and β: In this part, we study how the parameters λ and β affect the performance of the proposed LLRA-SLPG method. The radius r of the locality preserving graph is fixed at 2. Tables V–VIII show the OA of the LLRA-SLPG method with various values of λ and β on the four benchmark datasets.

For the *Indian Pines* dataset, as shown in Table V, the proposed LLRA-SLPG method achieves the highest OA (97.13%) with $\lambda = 0.1$ and $\beta = 5$. It can be observed that, with different values of λ, the corresponding highest OAs may not change too much, that is, with $\lambda = 0.05$ and 0.1, the corresponding highest OA equals 96.96% and 97.13%, respectively. According to the results, we can see that values of β that result in OA larger than 96.00% are in the interval [1,50].

For the *Salinas Valley* dataset, as shown in Table VI, the LLRA-SLPG method achieves the highest OA (98.10%) when $\lambda = 0.05$ and $\beta = 20$. With different values of λ, the corresponding best OAs are very close, that is, when λ equals 0.05 and 0.1, the best OA is 98.10% and 98.09%, respectively. Values of β that result in OA larger than 97% are in the interval [2,100].

For the *Pavia University* dataset, as shown in Table VII, the LLRA-SLPG method achieves the highest OA (90.79%) with $\lambda = 0.01$ and $\beta = 0.1$. Values of β that result in OA larger than 87% are in the interval [0.1,2].

For the *WHU-Hi-LongKou* dataset, as shown in Table VIII, the LLRA-SLPG method achieves the highest OA (98.42%) with

TABLE XII
COMPARISON AMONG ALL THE METHODS IN TERMS OF THE CLASSIFICATION ACCURACY ON THE *SALINAS VALLEY* DATASET, IN WHICH THE TRAINING PERCENTAGE PER CLASS EQUALS 0.5%

Class	Training	Testing	Original	RPCA[51]	LLRSSTV[52]	S^3LRR[28]	SS-LRR[30]	LSSTRPCA[35]	LPGTRPCA[53]	OLRT[54]	Proposed
1	11	1998	96.66	95.24	96.40	93.59	97.02	95.58	95.18	94.35	**100.00**
2	19	3707	98.36	98.35	98.87	95.45	98.52	98.53	98.20	95.95	**100.00**
3	10	1966	90.09	89.80	90.34	90.33	90.46	87.92	94.52	93.99	**99.87**
4	7	1387	**98.08**	97.84	96.65	84.62	97.89	96.87	96.36	97.50	92.73
5	14	2664	93.67	93.90	92.08	87.81	93.97	91.40	94.05	90.99	**94.28**
6	20	3939	99.18	99.15	99.53	98.47	99.22	99.53	99.23	99.61	**99.92**
7	18	3561	99.30	99.28	97.93	95.89	99.33	98.66	98.94	98.39	**99.61**
8	57	11214	80.49	81.87	96.33	92.60	86.23	94.18	92.78	93.52	**98.48**
9	32	6171	98.17	97.98	98.56	98.65	98.17	98.89	98.66	99.31	**99.97**
10	17	3261	85.28	84.39	81.12	89.26	86.19	87.48	89.15	84.93	**92.16**
11	6	1062	86.81	85.96	87.64	77.39	86.53	85.36	90.78	90.89	**93.08**
12	10	1917	99.26	99.08	97.98	81.71	**99.31**	97.56	97.71	92.38	99.07
13	5	911	97.65	97.55	90.93	77.30	97.59	96.51	**98.24**	78.63	**98.24**
14	6	1064	90.12	90.08	87.61	87.81	90.10	92.27	91.44	82.33	**94.39**
15	37	7231	63.82	65.49	96.17	93.55	73.13	92.42	91.90	91.90	**98.31**
16	10	1797	91.55	90.30	92.22	88.43	91.47	89.53	92.19	97.95	**99.24**
OA	-	-	88.11	88.41	95.22	92.39	90.65	94.61	94.86	93.99	**98.14**
AA	-	-	91.78	91.64	93.77	89.55	92.82	93.92	94.96	92.66	**97.46**
κ	-	-	86.75	87.09	94.69	91.53	89.58	94.00	94.28	93.31	**97.93**

The best result and the second-best result are shown in bold and with underline, respectively. The "original" item represents the result of applying the SVM classifier to original spectral features of an HSI directly.

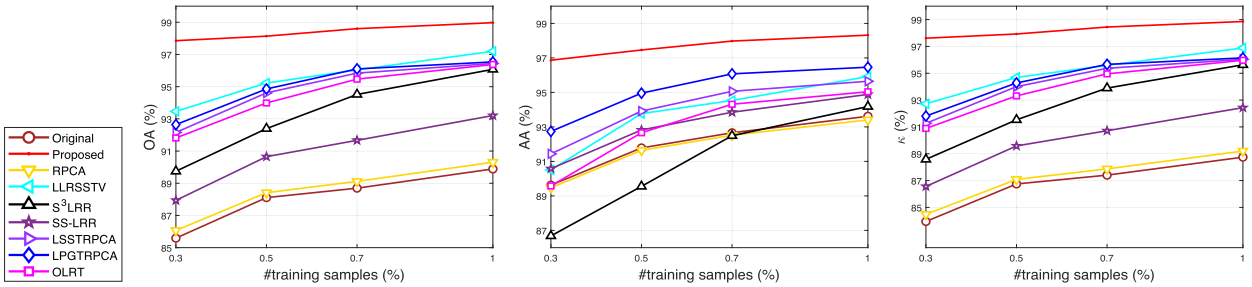

Fig. 3. Comparison among all the methods in terms of the classification accuracy under different percentages of training samples on the *Salinas Valley* dataset.

