Sparse and Low-Rank Representation With Key Connectivity for Hyperspectral Image Classification

Yun Ding, Yanwen Chong, and Shaoming Pan

***Abstract*—Combined techniques of sparse representation (SR) and low-rank representation (LRR) are commonly used for hyperspectral image (HSI) classification. Although they have the ability to capture the interclass representations of data for HSI classification, they ignore the adaptive key connectivity of the learned intraclass data representations in particular with the high-dimensional complex HSI data. It is well-known that the key connectivity of graph-based algorithms is crucial for subspace learning because of the guarantees of its good neighbors. For this purpose, a novel sparse and low-rank representation with key connectivity (SLRC) method is proposed for HSI classification. To be specific, the adaptive probability graph structure is developed to integrate the SR and LRR regularizations to formulate the SLRC model, which flexibly perform discriminative latent subspace construction and preserve the key connectivity of intraclass representations. Then, extensive experiments are executed based on three popular HSI datasets, which demonstrates that the SLRC method outperforms the other popular methods.**

***Index Terms*—Sparse representation (SR), low-rank representation (LRR), key connectivity, hyperspectral image (HSI).**

I. Introduction

HYPERSPECTRAL images (HSIs) provide detailed structural and spectral information because they comprise hundreds of narrow spectral bands [1]–[4], which can effectively capture the subtle differences between different materials and facilitate better land-cover classification. The classification of HSIs, where each pixel is assigned one thematic class in a scene, has attracted much attention in many studies because it plays a vital role in various applications [5], [6]. In fact, effective feature expression is important in land-cover classification since they greatly affect the analysis of numerous HSI applications. Recently, graph-based learning algorithms have got widespread attention in representation learning because of their interpretability and effectiveness in practice [7], [8]. The critical step in graph-based learning aims at building good graph to denote the original samples using nodes. The edge weights in graph-based learning algorithms represent the similarity of each pair of samples. In general, the edge weights can be obtained by using the representation coefficients that reveal the intrinsic properties in the new feature space. Thus, the most common graph-based learning algorithms construct the edge weights by using the representation coefficients of sparse representation (SR)-based methods and low-rank representation (LRR)-based methods [9], [10]. In addition, the edge weights of graph-based algorithm are also constructed by using collaborative representation (CR) coefficients based on collaborative representation methods in HSI analysis because it is able to obtain a closed-form solution via least-squares estimation and provide much lower complexity than that for SR [11], [12]. For the details of CR-based methods in HSI analysis, we refer readers to read the related works in [13] and [14]. Due to that sparse and low-rank properties have the potential capability to induce the connectivity with block-diagonal structure within each subspace, and the related sparse and low-rank based methods are introduced as follows.

SR-based graph methods are traditional graph-based learning methods. Notably, an SR model of a sample ideally can be obtained by the combination of a few samples from the same subspace via the underlying self-expressiveness property [15], [16]. Due to the powerful SR expression performance of these methods, they have been applied widely in HSI analysis [17]. Considering neighboring pixels centered at the pixel of interest, the classical joint sparse representation (JSR) model [18] was proposed for HSI classification. Furthermore, considering the SR classifier ignores the Euclidean distance relationship among samples, the class-dependent sparse representation classifier (cdSRC) was proposed for HSI classification [19]. Because l_1-based SR may yield unstable representation results, Tang *et al.* [20] incorporated manifold learning into SR to exploit the smoothness across neighboring samples. Moreover, considering that the limited labeled samples of different classes are unbalance and the learned representation is hard to reflect the particular characteristics of each class, the graph based context-aware elastic net [21] was proposed for HSI classification by taking full advantages of SR and CR, which promoted local and global consistency preserving. Recently, the combination of class probability and sparse representation has displayed excellent performance in HSI classification, mainly because class structure information helps the SR model to obtain a highly effective graphical expression. For example, the probabilistic class structure-regularized sparse representation (PCSSR) model was proposed for HSI classification [22] by combining the class probability and the SR model. Next, the spatial and class structure regularized sparse representation graph was proposed by Shao *et al.* [23] for HSI classification. However, the above SR-based regularization methods are sometimes overly sparse

Manuscript received June 16, 2020; revised August 20, 2020; accepted September 8, 2020. Date of publication September 11, 2020; date of current version September 30, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 62072345, Grant 41671382, and Grant 41271398, and in part by the National Key Research and Development Program of China under Grant 2017YFB0504202. *(Corresponding authors: Yanwen Chong; Shaoming Pan).*

The authors are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: yunding@whu.edu.cn; ywchong@whu.edu.cn; pansm@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3023483

such that the key connectivity (few "good" neighbors satisfied the requirement of connectivity can recover the subspace based on self-expressiveness property) of graph expression with each subspace (class) may be disconnected [24]. In fact, the nuclear norm-based regularizations such as LRR lead to the dense connectivity in intraclass representations [25]. The classical LRR-based methods are to seek the lowest rank expression by pursuing linear combinations between the samples and the given dictionary [26], [27]. Similarly, the LRR-based algorithms have been widely applied in the HSI analysis. The self-supervised LRR method was proposed to avoid the existence of noise information and the neglect of semantic information around HSI pixel [26]. Moreover, Wu *et al*. [28] proposed the kernel low-rank representation based on local similarity for HSI classification by taking homogenous region of pixels into consideration. In addition, considering the limitations of feature representations in LRR model and semi-supervised classification methods, the authors proposed the spatial–spectral locality constrained LRR with semi-supervised hypergraph learning (SSLR-HG) [27] model for HSI classification, which preserved global and local data structure simultaneously via nuclear norm regularized LRR and a spatial–spectral locality constraint. Although the connectivity within the intraclass samples may be guaranteed because of dense coefficients, the coefficients of the interclass samples are usually nonzero, whose key connectivity of interclass samples cannot be forbidden [25].

In fact, the connectivity is not only existing in sparse or low-rank graph models, but also existing in probabilistic graph models such as the classical Markov random field [29]. Markov property is the conditional independence property of the joint distribution over a set of random variables based on the link of nodes of undirected graph [30]. It plays an important role in simplifying both the structure of a model and the computations in probabilistic mode [30]. Generally, Markov property means that the current state of a thing relies on the state of local neighborhoods (local connectivity) that are close to it [31]. However, the connectivity of Markov property cannot discover the latent connectivity that is not necessarily close to it but belong to the same subspace.

Hence, a natural idea is to combine the SR and LRR based regularization to obtain a balanced graph connectivity based on self-expressiveness property that automatically picks a few other points that are not necessarily close to it but that belong to the same subspace to recover the corresponding subspace for HSI classification. For example, the sparse and low-rank graph-based discriminant analysis (SLGDA) method was proposed for HSI classification [32]. Indeed, it is commonly believed that the SR and LRR techniques can reveal the underlying data structures of the correlation patterns of different classes and tasks [33]. The combination of SR and LRR (abbreviated as CSRLRR) techniques ideally yields block-diagonal representations relying on the well-known self-expression property [34] and is useful in HSI classification because of the complementary advantages of both perspectives. However, it is hard to obtain good graph connectivity by striking a balance between the separation of interclass representations and the intrinsic connectivity of intraclass representations in HSI analysis based on CSRLRR techniques because of the following two reasons. First, the obtained sparse representation coefficients are usually able to recover a subspace representation but the connectivity of sparse representation may not be satisfied if the dimension of the subspace is larger than three [24] because the high dimension of the HSI subspace usually contains hundreds of narrow spectral bands and the high-dimensional HSI data is hard to satisfy the requirement of the connectivity via sparse representation. Second, another reason for this phenomenon is that the CSRLRR techniques may obtain a tradeoff of graph connectivity with strictly observing assumption of independent subspaces [35, Proposition 1], but different subspaces are overlapped with each other because of the complex multimodalities and nonlinear characteristics of real HSI dataset [36]. It is hard to obtain an adaptive good graph connectivity being as few as possible to recover the corresponding subspace based on self-expressiveness property that automatically picks a few other points that are not necessarily close to it but that belong to the same subspace to recover the corresponding subspace, according to the basic connectivity property. Thus, the self-expressiveness coefficients of the CSRLRR techniques are hard to obtain key connectivity of intraclass representations.

Furthermore, there are still some useful explorations in key connectivity of CSRLRR-based methods. For example, in order to preserve the effective block-diagonal structure via self-expressiveness, Wang *et al.* [9] proposed a locality and structure-regularized LRR (LSLRR) method for HSI classification. In addition, Ding *et al.* [37] proposed robust spatial—spectral block-diagonal structure representation with fuzzy class probability (SSBDFCP) method for HSI classification. Although they are exploiting the connectivity of intraclass representations, the connectivity structure of intraclass representations is manually fixed by utilizing fixed threshold value, which cannot adaptively deal with the complex and changing environment in HSI. Moreover, the apparent spatial–spectral information is directly used to measure the connectivity of intraclass representations of HSI pixels, which fails to excavate valuable information of underlying subspace. In short, there is no effective way to exploit the adaptive key connectivity of the intraclass representations in complex HSI classification.

Fortunately, the valuable information of the underlying subspace will be conductive to obtain good graph connectivity [38]. To address the hardness of the key connectivity in HSI analysis based on the CSRLRR techniques, motivated by [38] and [39], we propose a novel SLRC model for preserving key connectivity of intraclass representations in HSI classification based on CSRLRR techniques. Different from the proposed adaptive models in [39], we utilize the balanced valuable representation vectors of underlying subspace between SR regularization and LRR regularization as intrinsic information to seek adaptive probabilistic connectivity between each pair of samples in the SLRC model. And, the SLRC model has the potential capability to capture semantic information related to the separation of the interclass representations of a given dataset, and learn the key connectivity of the intraclass data representations based on our previous SSBDFCP model.

Thus, the main contributions with respect to this article can be concluded as follows.

1) We proposed the SLRC model for HSI classification. The intrinsic connectivity information associated with the adaptive probability graph structure is integrated to SR and

LRR regularizations to formulate SLRC model, which can adaptively preserve the key connectivity of intraclass representations and adaptively enables discriminative latent subspace construction for the given data.

2) We proposed the unified framework to adaptively learn the key connectivity of the intraclass representation based on representation learning and proved the usefulness of the SLRC method. Experimental results on three public HSI datasets illustrate that the SLRC model performs better against the related classification methods.

The remainder of this article is prepared as follows. Section II briefly introduces some related works. Then, the proposed SLRC method is presented in Section III in detail. Section IV provides relevant solutions with SLRC. Section V reports the extensive experimental results and analyses. Finally, the conclusions are drawn in Section VI.

II. Related Work

In Section II, some related works most closest to this article, including those LSLRR and SSBDFCP as well as SSLR-HG models, are briefly reviewed. Then, the definition of connectivity in graph-based algorithms is described. To begin with, some notations used in this article are given. The bold uppercase letters, e.g., $\boldsymbol{X}$ denotes matrices, and the lowercase letters, e.g., $\boldsymbol{x}$ denotes columns vectors. $\boldsymbol{x_{ij}}$ denotes the ith row and jth column of matrix $\boldsymbol{X}$, and i represents the identity matrix. The Frobenius norm of matrix $\boldsymbol{X}$ is defined as $\| \boldsymbol{X} \|_F^2 = \mathrm{tr}(\boldsymbol{X}^\tau \boldsymbol{X}) = \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^\tau)$, where $\mathrm{tr}(\bullet)$ is the trace operator, $\| \boldsymbol{X}\|_*$ denotes the nuclear norm of matrix $\boldsymbol{X}$, which can be obtained by the sum of the singular values of matrix $\boldsymbol{X}$, i.e., $\| \boldsymbol{X}\|_* = \sum_i |\sigma_i|$, and $\| \boldsymbol{X}\|_1$ is the l_1 norm of matrix $\boldsymbol{X}$, which can be obtained by the sum of the absolute value, i.e., $\| \boldsymbol{X}\|_1 = \sum_{ij} |\boldsymbol{x}_{ij}|$. Then, the related works are introduced as follows.

A. LSLRR

The LSLRR model develops a local structural constraint that utilizes both the spatial and spectral similarity to enhance the classification property of the LRR model. Then, the structure-preserving strategy of the self-expressiveness of HSI pixels is incorporated to formulate the LSLRR model, which can be defined as follows:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{E}} \quad & \| \mathbf{Z}\|_* + \lambda \| \mathbf{E}\|_{2,1} + \alpha \| \bar{\boldsymbol{M}} \odot \mathbf{Z}\|_1 + \beta \| \mathbf{Z} - \mathbf{R} \|_F^2 \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{X}_{tr}\mathbf{Z} + \mathbf{E}, \mathbf{1}\mathbf{Z} = \mathbf{1}, \mathbf{Z} \geq 0 \end{aligned} \tag{1}$$

where $\boldsymbol{X} = [\boldsymbol{X}_{\mathrm{tr}}, \boldsymbol{X}_{\mathrm{tt}}] \in \mathbb{R}^{d\times N}$ consists of training samples $\boldsymbol{X}_{\mathrm{tr}}$ and test samples $\boldsymbol{X}_{\mathrm{tt}}$, and $\bar{\boldsymbol{M}}_{ij} = \sqrt{\| \boldsymbol{x}_i - \boldsymbol{x}_j \|_2^2 + m \| \boldsymbol{l}_i - \boldsymbol{l}_j \|_2^2}$ is the local structure constraint combining the spectral and spatial similarity and $\boldsymbol{L} = [\boldsymbol{l}_i, \boldsymbol{l}_j, \ldots, \boldsymbol{l}_N] \in \mathbb{R}^{2\times N}$ denotes the position coordinates of HSI pixels, $\mathbf{Z}$ represents the self-expressive representation of data, $\mathbf{E}$ represents the noise component. $\| \cdot \|_*$ represents the nuclear norm of a matrix, $\| \mathbf{E}\|_{2,1} = \sum_{j=1}^{N} \sqrt{\sum_{i=1}^{d} (e_{ij})^2}$ represents the $l_{2,1}$ norm for characterizing the error term, and λ, β, α, m are the regularization parameters, $\mathbf{1}$ is the corresponding one vector. In addition, $\mathbf{R} = [\bar{\mathbf{R}}, \tilde{\mathbf{Q}}]$ represents the structure-preserving strategy, where $\bar{\mathbf{R}}$ is the class-wise block diagonal structure of intraclass representations of training samples, $\tilde{\mathbf{Q}}_{ij} = \exp(\| \boldsymbol{x}_{tr}^i - \boldsymbol{x}^j \|_2^2 + m \| \boldsymbol{l}_{tr}^i - \boldsymbol{l}^j \|_2^2/\phi)$ can be obtained by fixing a threshold value in spatial–spectral distance matrix $\bar{\boldsymbol{M}}_{ij}$. In essential, $\mathbf{R}$ is essentially the fixed connectivity by manually adjusting the corresponding threshold value.

B. SSBDFCP

Considering that the apparent spectral information to compute the similarity between HSI pixels cannot reveal the inherent characteristics of HSI pixels, the SSBDFCP model makes full use of fuzzy class probability structure to preserve block-diagonal structure of intraclass representations. And, the SSBDFCP model is written as follows:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{E},\mathbf{W}} \quad & \| \mathbf{Z}\|_* + \lambda_1 \| \tilde{\mathbf{A}} \odot \mathbf{Z} \|_F^2 + \lambda_2 \| \Theta \odot \mathbf{Z}\|_1 \\ & + \lambda_3 \| \mathbf{Z} - \mathbf{S} \|_F^2 + \lambda_4 \| \mathbf{E}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{X}_{\mathrm{tr}}\mathbf{Z} + \mathbf{E} \end{aligned} \tag{2}$$

where $\tilde{\mathbf{A}}$ represents the class-wise block diagonal structure of intraclass representations of training samples, $\| \tilde{\mathbf{A}} \odot \mathbf{Z} \|_F^2$ represents the spectral connectivity constraint of intraclass representations via self-expressiveness, $\mathbf{S}$ represents the spatial relationship within the $s \times s$ local window, where $\boldsymbol{S}_{ij} = \begin{cases} 1 & \text{if } \boldsymbol{x}_i \in N_s(\boldsymbol{x}_j) \text{ or } \boldsymbol{x}_j \in N_s(\boldsymbol{x}_i) \\ 0 & \text{otherwise.} \end{cases}$, and $N_{\mathrm{s}}(\boldsymbol{x}_i)$ represents the spatial neighbors. Specifically, $\boldsymbol{S} \in \{0, 1\}$ is a binary matrix in which the representation of spatially similar samples within in a spatial window is 1 and the representation of spatially dissimilar samples is 0. $\| \mathbf{Z} - \mathbf{S} \|_F^2$ represents the spatial connectivity of the neighboring samples, and λ_1, λ_2, λ_3, and λ_4 are several tradeoff parameters. In addition, the fuzzy class probability Θ is essentially the local constraint by means of inherent class structure of HSI samples. In fact, the key connectivity of the intraclass representations is manually fixed by expert knowledge and is not adaptive to reveal good neighbors in SSBDFCP model when facing complex HSI.