TABLE XIII
COMPARISON AMONG ALL THE METHODS IN TERMS OF THE CLASSIFICATION ACCURACY ON THE *PAVIA UNIVERSITY* DATASET, IN WHICH THE TRAINING PERCENTAGE PER CLASS EQUALS 0.2%

Class	Training	Testing	Original	RPCA[51]	LLRSSTV[52]	S^3LRR[28]	SS-LRR[30]	LSSTRPCA[35]	LPGTRPCA[53]	OLRT[54]	Proposed
1	14	6617	76.09	76.28	76.11	79.53	75.37	78.56	77.44	70.56	**91.79**
2	38	18611	91.88	92.24	95.85	91.26	93.12	94.14	95.19	94.86	**97.68**
3	5	2094	51.45	51.73	70.43	60.08	56.06	56.37	60.07	68.26	**85.73**
4	7	3057	74.17	73.22	70.74	70.13	75.11	**86.13**	83.91	50.14	72.08
5	3	1342	88.89	89.26	88.26	**91.21**	88.95	89.12	88.81	87.66	89.92
6	11	5018	54.52	55.33	84.50	71.49	61.90	66.23	66.36	83.89	**85.82**
7	3	1327	68.39	68.49	76.32	53.72	73.47	69.41	**79.92**	58.05	79.10
8	8	3674	73.56	73.21	80.81	66.34	70.85	74.13	76.05	70.22	**85.91**
9	2	945	**99.75**	**99.75**	99.58	99.46	**99.75**	99.66	99.72	40.98	82.59
OA	-	-	79.56	79.78	86.35	80.94	81.08	83.49	84.29	80.61	**90.79**
AA	-	-	75.41	75.50	82.51	75.91	77.17	79.31	80.83	69.40	**85.63**
κ	-	-	72.60	72.86	81.85	74.60	74.69	77.98	79.02	74.11	**87.69**

The best result and the second-best result are shown in bold and with underline, respectively. The "original" item represents the result of applying the SVM classifier to original spectral features of an HSI directly.

$\lambda = 0.01$ and $\beta = 2$. Values of β that result in OA larger than 98% are in the interval [0.1,100].

On all four datasets, with a fixed value of λ, the OA first increases and then decreases as β increases. For example, on the *Indian Pines* dataset, with $\lambda = 0.1$, the OA steadily increases from 79.02% to 97.13% as β changes from 0 to 5, and decreases to 90.01% as β increases from 5 to 10 000. Besides, when β varies in a specific range on those datasets (e.g., $0.01 \leq \beta \leq 20$ on the *Indian Pines* dataset), the LLRA-SLPG method performs consistently better than $\beta = 0$, which validates the effectiveness of the third term in problem (4) (i.e., $\beta \mathrm{Tr}(\mathbf{ZGZ}^{\mathrm{T}})$) to improve the classification performance.

2) Influence of Radius r of the Superpixel-Guided Locality Preserving Graph: In this part, we study the effect of the radius r. Table IX shows the classification results for the proposed LLRA-SLPG method with different values of r on the four benchmark datasets. Specifically, with the value of r fixed, we list the corresponding best classification result of LLRA-SLPG under different values of λ and β.

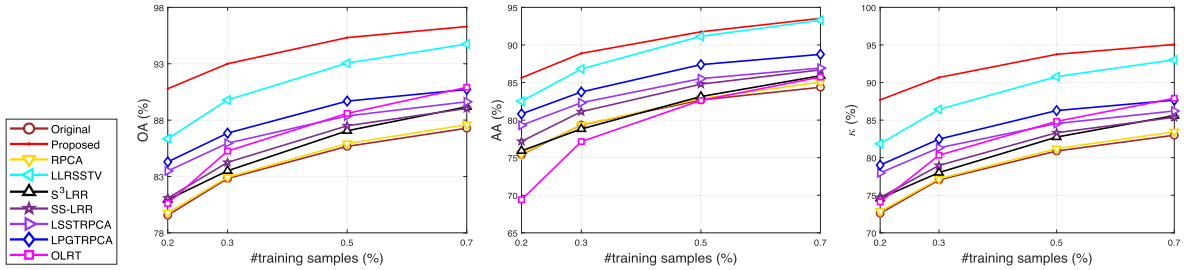

Fig. 4. Comparison among all the methods in terms of the classification accuracy under different percentages of training samples on the *Pavia University* dataset.

TABLE XIV
COMPARISON AMONG ALL THE METHODS IN TERMS OF THE CLASSIFICATION ACCURACY ON THE *WHU-HI-LONGKOU* DATASET, IN WHICH THE TRAINING PERCENTAGE PER CLASS EQUALS 1%

Class	Training	Testing	Original	RPCA[51]	LLRSSTV[52]	S^3LRR[28]	SS-LRR[30]	LSSTRPCA[35]	LPGTRPCA[53]	OLRT[54]	Proposed
1	76	7507	97.30	97.30	97.50	99.18	97.42	97.95	98.77	97.37	**99.54**
2	84	8290	88.75	88.75	93.27	96.97	90.56	91.35	98.18	96.35	**99.28**
3	25	2472	82.21	82.22	90.06	94.55	85.85	85.90	95.26	93.93	**99.96**
4	200	19788	94.11	94.11	94.75	97.99	94.86	95.22	98.57	97.08	**98.58**
5	42	4109	84.19	84.19	85.61	86.76	85.35	85.68	94.63	87.98	**98.20**
6	30	2947	97.02	97.02	96.15	99.32	97.09	97.17	96.93	98.26	**99.48**
7	244	24077	**99.98**	**99.98**	99.97	99.72	99.97	**99.98**	99.84	99.67	99.95
8	29	2809	78.46	78.46	81.42	85.16	77.98	80.77	87.99	77.56	**93.84**
9	18	1727	64.80	64.79	40.20	72.88	66.29	66.00	**85.23**	64.61	76.47
OA	-	-	93.63	93.63	94.17	96.80	94.25	94.61	97.85	95.80	**98.57**
AA	-	-	87.42	87.43	86.55	92.50	88.37	88.89	95.05	90.31	**96.15**
κ	-	-	91.92	91.93	92.62	95.94	92.72	93.18	97.28	94.68	**98.19**

The best result and the second-best result are shown in bold and with underline, respectively. The "original" item represents the result of applying the SVM classifier to original spectral features of an HSI directly.

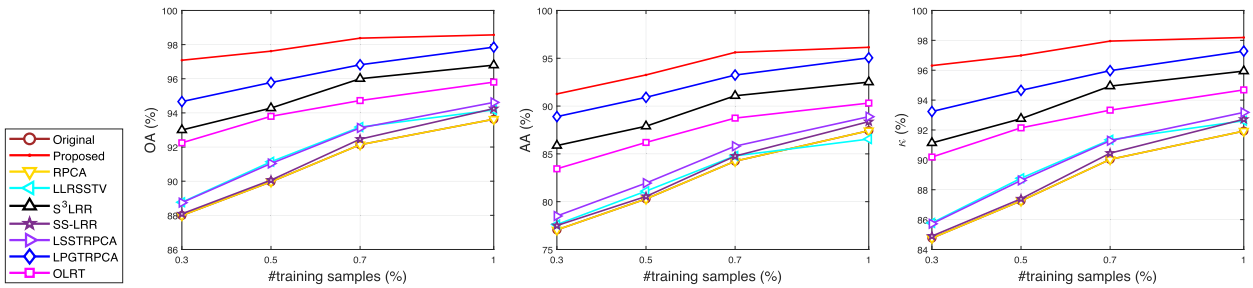

Fig. 5. Comparison among all the methods in terms of the classification accuracy under different percentages of training samples on the *WHU-Hi-LongKou* dataset.

TABLE XV
COMPARISON OF RUNNING TIME (IN SECONDS) FOR ALL THE METHODS ON THE FOUR BENCHMARK DATASETS

	P	RPCA[51]	LLRSSTV[52]	S^3LRR[28]	SS-LRR[30]	LSSTRPCA[35]	LPGTRPCA[53]	OLRT[54]	Proposed
Indian Pines	5%	**56.37**	92.90	292.63	144.40	203.29	166.58	843.60	177.92
Salinas Valley	0.5%	343.79	466.52	1250.60	**319.65**	862.54	1960.57	3793.64	1046.90
Pavia University	0.2%	259.33	513.75	1533.30	**241.84**	960.11	4376.81	4059.19	1093.20
WHU-Hi-LongKou	1%	274.48	393.04	1235.49	**164.74**	918.68	1283.86	3779.87	1094.34

The best result and the second-best result are shown in bold and with underline, respectively.

As shown in Table IX, the performance of the LLRA-SLPG method on the *Indian Pines*, *Salinas Valley* and *WHU-Hi-LongKou* datasets is insensitive to r. For example, on the *Salinas Valley* dataset, the highest OA (98.14%) at $r = 3$ is only 0.21% larger than the lowest OA (97.93%) at $r = 1$. Differently, the performance of the LLRA-SLPG method on the *Pavia University* dataset fluctuates as r increases, that is, the OA values obtained by the LLRA-SLPG method are 89.07%, 90.79%, 89.02%, and 89.90% for $r = 1$, 2, 3, and 4, respectively.

Accordingly, in the subsequent experiments, the parameters r, λ, and β are set to the corresponding values which could achieve the best performance (i.e., the highest OA) according to Table IX. Specifically, r is set to 1, 3, 2 and 4 for the *Indian Pines*, *Salinas Valley*, *Pavia University*, and *WHU-Hi-LongKou* datasets, respectively.

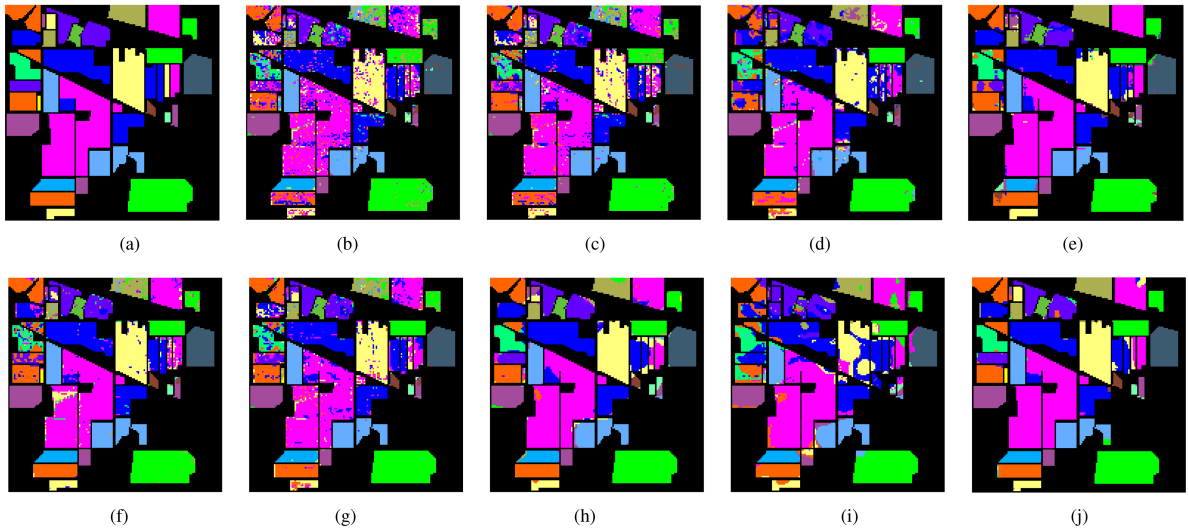

Fig. 6. Classification maps of tested methods on the *Indian Pines* dataset with $P = 5\%$ (the OAs are reported in the parentheses). (a) Groundtruth. (b) Original (76.65%). (c) RPCA (82.85%). (d) LLRSSTV (83.51%). (e) S3LRR (93.38%). (f) SS-LRR. (90.66%). (g) LSSTRPCA (87.02%). (h) LPGTRPCA (94.17%). (i) OLRT (79.37%). (j) Proposed (97.04%).