Besides, some works related to SLRC model such as SSLR-HG model are briefly introduced as follows. First, considering the limitations of feature representations in LRR model and semi-supervised classification methods, the authors proposed the spatial–spectral locality constrained LRR with SSLR-HG [27] model for HSI classification, which preserved global and local data structure simultaneously via nuclear norm regularized LRR and a spatial–spectral locality constraint. Although SSLR-HG model tries to utilize the KNN largest absolute values to preserve the connectivity of SSLR coefficients, simply preserving larger coefficients of SSLR coefficients does not guarantee the key connectivity of each subspace [24].

C. Connectivity

An extremely sparse representation results in unsatisfied connectivity if the non-zero representation coefficients cannot comprise enough correlations for the connectivity with each subspace. And low-rank representation usually generates a block-diagonal structure with dense connections of interclass samples [24]. The definition of connectivity is as follows.

Definition 1. (Connectivity) [24], [40]: The connectivity property requires that the samples within same subspace form the corresponding connected components in the affinity graph $\mathbf{G} = (W, E)$. Given K subspaces $\{\mathbf{\Omega}_i\}_{i=1}^{K}$, an ideal $\mathbf{G}$ meets the following three requirements.

1) If two samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are connected with weight being $w_{ij} \neq 0$, then we have $\boldsymbol{x}_i \in \mathbf{\Omega}$ and $\boldsymbol{x}_j \in \mathbf{\Omega}$.
2) If we have $\boldsymbol{x}_i \in \mathbf{\Omega}$ and $\boldsymbol{x}_j \notin \mathbf{\Omega}$, then $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are disconnected.
3) The connections should be as few as possible, and yet guarantee the connectivity property within each subspace.

As a result, an ideal $\mathbf{G}$ has K connected components corresponding to K subspaces, which guarantee the subspace-preserving property of intraclass samples [41].

III. Proposed Method

In this section, we will describe the SLRC method in detail. The adaptive probability connectivity graph is integrated to our previous basic SSBDFCP model to formulate SLRC model for capturing the key connectivity of the intraclass representations.

A. Learning Based on an Adaptive Probability Graph Structure

Suppose that an HSI dataset is denoted as $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{\mathrm{C}}] \in \mathbb{R}^{d \times N}$, where C denotes the number of land cover types, d represents the number of bands, and$N$ represents the number of HSI pixels. Moreover, $\boldsymbol{X}_{\mathrm{tr}} \in \mathbb{R}^{d \times l}$ denotes the labeled samples matrix and $\boldsymbol{X}_{\mathrm{tt}} \in \mathbb{R}^{d \times u}$ denotes the unlabeled samples matrix, where l denotes the number of the training samples and u denotes the number of the testing samples. Both the training and testing samples can be represented as $\boldsymbol{X} = [\boldsymbol{X}_{\mathrm{tr}}, \boldsymbol{X}_{\mathrm{tt}}] \in \mathbb{R}^{d \times N}$, where $N = l + u$.

The adaptive probability graph structure is introduced to highlight the key connectivity of the learned intraclass representations, which is conductive to the classification of land-cover samples. Based on the problems (1) and (2), the balanced representation vectors $\mathbf{z}_i$ and $\mathbf{z}_j$ of underlying subspace can be obtained via SR and LRR regularizations. $\mathbf{z}_i$ represents the ith column of $\mathbf{Z}$ and also represents the new feature embedding of $\boldsymbol{x}_i$. Their embeddings in a new feature space should be close to each other if two samples are close in the sense of the intrinsic geometry based on graph embedding [42]. Motivated by the underlying assumption that if the pixels $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are similar, the representation coefficient vectors $\mathbf{z}_i$ and $\mathbf{z}_j$ are similar. Therefore, the representation vectors are used to construct an adaptive probabilistic graph to preserve the key connectivity of the near-optimal neighbor correlation because the projected representation vectors via the balanced representations between SR and LRR regularizations are discriminative (i.e., the key connectivity of intraclass representation and the separation of interclass representation) to separation of interclass samples. The adaptive graph learning model is built by assigning optimal neighbors for HSI samples based on local measurement. It is assumed that the nearby samples have high probabilities of being connected in adaptive graph learning, and this relation can be written as follows:

$$\psi(\mathbf{z}) = \sum_{i,j} \mathrm{dist}(\mathbf{z}_i, \mathbf{z}_j) \times a_{ij}$$

$$\text{s.t. } 0 < a_{ij} < 1, \mathbf{A}\mathbf{1} = \mathbf{1} \tag{3}$$

where $\mathbf{A}\mathbf{1} = \mathbf{1}$ represents $\mathbf{A} \times \mathbf{1} = \mathbf{1}$, a_{ij} represents the probability that the ith sample and the jth sample are in the same class. Note that minimizing this item $\sum_{i,j} \mathrm{dist}(\mathbf{z}_i, \mathbf{z}_j) \times a_{ij}$ has a potential capability to distinguish the correlation of different class samples for obtaining an ideal adaptive graph structure compared with minimizing $\sum_{i,j} \mathrm{dist}(\mathbf{x}_i, \mathbf{x_j}) \times a_{ij}$ because the new feature vector $\mathbf{z}_i$ is of overall expression capability that is robust to noise and can discover the valuable connections of underlying different subspaces, which will be conductive to obtain good graph connectivity. Specifically, $\mathbf{A}$ reflects the probabilistic connectivity of each pair of samples [43], [44] belonging to the same class, which aims to capture data connectivity. $\mathrm{dist}(\mathbf{z}_i, \mathbf{z}_j)$ measures the distance between the vector $\mathbf{z}_i$ and the vector $\mathbf{z}_j$. The probability matrix $\mathbf{A}$ is a transition probability matrix, and each of its rows is a probability distribution. Notably, the probability matrix $\mathbf{A}$ is a feasible similarity measurement adaptively determined using a self-tuning technique according to the common Euclidean distance between each pair of representation coefficient vectors. In fact, the adaptive probability graph structure exploits the probabilistic data connectivity to enhance the subspace-preserving property of the intraclass representation of HSI samples. Therefore, the adaptive probability graph structure is integrated to formulate the SLRC method, which enhances the connectivity of the intraclass representations. Thus, based on the SSBDFCP model, the SLRC model can be defined as follows:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{A},\mathbf{E}} & \parallel \mathbf{Z} \parallel_* + \lambda_1 \parallel \mathbf{\Theta} \odot \mathbf{Z} \parallel_1 \\ & + \lambda_2 \left(\parallel \mathbf{Z} - \mathbf{A} \parallel_F^2 + \gamma \sum_{i=1}^{l} \sum_{j=1}^{N} \parallel \mathbf{z}_i - \mathbf{z}_j \parallel_2^2 \boldsymbol{a}_{ij} \right) \\ & + \lambda_3 \parallel \mathbf{Z} - \mathbf{S} \parallel_F^2 + \lambda_4 \parallel \mathbf{E} \parallel_{2,1} \\ \text{s.t. } & \mathbf{X} = \mathbf{X}_{tr}\mathbf{Z} + \mathbf{E} \end{aligned} \tag{4}$$

where $\mathbf{A} \in \mathbb{R}^{l \times N}$ and γ is the parameter used to balance the self-expressiveness of spectral information and the probabilistic connectivity of HSI samples. Notably, the regularization $\parallel \mathbf{Z} - \mathbf{A} \parallel_F^2$ naturally has the following three attributes.