3) Ablation Study for the Superpixel-Guided Locality Preserving Graph: In this part, an ablation study is conducted to study the effectiveness of the superpixel-guided locality preserving graph. First, we define the objective function in problem (15), namely M1, which removes the superpixel-guided locality preserving graph term (i.e., $\beta \mathrm{Tr}(\mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}})$) in problem (15). The candidate values of λ for M1 are the same as the proposed LLRA-SLPG method. The comparison results on all the four datasets are shown in Table X, which can be seen that the performance of the proposed LLRA-SLPG method achieves much better performance than the M1 method on all the evaluation metrics (i.e., OA, AA, and κ). Such an observation validates that the superpixel-guided locality preserving graph term (i.e., $\beta \mathrm{Tr}(\mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathrm{T}})$) can significantly enhance the classification performance

$$\begin{aligned} &\min_{[\mathbf{Z},\mathbf{N}]} \sum_{i=1}^{S} \|\mathbf{Z}_i\|_* + \lambda \|\mathbf{N}\|_1 \\ &\text{s.t. } \mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_S],\ \mathbf{N} = [\mathbf{N}_1, \mathbf{N}_2, \ldots, \mathbf{N}_S] \\ &\quad \mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_S],\ \mathbf{Y} = \mathbf{Z} + \mathbf{N}. \end{aligned} \tag{15}$$

4) Convergence Analysis: As shown in Fig. 1, we can see that problem (6) converges after about 150 iterations on the *Indian Pines* and *Salinas Valley* datasets. In addition, on the *Pavia University* and *WHU-Hi-LongKou* datasets, problem (6) converges faster (i.e., in about 100 iterations). All of those results validate the convergence of the proposed LLRA-SLPG method.

C. Comparison With State-of-the-Art Methods

We compare the proposed LLRA-SLPG method with state-of-the-art HSI classification methods, including seven LRA-based methods for HSIs, i.e., RPCA [51], LLRSSTV [52], S^3 LRR [28], SS-LRR [30], LSSTRPCA [35], LPGTRPCA [53], OLRT [54].

We adopt the suggested settings in the original papers for all those baseline methods. We repeat the experiments for ten times with randomly sampled training pixels for all the methods and report the average results.

According to results on the *Indian Pines* dataset shown in Table XI where the training percentage per class equals 5%, we can see that the LLRA-SLPG method significantly outperforms baseline methods. For example, the proposed LLRA-SLPG method achieves the highest accuracies (i.e., OA (97.18%), AA (96.52%), and κ (96.79%)), which is much better than the second-best one (i.e., OA (93.99%), AA (90.45%), and κ (93.15%)). Moreover, the proposed LLRA-SLPG method achieves the best performance in 12 out of 16 classes, especially for classes 1, 2, 3, 5, 7, 9, 10, 11, 12, 14. For classes with limited samples, i.e., classes 1, 7, and 9, the LLRA-SLPG method achieves the best performance and shows a remarkable margin over the second-best one. We also compare the classification performance for all the methods with various training percentages per class, i.e., $P = 1\%, 3\%, 5\%$, and 7%. As shown in Fig. 2, the proposed LLRA-SLPG method is consistently better than baseline methods in OA, AA, and κ. The classification accuracies of all the methods gradually improve as the training percentage increases. Moreover, our method performs significantly better than the compared ones when the number of training pixels is extremely small, e.g., 1% and 3%.

Table XII shows the performance comparison on the *Salinas Valley* dataset, where the training percentage per class is set to 0.5%. We can observe that the proposed LLRA-SLPG method also achieves the highest values in all the accuracy metrics. Specifically, the proposed LLRA-SLPG method achieves the best performance in most categories, i.e., the best performance for 14 out of 16 classes. Especially for the classes 1, 2, 3, 8, 10, 11, 14, 15, and 16, the proposed LLRA-SLPG method shows remarkable margins over baseline methods. Furthermore, as illustrated in Fig. 3, the proposed LLRA-SLPG method performs well under various training percentages, which shows its

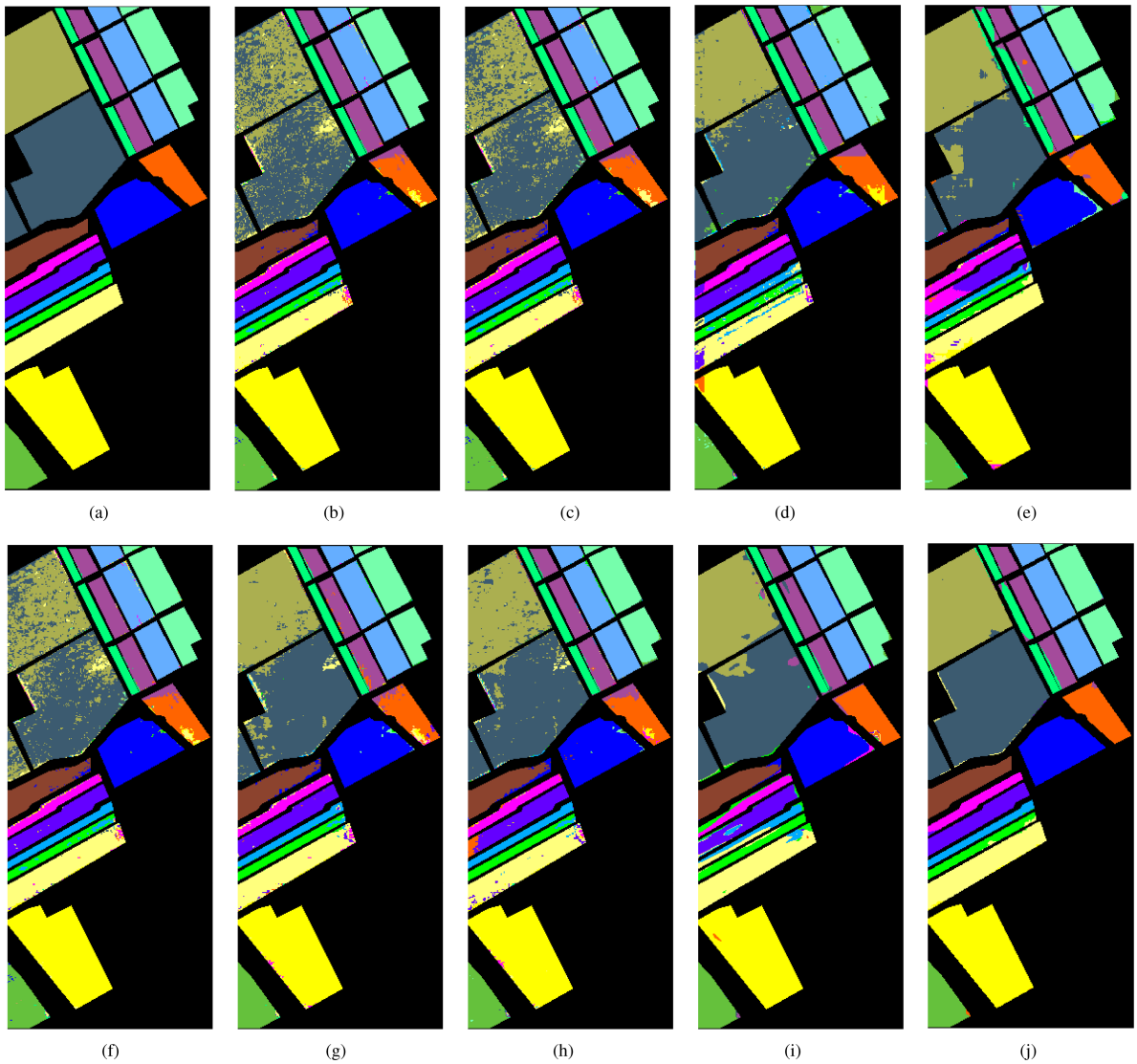

Fig. 7. Classification maps of tested methods on the *Salinas Valley* dataset with $P = 0.5\%$ (the OAs are reported in the parentheses).(a) Groundtruth. (b) Original (87.79%). (c) RPCA (88.21%). (d) LLRSSTV (94.03%). (e) S^3LRR (92.98%). (f) SS-LRR (89.49%). (g) LSSTRPCA (94.64%). (h) LPGTRPCA (94.10%). (i) OLRT (94.36%). (j) Proposed (98.79%).

superiority. It can be observed that the classification performance of all the methods gradually improves with the training percentage increasing.

Table XIII shows the performance of all the methods in comparison on the *Pavia University* dataset, in which the training percentage per class equals 0.2%. According to the results, we can see that the proposed LLRA-SLPG method achieves better performance. Specifically, the proposed LLRA-SLPG method performs remarkably better than baseline methods, especially in classes 1, 2, 3, 6, and 8. In addition, as shown in Fig. 4, the proposed LLRA-SLPG method consistently performs better than baseline methods under different training percentages, especially for small training percentages, e.g., 0.2%.

Table XIV shows the performance comparison on the *WHU-Hi-LongKou* dataset, in which the training percentage per class is 1%. We can observe that the proposed LLRA-SLPG method outperforms the other methods in OA, AA, and κ. Specifically, the proposed LLRA-SLPG method achieves the best performance in 7 out of 9 classes, especially for classes 3, 5, 8. Moreover, as shown in Fig. 5, the proposed LLRA-SLPG method shows significant superiority over other methods on different training percentages, especially for small training percentages, e.g., 0.3% and 0.5%.

The classification maps of all the methods on the four benchmark datasets are given in Figs. 6–9, which further manifest the advantage of the proposed LLRA-SLPG method.

D. Comparison of Running Time

Table XV shows the running time of all methods on the four datasets, where all the methods are running on a laptop with i5-8500 CPU @ 3.00GHz. Specifically, the training percentages