1) The regularization $\parallel \mathbf{Z} - \mathbf{A} \parallel_F^2$ encompasses a shift toward a unified optimization framework by combining the self-expressiveness and adaptive probabilistic connectivity of the learned representation of the HSI samples.
2) The representation learned by adaptive probabilistic graph learning displays connectivity consistent with that of the corresponding intraclass representation, which can improve the subspace connectivity of intraclass representation in pursuing sparse and low-rank representations of HSI samples. Moreover, the regularization $\parallel \mathbf{Z} - \mathbf{A} \parallel_F^2$ avoids a trivial solution of $\mathbf{A}$ from the optimization perspective.
3) Different from [33], the adaptive probability connectivity graph can be obtained by valuable representation vectors of underlying subspace via SR and LRR regularizations, which is discriminative to pursuing the key connectivity of intraclass to a certain extent.

Compared with fuzzy class probability of the SSBDFCP model, the local constraint Θ of the SLRC model is the simple local similarity measurement by Euclidean distance, that is, $\boldsymbol{\Theta}_{ij} = \sqrt{\| \boldsymbol{x}_i - \boldsymbol{x}_j \|_2^2}$. The constraint of spatial information of the SLRC model is the same as the SSBDFCP model.

IV. Optimization and Algorithm Analysis

In this section, we adopt the alternating direction method of multipliers (ADMM) [45] to solve the optimization of SLRC model in (4), which can be briefly introduced as follows.

A. Optimization

First, the two auxiliary variables $\mathbf{P}$ and $\mathbf{Q}$ are introduced to separate the problem (4), which can be written as follows:

$$\begin{aligned}\min_{\mathbf{Z},\mathbf{A},\mathbf{E}} &\| \mathbf{P}\|_* + \lambda_1 \| \boldsymbol{\Theta} \odot \mathbf{Q}\|_1 + \frac{\lambda_3}{2} \| \mathbf{Z} - \mathbf{S} \|_F^2 + \lambda_4 \| \mathbf{E}\|_{2,1} \\ &+ \frac{\lambda_2}{2} \left(\| \mathbf{Z} - \mathbf{A} \|_F^2 + \gamma \sum_{i=1}^{l} \sum_{j=1}^{N} \| \mathbf{z}_i - \mathbf{z}_j \|_2^2 \, \boldsymbol{a}_{ij} \right) \\ \text{s.t. } & \mathbf{X} = \mathbf{X}_{tr}\mathbf{Z} + \mathbf{E}, \mathbf{P} = \mathbf{Z}, \mathbf{Q} = \mathbf{Z}. \end{aligned} \tag{5}$$

Then, the following augmented Lagrangian function can be obtained as follows:

$$\begin{aligned}\min_{\mathbf{Z},\mathbf{A},\mathbf{E},\mathbf{P},\mathbf{Q}} &\| \mathbf{P}\|_* + \lambda_1 \| \boldsymbol{\Theta} \odot \mathbf{Q}\|_1 \\ &+ \frac{\lambda_3}{2} \| \mathbf{Z} - \mathbf{S} \|_F^2 + \lambda_4 \| \mathbf{E}\|_{2,1} \\ &+ \frac{\lambda_2}{2} \left(\| \mathbf{Z} - \mathbf{A} \|_F^2 + \gamma \sum_{i=1}^{l} \sum_{j=1}^{N} \| \mathbf{z}_i - \mathbf{z}_j \|_2^2 \, \boldsymbol{a}_{ij} \right) \\ &+ \left\langle \hat{\mathbf{Y}}_1, \mathbf{X} - \mathbf{X}_{\mathrm{tr}}\mathbf{Z} - \mathbf{E} \right\rangle + \left\langle \hat{\mathbf{Y}}_2, \mathbf{P} - \mathbf{Z} \right\rangle \\ &+ \left\langle \hat{\mathbf{Y}}_3, \mathbf{Q} - \mathbf{Z} \right\rangle + \frac{\mu}{2} (\| \mathbf{P} - \mathbf{Z} \|_F^2 \\ &+ \| \mathbf{Q} - \mathbf{Z} \|_F^2 + \| \mathbf{X} - \mathbf{X}_{\mathrm{tr}}\mathbf{Z} - \mathbf{E} \|_F^2) \end{aligned} \tag{6}$$

where $\langle \hat{\mathbf{Y}}_2, \mathbf{P} - \mathbf{Z} \rangle = \mathrm{tr}(\hat{\mathbf{Y}}_2^{\tau}(\mathbf{P} - \mathbf{Z}))$. $\hat{\mathbf{Y}}_1$, $\hat{\mathbf{Y}}_2$, and $\hat{\mathbf{Y}}_3$ represents the Lagrangian multipliers, respectively. Moreover, $\mu > 0$ represents the corresponding penalty parameter. Then, the loss (6) can be minimized when one variable is changing and the remaining variables are fixed. And, the details are introduced as follows.

1) Update $\mathbf{A}$: Consider $\mathbf{A}$ as a variable and fix the other variables

$$\begin{aligned}\min_{\mathbf{A}} \frac{\lambda_2}{2} &\left(\| \mathbf{Z} - \mathbf{A} \|_F^2 + \gamma \sum_{i=1}^{l} \sum_{j=1}^{N} \| \mathbf{z}_i - \mathbf{z}_j \|_2^2 \, \boldsymbol{a}_{ij} \right) \\ \text{s.t. } & \mathbf{A}\mathbf{1} = \mathbf{1}, 0 < \boldsymbol{a}_{ij} < 1. \end{aligned} \tag{7}$$

From (7), we have

$$\begin{aligned}\min_{\mathbf{A}\mathbf{1}=\mathbf{1},0<\boldsymbol{a}_{ij}<1} & \frac{\lambda_2}{2} \| \mathbf{Z} - \mathbf{A} \|_F^2 + \lambda_5 \sum_{i=1}^{l} \sum_{j=1}^{N} \| \mathbf{z}_i - \mathbf{z}_j \|_2^2 \, \boldsymbol{a}_{ij} \\ &= \frac{\lambda_2}{2} \| \mathbf{Z} - \mathbf{A} \|_F^2 + \lambda_5 \sum_{i=1}^{l} \sum_{j=1}^{N} h_{ij} \boldsymbol{a}_{ij} \\ &= \frac{\lambda_2}{2} \| \mathbf{A} - (\mathbf{Z} - \frac{\lambda_5}{\lambda_2}\mathbf{H}) \|_2^2 \end{aligned} \tag{8}$$

where each element of $\mathbf{H}$ is calculated by $h_{ij} = \| \mathbf{z}_i - \mathbf{z}_j \|_2^2$ and $\lambda_5 = \frac{\lambda_2 \times \gamma}{2}$. From (8), it is obvious to see that the problem (8) is equivalent in solving the following problem:

$$\min_{\boldsymbol{a}_i\mathbf{1}=\mathbf{1},\boldsymbol{a}_i \geq 0} \frac{\lambda_2}{2} \| \boldsymbol{a}_i - \left(\mathbf{z}_i - \frac{\lambda_5}{\lambda_2}\mathbf{h}_i \right) \|_2^2 \tag{9}$$

where $\boldsymbol{a}_i$, $\mathbf{z}_i$, and $\mathbf{h}_i$ denote the ith rows of $\mathbf{A}$, $\mathbf{Z}$, and $\mathbf{H}$, respectively. Problem (9) can be quickly solved with the efficient algorithm presented in [46].