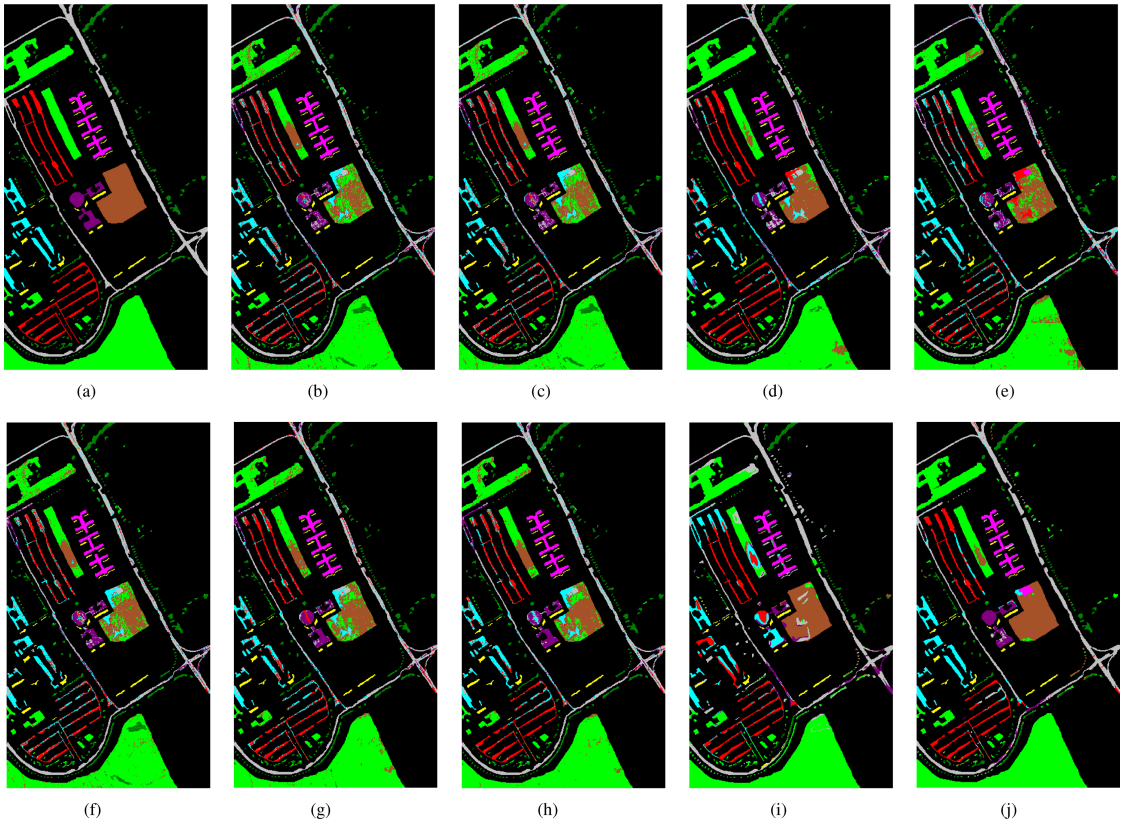

Fig. 8. Classification maps of tested methods on the *Pavia University* dataset with $P = 0.2\%$ (the OAs are reported in the parentheses). (a) Groundtruth. (b) Original (79.75%). (c) RPCA (79.84%). (d) LLRSSTV (86.22%). (e) S^3LRR (80.49%). (f) SS-LRR (80.98%). g) LSSTRPCA (83.65%). (h) LPGTRPCA (84.87%). (i) OLRT (80.39%). (j) Proposed (92.64%).

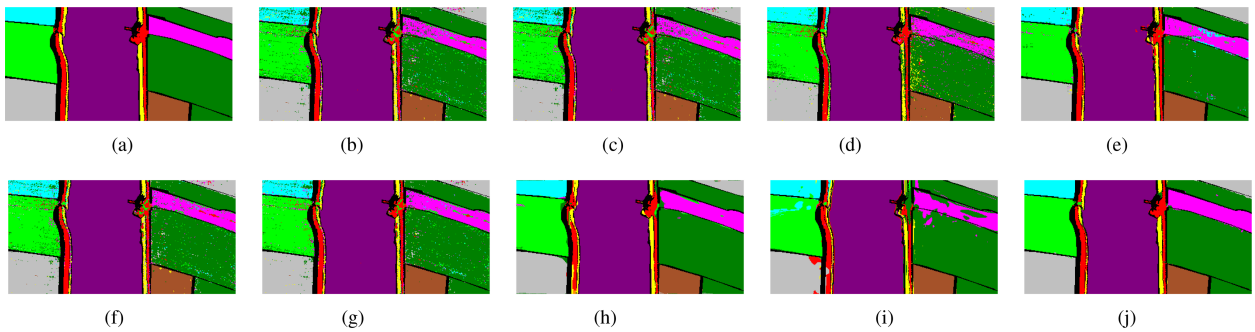

Fig. 9. Classification maps of tested methods on the *WHU-Hi-LongKou* dataset with $P = 1\%$ (the OAs are reported in the parentheses). (a) Groundtruth. (b) Original (93.13%). (c) RPCA (93.14%). (d) LLRSSTV (94.20%). (e) S^3LRR (97.20%) (f) SS-LRR (94.28%). (g) LSSTRPCA (94.85%). (h) LPGTRPCA (97.71%). (i) OLRT (93.90%). (j) Proposed (99.06%).

per class of the four datasets (i.e., *Indian Pines*, *Salinas Valley*, *Pavia University*, and *WHU-Hi-LongKou*) are 5%, 0.5%, 0.2%, and 1%, respectively. It can be seen that the proposed LLRA-SLPG method has a reasonable running time compared with baseline methods.

V. Conclusion

In this article, we propose the LLRA-SLPG method to improve the discriminative representation of pixels in an HSI. The LLRA-SLPG method can reduce the spectral variations and promote the local manifold structure to improve the discriminability of features in the low-rank component. Experiments on four benchmark datasets demonstrate that the proposed LLRA-SLPG method outperforms state-of-the-art methods. In addition, the proposed LLRA-SLPG method shows remarkable performance improvement with extremely few training samples. In the future, we would like to improve the efficiency of the proposed LLRA-SLPG method by using Anderson acceleration [55], and adaptively determine the value of parameters, e.g. using maximum a posteriori (MAP) to estimate them [56], [57], [58], [59] according to the characteristic of input samples.

APPENDIX

A. Pseudocode of the Algorithm for the Proposed LLRA-SLPG Method

Algorithm 1: Algorithm for the LLRA-SLPG Model.

Input: $\mathbf{Y}$, $\mathbf{Y}_i$, Laplacian graph $\mathbf{G}$, and parameters λ, β, S.
Output: $\mathbf{Z}$, $\mathbf{N}$.