2) Update $\mathbf{Z}$: Consider $\mathbf{Z}$ as a variable and fix the other variables

$$\begin{aligned}\min_{\mathbf{Z}} &\frac{\lambda_2}{2} \| \mathbf{Z} - \mathbf{A} \|_F^2 + \frac{\lambda_3}{2} \| \mathbf{Z} - \mathbf{S} \|_F^2 \\ &+ \left\langle \hat{\mathbf{Y}}_1, \mathbf{X} - \mathbf{X}_{\mathrm{tr}}\mathbf{Z} - \mathbf{E} \right\rangle + \left\langle \hat{\mathbf{Y}}_2, \mathbf{P} - \mathbf{Z} \right\rangle \\ &+ \left\langle \hat{\mathbf{Y}}_3, \mathbf{Q} - \mathbf{Z} \right\rangle + \frac{\mu}{2} (\| \mathbf{P} - \mathbf{Z} \|_F^2 \\ &+ \| \mathbf{Q} - \mathbf{Z} \|_F^2 + \| \mathbf{X} - \mathbf{X}_{\mathrm{tr}}\mathbf{Z} - \mathbf{E} \|_F^2). \end{aligned} \tag{10}$$

From problem (10), this item $\sum_{i=1}^{l} \sum_{j=1}^{N} \| \mathbf{z}_i - \mathbf{z}_j \|_2^2 \, \boldsymbol{a}_{ij}$ is not considered into this process of updating $\mathbf{Z}$ because the high computational complexity and the main role to obtain an ideal $\mathbf{A}$ of the item $\sum_{i=1}^{l} \sum_{j=1}^{N} \| \mathbf{z}_i - \mathbf{z}_j \|_2^2 \, \boldsymbol{a}_{ij}$. Thus, the problem (10) can be equivalent to solve the following problem:

$$\begin{aligned}\min_{\mathbf{Z}} &\frac{\lambda_2}{2} \| \mathbf{Z} - \mathbf{A} \|_F^2 + \frac{\lambda_3}{2} \| \mathbf{Z} - \mathbf{S} \|_F^2 \\ &+ \frac{\mu}{2} \left(\| \mathbf{X} - \mathbf{X}_{\mathrm{tr}}\mathbf{Z} - \mathbf{E} + \frac{\hat{\mathbf{Y}}_1}{\mu^t} \|_F^2 \right. \\ &\left. + \| \mathbf{P} - \mathbf{Z} + \frac{\hat{\mathbf{Y}}_2}{\mu^t} \|_F^2 + \| \mathbf{Q} - \mathbf{Z} + \frac{\hat{\mathbf{Y}}_3}{\mu^t} \|_F^2 \right). \end{aligned} \tag{11}$$

From (11), we can calculate the derivative of problem (11) with $\mathbf{Z}$ and obtain the corresponding solution:

$$\begin{aligned}\boldsymbol{Z} = &\left[(2 + \frac{\lambda_2}{\mu} + \frac{\lambda_3}{\mu}) i + \boldsymbol{X}_{tr}^{\tau} \boldsymbol{X}_{tr} \right]^{-1} \\ &\left(\boldsymbol{X}_{tr}^{\tau} \bar{R}_1 + \bar{R}_2 + \bar{R}_3 + \frac{\lambda_2}{\mu}\mathbf{A} + \frac{\lambda_3}{\mu}\mathbf{S} \right) \end{aligned} \tag{12}$$

where $\bar{\boldsymbol{R}}_1 = \boldsymbol{X} - \mathbf{E}^t + (\hat{\mathbf{Y}}_1/\mu)$, $\bar{\boldsymbol{R}}_2 = \mathrm{P} + (\hat{\mathbf{Y}}_2/\mu)$, and $\bar{\boldsymbol{R}}_3 = \mathbf{z} + (\hat{\mathbf{Y}}_3/\mu)$.

3) Updating P: Fix the other variables and update P by solving the following problem:

$$\begin{aligned}\min_{\mathbf{P}} &\| \mathbf{P}\|_* + \left\langle \hat{\mathbf{Y}}_2, \mathbf{P} - \mathbf{Z} \right\rangle + \frac{\mu}{2} \| \mathbf{P} - \mathbf{Z} \|_F^2 \\ &= \| \mathbf{P}\|_* + \frac{\mu}{2} \| \mathbf{P} - (\mathbf{Z} - \frac{\hat{\mathbf{Y}}_2}{\mu}) \|_F^2 . \end{aligned} \tag{13}$$

Problem (13) can be solved via the singular-value thresholding operator [47], which can be written as follows:

$$\mathbf{P} = S_{1/\mu}\left(\mathbf{Z} - \hat{\mathbf{Y}}_2 \Big/ \mu\right) = \mathbf{U}\hat{S}_{1/\mu}(\mathbf{\Sigma})\mathbf{V}^{\tau}, \tag{14}$$

where $\mathbf{U\Sigma V}^{\tau}$ represents the singular-value of decomposition (SVD) of $(\mathbf{Z} - \hat{\mathbf{Y}}_2/\mu)$ and $\hat{S}_{1/\mu}(\mathbf{\Sigma}) = \mathrm{sgn}(\mathbf{\Sigma})\max(|\mathbf{\Sigma}| - 1/\mu, 0)$ is the soft-thresholding operator [25].

4) Updating $\mathbf{Q}$: Consider $\mathbf{Q}$ as a variable and fix the other variables

$$\begin{aligned}\min_{\mathbf{Q}} \lambda_1 \parallel \mathbf{\Theta} \odot \mathbf{Q}\parallel_1 + \left\langle \hat{\mathbf{Y}}_3, \mathbf{Q} - \mathbf{Z}\right\rangle + \mu/2 \parallel \mathbf{Q} - \mathbf{Z} \parallel_F^2 \\ = \lambda_1 \parallel \mathbf{\Theta} \odot \mathbf{Q}\parallel_1 + \mu/2 \parallel \mathbf{Q} - (\mathbf{Z} - \hat{\mathbf{Y}}_3/\mu) \parallel_F^2 .\end{aligned} \tag{15}$$

Problem (15) can be equivalently decomposed into $l \times N$ subproblems by means of an elementwise strategy [33] and. Thus, the optimal solution of problem (15) is

$$\mathbf{q}_{ij} = \hat{S}_{\lambda_2\Theta_{ij}/\mu}(\mathbf{z}_{ij} - (\hat{Y}_3)_{ij}/\mu). \tag{16}$$

5) Updating $\mathbf{E}$: Consider $\mathbf{E}$ as a variable and fix the other variables

$$\begin{aligned}&\min_{\mathbf{E}} \lambda_4 \parallel \mathbf{E}\|_{2,1} \\ &+ \left\langle \hat{\mathbf{Y}}_1, \mathbf{X} - \mathbf{X}_{tr}\mathbf{Z} - \mathbf{E}\right\rangle + \frac{\mu}{2} \parallel \mathbf{X} - \mathbf{X}_{tr}\mathbf{Z} - \mathbf{E} \parallel_F^2 \\ &= \frac{\lambda_4}{\mu} \parallel \mathbf{E}\|_{2,1} + \frac{1}{2} \parallel \mathbf{E} - (\mathbf{X} - \mathbf{X}_{tr}\mathbf{Z} + \frac{\hat{\mathbf{Y}}_1}{\mu}) \parallel_F^2\end{aligned} \tag{17}$$

which can be solved with the method in [48]. Denoting $\mathbf{W} = \mathbf{X} - \mathbf{X}_{tr}\mathbf{Z} + \hat{\mathbf{Y}}_1/\mu$, then the $i$the column of the optimal $\mathbf{E}$ as follows:

$$\mathbf{E}^i = \begin{cases} \frac{\|\mathbf{W}^i\|_2 - \lambda_4/\mu}{\|\mathbf{W}^i\|_2} \text{ if } \lambda_4/\mu < \parallel \mathbf{W}^i\|_2 \\ 0 \qquad \text{otherwise} \end{cases} \tag{18}$$

where $\mathbf{W}^i$ represents the ith column. After optimizing the $\mathbf{A}$, $\mathbf{Z}$, P, $\mathbf{Q}$, $\mathbf{E}$, the Lagrange multipliers $\hat{\mathbf{Y}}_1$, $\hat{\mathbf{Y}}_2$, and $\hat{\mathbf{Y}}_3$ are also updated as follows:

$$\begin{cases} \hat{\mathbf{Y}}_1^{t+1} = \hat{\mathbf{Y}}_1^t + \mu^t(\mathbf{X} - \mathbf{X}_{\mathrm{tr}}\mathbf{Z}^{t+1} - \mathbf{E}^{t+1}) \\ \hat{\mathbf{Y}}_2^{t+1} = \hat{\mathbf{Y}}_2^t + \mu^t(\mathbf{P}^{t+1} - \mathbf{Z}^{t+1}) \\ \hat{\mathbf{Y}}_3^{t+1} = \hat{\mathbf{Y}}_3^t + \mu^t(\mathbf{Q}^{t+1} - \mathbf{Z}^{t+1}) \end{cases} \tag{19}$$

where t represents the number of iterations in the SLRC method. The detailed procedure based on ADMM optimization used to solve the SLRC model (4) is concluded as Algorithm 1.

Algorithm 1: Optimize the SLRC by Means of the ADMM.

Input: The HSIs data matrix $\mathbf{X} = [\mathbf{X}_{tr}, \mathbf{X}_{tt}]$; Parameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4$; the measurement matrix $\mathbf{\Theta}$.
Initialization: $\mathbf{A} = \mathbf{0}$, $\mathbf{Z} = \mathbf{0}$, $\mathbf{P} = \mathbf{0}$, $\mathbf{Q} = \mathbf{0}$, $\mathbf{E} = \mathbf{0}$, $\hat{\mathbf{Y}}_1 = \mathbf{0}$, $\hat{\mathbf{Y}}_2 = \mathbf{0}$, $\hat{\mathbf{Y}}_3 = \mathbf{0}$, $\lambda_1, \lambda_2, \lambda_3, \lambda_4 > 0$, $\mu_{\max} = 1e8$, $\rho = 1.15$, $tol = 1e-4$.
While not converged do
1) Update $\mathbf{A}$ by optimizing (9);
2) Update $\mathbf{Z}$ by optimizing (12);
3) Update P by optimizing (14);
4) Update $\mathbf{Q}$ by optimizing (16);
5) Update $\mathbf{E}$ by optimizing (18);
6) Update $\hat{\mathbf{Y}}_1$, $\hat{\mathbf{Y}}_2$ and $\hat{\mathbf{Y}}_3$ by optimizing (19);
7) Update μ;
$\mu^{t+1} = \min(\mu_{\max}, \rho\mu^t)$.
9) Examine the convergence conditions
$\max(\frac{\parallel \mathbf{X} - \mathbf{X}_{tr}\mathbf{Z} - \mathbf{E}\|_\infty}{\parallel \mathbf{P} - \mathbf{Z}\|_\infty, \parallel \mathbf{Q} - \mathbf{Z}\|_\infty}) \le tol$.
8) Update t: $t \leftarrow t + 1$.
End
Outputs: $\mathbf{A}$, $\mathbf{Z}$, $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{E}$.

Algorithm 2: SLRC Model for HSIs Classification.

Input: HSI data matrix $\mathbf{X} = [\mathbf{X}_{tr}, \mathbf{X}_{tt}]$ with label indicator matrix Y.
1) Build the SLRC model (4) according to the HSI spectral information.
2) Solve the SLRC model via **Algorithm 1,** and obtain the feature representation $\mathbf{Z} = [\mathbf{Z}_{tr}, \mathbf{Z}_{tt}]$.
3) Construct an optimal linear classifier $\mathbf{B}^*$ by optimize problem (21).
4) Predict the labels $\mathbf{Y}_{tt}$ of testing samples by solving problem (23).
Output: Classification results $\mathbf{Y}_{tt} \in \mathbb{R}^{u\times C}$ of the test samples

B. HSI Classification via Linear Classifier

After the ADMM optimization for the SLRC model (4), the HSI representation can be obtained as $\mathbf{Z} = [\mathbf{Z}_{tr}, \mathbf{Z}_{tt}]$, where $\mathbf{Z}_{tr} \in \mathbb{R}^{l\times l}$ and $\mathbf{Z}_{tt} \in \mathbb{R}^{l\times u}$. Then, $\mathrm{Y} \in \mathbb{R}^{l\times C}$ represents the labels of the training samples, and the final pixel-based classification results are obtained by using the linear classifier [49]. The detailed procedure is described as follows.

The linear classifier $\mathbf{B}$ can be acquired by utilizing spectral representation $\mathbf{Z}_{tr}$ and labels $\mathrm{Y} \in \mathbb{R}^{l\times C}$ of the training samples, which can be written as follows:

$$\mathbf{B}^* = \underset{\mathbf{B}}{\mathrm{argmin}} \parallel \mathbf{Y} - \mathbf{Z}_{tr}^{\tau}\mathbf{B} \parallel_F^2 + \eta \parallel \mathbf{B} \parallel_F^2, \tag{20}$$

where η denotes the corresponding regularization parameter. And, the optimal solution can be acquired as follows:

$$\mathbf{B}^* = \mathbf{Z}_{tr}^{\tau}\mathbf{Y}(\mathbf{Z}_{tr}^{\tau}\mathbf{Z}_{tr} + \eta\mathbf{I})^{-1}. \tag{21}$$

Thus, given the spectral representation of the testing samples, the final recognition is determined by judging

$$\mathrm{label}(\mathbf{Y}_{tt}) = \mathrm{argmax}(\mathbf{Z}_{tt}^{\tau}\mathbf{B}^*) \tag{22}$$

where $\mathbf{Y}_{tt} \in \mathbb{R}^{u\times C}$ represents the labels of the testing samples. The complete process of HSI classification for the SLRC model is concluded as Algorithm 2.

C. Convergence

For the ADMM optimization process of the SLRC model listed in Algorithm 1, the optimal solution for the $\mathbf{A}$, $\mathbf{Z}$, $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{E}$ can be solved by optimizing corresponding functions. First, the closed-form solution of $\mathbf{A}$ can be quickly determined with an efficient algorithm [46] that guarantees convergence. Moreover, the convergence of $\mathbf{Z}$ can be guaranteed by introducing the corresponding auxiliary variables $\mathbf{P}$ and $\mathbf{Q}$, which

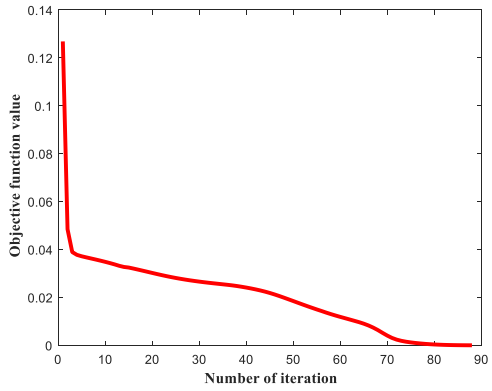


Fig. 1. Convergence curve of SLRC model.

Fig. 2. Indian Pines. (a) False-color image. (b) Ground truth image (adopted bands in false color are 20, 69, and 178, respectively).

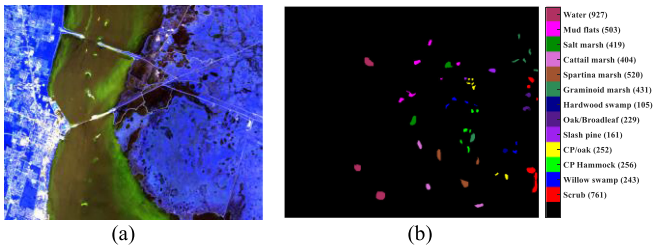


Fig. 3. KSC data. (a) False color image. (b) Ground truth image (adopted bands in false color are 5, 22, and 55, respectively).

is solved by using classical singular-value thresholding. The closed-form of the variable $\mathbf{E}$ can be solved by [25]. Therefore, the convergence of the SLRC model can be guaranteed and the convergence curves of the SLRC model is shown in Fig. 1.

D. Analysis of Theoretical Convergence

Via our simple analysis in Section IV-C, the convergence curve of the SLRC model has been described as Fig. 1 with Indian Pines dataset. Next, the deeply theoretical analysis will be conducted in terms of convergence of the SLRC model. To address the main model (4), the ADMM algorithm was applied to obtain the solution of problem (4).