1: **Initialize:** $\mathbf{Z} = \mathbf{N} = \mathbf{Q} = \mathbf{\Gamma}_1 = \mathbf{\Gamma}_2 = \mathbf{0}$, $\rho = 10^{-4}$, $\rho_{\max} = 10^{12}$, $\epsilon = 10^{-6}$, $\mu = 1.1$;
2: **while** not converged **do**
3: Update $\{\mathbf{Z}_i\}_{i=1}^{S}$ via (9);
4: Update $\mathbf{N}$ via (11);
5: Update $\mathbf{Q}$ via (13);
6: Update $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, and ρ via (14);
7: Check the convergence condition: $\|\mathbf{Y} - \mathbf{Z} - \mathbf{N}\|_\infty \leq \epsilon$ and $\|\mathbf{Z} - \mathbf{Q}\|_\infty \leq \epsilon$
8: **end while**

REFERENCES

[1] A. Plaza et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, 2009.
[2] B. Hörig, F. Kühn, F. Oschütz, and F. Lehmann, "Hymap hyperspectral remote sensing to detect hydrocarbons," *Int. J. Remote Sens.*, vol. 22, no. 8, pp. 1413–1422, 2001.
[3] N. Patel, C. Patnaik, S. Dutta, A. Shekh, and A. Dave, "Study of crop growth parameters using airborne imaging spectrometer data," *Int. J. Remote Sens.*, vol. 22, no. 12, pp. 2401–2411, 2001.
[4] B. Datt, T. R. McVicar, T. G. Van Niel, D. L. Jupp, and J. S. Pearlman, "Preprocessing EO-1 hyperion hyperspectral data to support the application of agricultural indexes," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1246–1259, Jun. 2003.
[5] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing: Principles and Applications*. Boca Raton, FL, USA: CRC Press, 2007.
[6] C. Weber et al., "Hyperspectral imagery for environmental urban planning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 1628–1631.
[7] M. Moroni, E. Lupo, E. Marra, and A. Cenedese, "Hyperspectral image analysis in environmental monitoring: Setup of a new tunable filter platform," *Procedia Environ. Sci.*, vol. 19, pp. 885–894, 2013.
[8] M. B. Stuart, A. J. McGonigle, and J. R. Willmott, "Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems," *Sensors*, vol. 19, no. 14, 2019, Art. no. 3071.
[9] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.
[10] H. Wang, Y. Cen, Z. He, R. Zhao, Y. Cen, and F. Zhang, "Robust generalized low-rank decomposition of multimatrices for image recovery," *IEEE Trans. Multimedia*, vol. 19, no. 5, pp. 969–983, May 2017.
[11] S. Wang, Y. Chen, Y. Jin, Y. Cen, Y. Li, and L. Zhang, "Error-robust low-rank tensor approximation for multi-view clustering," *Knowl.-Based Syst.*, vol. 215, 2021, Art. no. 106745.
[12] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 193–200.
[13] J. M. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
[14] S. Mei, Q. Bi, J. Ji, J. Hou, and Q. Du, "Spectral variation alleviation by low-rank matrix approximation for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 796–800, Jun. 2016.
[15] W. Sun and Q. Du, "Graph-regularized fast and robust principal component analysis for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3185–3195, Jun. 2018.
[16] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
[17] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.
[18] A. Sumarsono and Q. Du, "Low-rank subspace representation for supervised and unsupervised classification of hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4188–4195, Sep. 2016.
[19] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 911–923, Feb. 2019.
[20] Y. Wang et al., "Self-supervised low-rank representation (SSLRR) for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5658–5672, Oct. 2018.
[21] J. Mei et al., "PSASL: Pixel-level and superpixel-level aware subspace learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4278–4293, Jul. 2019.
[22] J. Xu, J. E. Fowler, and L. Xiao, "Hypergraph-regularized low-rank subspace clustering using superpixels for unsupervised spatial-spectral hyperspectral classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 871–875, May 2021.
[23] C. Xing, M. Wang, Z. Wang, C. Duan, and Y. Liu, "Diagonalized low-rank learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2021, Art. no. 5507812.
[24] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.
[25] R. Zhu, M. Dong, and J.-H. Xue, "Spectral nonlocal restoration of hyperspectral images with low-rank property," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3062–3067, Jun. 2015.
[26] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.
[27] S. Mei, Q. Bi, J. Ji, J. Hou, and Q. Du, "Hyperspectral image classification by exploring low-rank property in spectral or/and spatial domain," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2910–2921, Jun. 2017.
[28] S. Mei, J. Hou, J. Chen, L.-P. Chau, and Q. Du, "Simultaneous spatial and spectral low-rank representation of hyperspectral images for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2872–2886, May 2018.
[29] Y. Xu, Z. Wu, and Z. Wei, "Spectral–spatial classification of hyperspectral image based on low-rank decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2370–2380, Jun. 2015.
[30] F. Fan, Y. Ma, C. Li, X. Mei, J. Huang, and J. Ma, "Hyperspectral image denoising with superpixel segmentation and low-rank representation," *Inf. Sci.*, vol. 397, pp. 48–68, 2017.
[31] S. Yang, J. Hou, Y. Jia, S. Mei, and Q. Du, "Superpixel-guided discriminative low-rank representation of hyperspectral images for classification," *IEEE Trans. Image Process.*, vol. 30, pp. 8823–8835, 2021.
[32] J. An, X. Zhang, H. Zhou, and L. Jiao, "Tensor-based low-rank graph with multimanifold regularization for dimensionality reduction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4731–4746, Aug. 2018.
[33] Y.-J. Deng, H.-C. Li, K. Fu, Q. Du, and W. J. Emery, "Tensor low-rank discriminant embedding for hyperspectral image dimensionality reduction," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7183–7194, Dec. 2018.
[34] Y.-J. Deng, H.-C. Li, X. Song, Y.-J. Sun, X.-R. Zhang, and Q. Du, "Patch tensor-based multigraph embedding framework for dimensionality reduction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1630–1643, Mar. 2020.
[35] W. Sun, G. Yang, J. Peng, and Q. Du, "Lateral-slice sparse tensor robust principal component analysis for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 107–111, Jan. 2020.
[36] H. Liu, Y. Jia, J. Hou, and Q. Zhang, "Global-local balanced low-rank approximation of hyperspectral images for classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 2013–2024, Apr. 2022.
[37] Q. Zhang, Q. Yuan, J. Li, X. Liu, H. Shen, and L. Zhang, "Hybrid noise removal in hyperspectral imagery with a spatial–spectral gradient network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7317–7329, Oct. 2019.