Motivated by [33], the famous ADMM algorithm is able to address the following problem:

$$\min_{\mathbf{z}\in\mathbb{R}^l,\mathbf{r}\in\mathbb{R}^m} f_1(\mathbf{z}) + f_2(\mathbf{r}) \ \text{ s.t. } \mathbf{W}\mathbf{z} + \mathbf{M}\mathbf{r} = \mathbf{u} \qquad (23)$$

where $\mathbf{W} \in \mathbb{R}^{p\times l}$, $\mathbf{M} \in \mathbb{R}^{p\times m}$, $\mathbf{u} \in \mathbb{R}^p$, and f_1, f_2 are convex functions. The corresponding matrix optimization can be written as follows:

$$\min_{\mathbf{Z}\in\mathbb{R}^{l\times N},\mathbf{R}\in\mathbb{R}^{m\times N}} f_1(\mathbf{Z}) + f_2(\mathbf{R}) \text{s.t.} \mathbf{W}\mathbf{Z} + \mathbf{M}\mathbf{R} = \mathbf{U} \qquad (24)$$

where $\mathbf{U} \in \mathbb{R}^{p\times N}$. The corresponding augmented Lagrangian of problem (24) with multiplier can be written as follows:

$$\varphi(\mathbf{Z},\mathbf{R},\hat{\mathbf{Y}}) = f_1(\mathbf{Z}) + f_2(\mathbf{R}) + \frac{\mu}{2} \parallel \mathbf{W}\mathbf{Z} + \mathbf{M}\mathbf{R} - \mathbf{U} \parallel_F^2 + \langle \hat{\mathbf{Y}}, \mathbf{W}\mathbf{Z} + \mathbf{M}\mathbf{R} - \mathbf{U} \rangle \qquad (25)$$

where $\hat{\mathbf{Y}} \in \mathbb{R}^{p\times N}$ represents the corresponding Lagrangian multiplier and μ represents the penalty parameter. It is worth noting that the main model (4) can be formulated as a special case of problem (24). To be specific, the constraints of problem (4) can be written as $\mathbf{W}\mathbf{Z} + \mathbf{M}\mathbf{R} = \mathbf{U}$. Let $\mathbf{W} = \begin{pmatrix} -\mathbf{I}_p \\ -\mathbf{I}_p \\ \mathbf{X}_{tr} \end{pmatrix}$, $\mathbf{M} = \begin{bmatrix} \mathbf{I}_p & & \\ & \mathbf{I}_p & \\ & & \mathbf{I}_d \end{bmatrix}$, $\mathbf{R} = \begin{pmatrix} \mathbf{P} \\ \mathbf{Q} \\ \mathbf{E} \end{pmatrix}$, $\mathbf{U} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{X} \end{pmatrix}$, and $\mathbf{I}_p$ represents the $p \times p$ identity matrix. It is apparent that the main model (4) can be reformulated as problem (24). According to the optimization procedure of the SLRC model, the SLRC model is equivalent to the two-block ADMM [33], [50] via the analysis of problem (24). Therefore, the global convergence of the SLRC model is theoretically guaranteed. For the issue of the two-block ADMM, there exists a saddle point of $\varphi(\mathbf{Z},\mathbf{R},\hat{\mathbf{Y}})$, that is, $\mathbf{Z}^t \to \mathbf{Z}^*$, $\mathbf{R}^t \to \mathbf{R}^*$, $\hat{\mathbf{Y}}^t \to \hat{\mathbf{Y}}^*$, where $(\mathbf{Z}^*, \mathbf{R}^*, \hat{\mathbf{Y}}^*)$ is the saddle point according to [51, Proposition 1.1.5]. Obviously, the convergence of the proposed SLRC model can be theoretically guaranteed and guaranteed with experimental curve in Fig. 1.

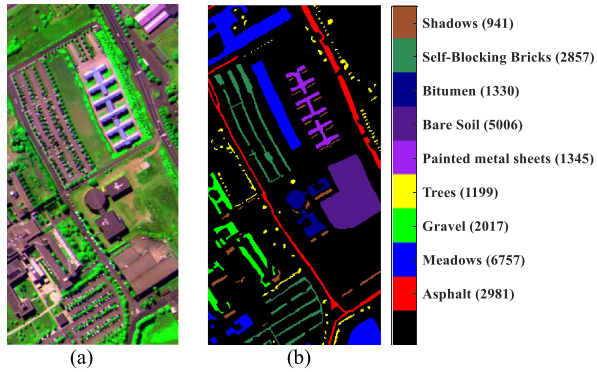


Fig. 4. PaviaU data. (a) False-color image. (b) Ground truth image (adopted bands in false-color are 52, 99, and 4, respectively).

E. Complexity Analysis

From Algorithm 1, the computational costs are mainly from the inverse operation and singular-value thresholding in the calculation of the $\mathbf{Z}$, $\mathbf{P}$, and $\mathbf{Q}$. The computational complexity of calculating $\mathbf{Z}$ is $\mathrm{O}(dl^2 + Nl^2)$ owing to the calculation of the inverse matrix. In addition, the complexity with optimizing $\mathbf{P}$ is nearly $\mathrm{O}(Nl^2), (N > l)$ based on the classical SVD operation. To be specific, the calculation of the SVD usually takes a lot of time when the l and N are very large. Next, the complexity of optimizing $\mathbf{Q}$ is approximately $\mathrm{O}(Nl)$ owing to the calculation of the l_1 problem. Moreover, the complexity of optimizing $\mathbf{E}$ is

TABLE I
NUMBERS OF TRAINING AND TESTING SAMPLES ON THREE DATASETS

	Indian Pines			KSC				PaviaU			
Class	Name	train	test	Class	Name	train	test	Class	Name	train	test
1	Alfalfa	4	42	1	Scrub	38	723	1	Asphalt	149	2832
2	Corn-notill	142	1286	2	Willow swamp	12	231	2	Meadows	337	6420
3	Corn-mintill	83	747	3	Cabbage palm hammock	12	244	3	Gravel	100	1917
4	Corn	23	214	4	Cabbage palm/oak	12	240	4	Trees	59	1140
5	Grass-pasture	48	435	5	Slash pine	8	153	5	Painted metal sheets	67	1278
6	Grass-trees	73	657	6	Oak/broadleaf hammock	11	218	6	Bare Soil	250	4756
7	Grass-pasture-mowed	2	26	7	Hardwood swamp	5	100	7	Bitumen	66	1264
8	Hay-windrowed	47	431	8	Graminoid marsh	21	410	8	Self-Blocking Bricks	142	2715
9	Oats	2	18	9	Spartina marsh	26	494	9	Shadows	47	894
10	Soybean-notill	97	875	10	Cattail marsh	20	384		Total	1217	23216
11	Soybean-mintill	245	2210	11	Salt marsh	20	399				
12	Soybean-clean	59	534	12	Mud flats	25	478				
13	Wheat	20	185	13	Water	46	881				
14	Woods	126	1139		Total	256	4955				
15	Buildings-Grass-Trees-Drives	38	348								
16	Stone-Steel-Towers	9	84								
	Total	1018	9231								

nearly $\mathrm{O}(dN)$. In short, the total complexity of optimizing the SLRC method is nearly $\mathrm{O}(t(dl^2 + 2Nl^2 + Nl + dN))$ via our analysis.

V. Experimental Result and Analysis

In Section V, some comprehensive experiments are conducted to verify the usefulness of our proposed SLRC model in HSI classification. The experiments of the SLRC model and the related popular classification methods can be quantitatively and qualitatively evaluated based on three real HSI datasets.

A. Dataset Description

Some experimental results were obtained based on three typical HSI datasets including Indian Pines, Kennedy Space Center (KSC), Pavia University (PaviaU) [52]. This Indian Pines scene covers an agricultural field and contains 16 classes, which consists of 145×145 pixels with 220 spectral bands. The KSC scene has a spatial size of 512×614 with 176 spectral bands and consist 13 land cover types. The PaviaU has a spatial size of 457×231 with 103 spectral bands and consists of 9 land cover types. Moreover, the false-color image and the corresponding ground truth of three datasets are given in Figs. 2–4.

B. Experimental Settings

To verify the superiority of our proposed SLRC model, some related popular methods, including the SVMCK [53], JSR [18], cdSRC [19], EPF [54], PCSSR [22], LSLRR [9], SLGDA [32], KLRRLSCK [28], and SSBDFCP [37] methods, are compared with the SLRC model for HSI classification. In the experiment, the maximum noise fraction (MNF) [55] is first applied to reduce the dimensionality of HSIs for alleviating the computational burden. For fairness, all the related comparison methods have the same dimension reduction settings. Then, for the size of training sets, we randomly picked 10% samples, 5% samples, and 5% samples in each class as training samples for the Indian Pines, KSC, and PaviaU datasets, respectively, and the remaining samples for three datasets were used as the testing set to evaluate the proposed methods in HSI classification. The detailed numbers of the training and testing samples used in the experiment are reported in Table I. All the classification methods were repeated ten times to circumvent the bias owing to random sampling. For EPF, the optimal parameter setting was based on that in [54]. For the other compared methods, all the optimal parameter settings are reported in Table II. Finally, the overall accuracy (OA), kappa coefficient (KC), and class accuracy (CA) were employed to evaluate all approaches in HSI classification.