[38] X. Jiang, H. Xue, L. Zhang, X. Gao, Y. Zhou, and J. Bai, "Hyperspectral data feature extraction using deep learning hybrid model," *Wireless Pers. Commun.*, vol. 102, no. 4, pp. 3529–3543, 2018.
[39] C. Wang, L. Zhang, W. Wei, and Y. Zhang, "When low rank representation based hyperspectral imagery classification meets segmented stacked denoising auto-encoder based spatial-spectral feature," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 284.
[40] H. Zhang, H. Chen, G. Yang, and L. Zhang, "LR-Net: Low-rank spatial-spectral network for hyperspectral image denoising," *IEEE Trans. Image Process.*, vol. 30, pp. 8743–8758, 2021.
[41] Y. Duan, H. Huang, and T. Wang, "Semisupervised feature extraction of hyperspectral image using nonlinear geodesic sparse hypergraphs," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.
[42] F. Luo, Z. Zou, J. Liu, and Z. Lin, "Dimensionality reduction and classification of hyperspectral image via multistructure unified discriminative embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2021.
[43] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2097–2104.
[44] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *Unpublished paper*, 2010. [Online]. Available: https://arxiv.org/abs/1009.5055
[45] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[46] Z. Lin and H. Zhang, *Low-Rank Models in Visual Analysis: Theories, Algorithms, and Applications*. New York, NY, USA: Academic Press, 2017.
[47] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," 2011, *arXiv:1109.0367*.
[48] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.
[49] Y. Zhang, "Recent advances in alternating direction methods: Practice and theory," 2010, pp. 1–33.
[50] F. Cao, Z. Yang, J. Ren, M. Jiang, and W.-K. Ling, "Does normalization methods play a role for hyperspectral image classification?" 2017, *arXiv:1710.02939*.
[51] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," *J. ACM*, vol. 58, no. 3, 2011, Art. no. 11.
[52] W. He, H. Zhang, H. Shen, and L. Zhang, "Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 713–729, Mar. 2018.
[53] Y. Wang, T. Li, L. Chen, Y. Yu, Y. Zhao, and J. Zhou, "Tensor-based robust principal component analysis with locality preserving graph and frontal slice sparsity for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2022.
[54] Y. Chang, L. Yan, B. Chen, S. Zhong, and Y. Tian, "Hyperspectral image restoration: Where does the low-rank property exist," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6869–6884, Aug. 2021.
[55] J. Zhang, Y. Peng, W. Ouyang, and B. Deng, "Accelerating admm for efficient simulation and optimization," *ACM Trans. Graph.*, vol. 38, no. 6, pp. 1–21, 2019.
[56] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, Aug. 2012.
[57] P. Alquier, "Bayesian methods for low-rank matrix estimation: Short survey and theoretical study," in *Proc. Int. Conf. Algorithmic Learn. Theory*, 2013, pp. 309–323.
[58] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3419–3430, Dec. 2011.
[59] V. Y. Tan and C. Févotte, "Automatic relevance determination in nonnegative matrix factorization with the/spl beta/-divergence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1592–1605, Jul. 2013.

Shujun Yang (Member, IEEE) received the B.S. degree from JiangSu University, ZhenJiang, China, in 2011, the M.S. degree from NanJing University, NanJing, China, in 2014, and the Ph.D. degree from City University of Hong Kong, Hong Kong,, in 2020.

She is currently a postdoctoral with the Department of Computer Science, Southern University of Science and Technology, China. Her research interests include hyperspectral remote sensing image processing and applications, pattern recognition, and machine learning.

Yu Zhang (Member, IEEE) is an Associate Professor with the Department of Computer Science and Engineering, Southern University of Science and Technology. His research interests mainly include artificial intelligence and machine learning, especially in multitask learning, transfer learning, dimensionality reduction, metric learning, and semisupervised learning. He has published a book *Transfer Learning* and about 80 papers on top-tier conferences and journals. He serves as a Reviewer for various journals and area chairs/(senior) program committee members for several top-tier conferences.

Prof. Zhang has won the Best Paper Awards in UAI 2010 and PAKDD 2019, and the best Student Paper Award in WI 2013.

Yuheng Jia (Member, IEEE) received the B.S. degree in automation and the M.S. degree in control theory and engineering from Zhengzhou University, Zhengzhou, China, in 2012 and 2015, respectively, and the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2019.

He is currently an Associate Professor with the School of Computer Science and Engineering, Southeast University, China. His research interests broadly include topics in machine learning and data representation, such as semisupervised learning, high-dimensional data modeling and analysis, low-rank tensor/matrix approximation and factorization, graph signal processing and machine learning on graphs, and deep learning.

Weijia Zhang (Member, IEEE) received the B.Sc. degree in mathematics and the M.Sc. degree in computer science from Nanjing University, Nanjing, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from University of South Australia, Adelaide, South Australia, in 2018.

From 2018 to 2020, he worked as a Research Fellow with the University of South Australia. He is currently an Associate Professor with the School of Computer Science and Engineering, Southeast University, China. His main research interests include causal inference and machine learning.