C. Experimental Results and Analysis

Tables III–V demonstrate the final classification performance of different models in three datasets and their corresponding classification maps are given in Figs. 5–7. Among these compared methods of HSI classification, the PCSSR and SLGDA are the pixelwise classification methods which only utilize the spectral information. The other methods, including SVMCK, JSR, cdSRC, EPF, LSLRR, KLRRLSCK, SSBDFCP, and SLRC, combine spectral–spatial information to perform classification. From Table III, the classification performance of the PCSSR method is worse than the other methods in terms of the OA and KC values. This result indicates that spatial contextual information is good for HSI classification. In terms of OA and KC, the SLRC method achieves values of 98.86% and 98.70%, respectively, which are higher than those of the other methods, which verifies the effectiveness of this approach in HSI classification. From Table IV, only a small number of HSI pixels are incorrectly classified, and the classification accuracy of the SLRC method yields the highest in OA and KC values compared with those of the other methods. Compared with the SVMCK, PCSSR, EPF, LSLRR, SLGDA, KLRRLSCK, SSBDFCP methods, the SLRC achieves the best classification results for most classes, which indicates that it is an effective and superior approach for HSI classification. From Table V, the average precision of all land cover types in PaviaU are more than 95% via the SLRC method. The pixels of most land cover types can be correctly classified by the SLRC method. Moreover, the SLRC yields the highest in OA and KC values in pixel classification, which indicates the effectiveness of the SLRC method. From Figs. 5–7, the classification map of the SLRC method is closest to the ground truth data and can be beneficial for visually distinguishing the different classes. The main reason

TABLE II
OPTIMAL PARAMETERS OF DIFFERENT METHODS FOR THREE DATASETS

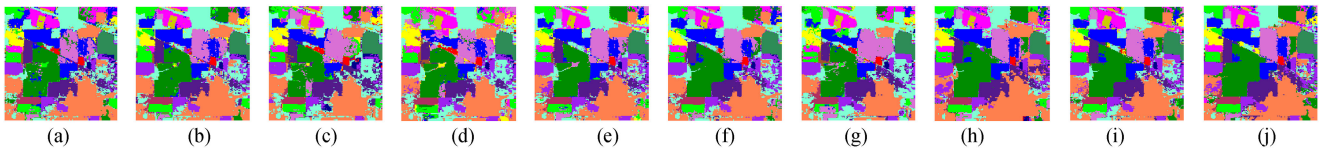
Methods	Indian Pines	KSC	PaviaU
SVMCK	$\mu=0.5$, window $=21\times21$	$\mu=0.5$, window $=19\times19$	$\mu=0.5$, window $=27\times27$
JSR	$K_0=30$, window $=5\times5$	$K_0=30$, window $=13\times13$	$K_0=30$, window $=11\times11$
cdSRC	$\lambda=0.02$, S $=10$	$\lambda=0.05$, S $=10$	$\lambda=0.05$, S $=10$
PCSSR	$\lambda_1=0.1, \lambda_2=0.06$	$\lambda_1=0.5, \lambda_2=0.001$	$\lambda_1=0.5, \lambda_2=0.001$
LSLRR	$\lambda=20, \alpha=0.8, \beta=0.6, m=25$	$\lambda=20, \alpha=0.8, \beta=0.6, m=30$	$\lambda=20, \alpha=0.8, \beta=0.6, m=25$
SLGDA	$\beta=0.1, \lambda=0.001$	$\beta=0.1, \lambda=0.001$	$\beta=0.1, \lambda=0.001$
KLRRLSCK	$\lambda=1e-4$	$\lambda=1e-2$	$\lambda=1e-4$
SSBDFCP	$\lambda_1=0.1, \lambda_2=0.06, \lambda_3=1,$ $\lambda_4=125$, window $=13\times13$	$\lambda_1=0.1, \lambda_2=0.01, \lambda_3=1,$ $\lambda_4=125$, window $=21\times21$	$\lambda_1=0.1, \lambda_2=0.06, \lambda_3=1,$ $\lambda_4=125$, window $=11\times11$
SLRC	$\lambda_1=0.1, \lambda_2=10, \lambda_3=1,$ $\lambda_4=30$, window $=13\times13$	$\lambda_1=0.1, \lambda_2=1, \lambda_3=1,$ $\lambda_4=30$, window $=25\times25$	$\lambda_1=0.1, \lambda_2=10, \lambda_3=1,$ $\lambda_4=30$, window $=35\times35$

TABLE III
OVERALL ACCURACY (%) OF DIFFERENT CLASSIFICATION METHODS FOR INDIAN PINES DATASET

Class	SVMCK	JSR	cdSRC	PCSSR	EPF	LSLRR	SLGDA	KLRRLSCK	SSBDFCP	SLRC
Alfalfa	56.19 (39.65)	82.27 (8.90)	**100** (0)	69.05 (0)	**100** (0)	60.95(27.27)	86.41 (14.33)	95.24 (0)	95.83 (4.91)	98.57 (2.3)
Corn-no till	85.30 (4.6)	89.80 (1.33)	82.77 (2.05)	76.52 (0.15)	95.83 (2.03)	88.24 (3.26)	73.33 (3.75)	90.99 (0.07)	97.80 (0.82)	**98.93** (0.63)
Corn-min till	71.97 (8.21)	86.11 (2.81)	85.74 (3.81)	54.80 (0.58)	97.82 (0.71)	91.37 (0.73)	76.87 (3.08)	**99.50** (0.57)	95.65 (2.13)	97.48 (1.5)
Corn	65.19 (11.2)	82.30 (4.07)	80.88 (4.42)	31.85 (0.46)	90.67 (7.28)	63.97 (17.11)	69.78 (3.24)	92.52 (0.76)	**95.21** (4.79)	94.81 (3.73)
Grass-pasture	94.71 (2.05)	92.48 (2.84)	93.24 (1.51)	91.76 (0.37)	**97.92** (1.32)	86.69 (4.71)	91.98 (2.89)	96.38 (1.80)	91.55 (3.31)	97.2 (1.86)
Grass-trees	98.95 (0.84)	98.12 (0.60)	87.86 (1.51)	97.34 (0.23)	99.26 (0.73)	95.72 (2.15)	90.93 (1.82)	**100** (0)	95.13 (1.10)	**99.89** (0.13)
Grass-pasture-mowed	43.46 (45.11)	77.58 (13.04)	**100** (0)	26.28 (1.57)	**100** (0)	21.15 (23.31)	97.31 (3.16)	96.15 (0)	51.92 (27.11)	62.69 (20.11)
Hay-windrowed	99.79 (0.3)	98.33 (1.07)	90.66 (1.42)	**100** (0)	97.99 (2.08)	**100** (0)	94.41 (1.35)	**100** (0)	99.71 (0.35)	**100** (0)
Oats	18.33 (29.29)	93.59 (6.81)	0 (0)	24.07 (2.87)	90.00 (31.62)	0 (0)	90.50 (12.96)	**100** (0)	91.67 (13.22)	82.78 (14.92)
Soybean-no till	62.96 (16.11)	92.93 (1.30)	82.48 (2.13)	62.76 (0.24)	95.69 (1.87)	87.09 (3.04)	76.90 (2.81)	93.00 (0.20)	96.63 (0.58)	**97.73** (1.56)
Soybean-min till	91.61 (1.58)	91.54 (1.54)	75.6 (1.69)	85.06 (0.17)	90.50 (3.93)	97.77 (1.00)	75.93 (0.74)	**99.66** (0.03)	97.40 (1.23)	99.44 (0.38)
Soybean-clean	81.91 (7.94)	76.81 (4.67)	87.52 (2.74)	67.85 (0.35)	97.93 (1.13)	91.35 (3.72)	80.82 (4.45)	95.65 (0.28)	93.26 (3.37)	**99.19** (0.43)
Wheat	99.41 (0.31)	93.10 (2.39)	94.54 (1.69)	98.92 (0)	**100** (0)	99.46 (0)	97.72 (1.38)	99.59 (0.27)	97.97 (2.09)	98.76 (1.55)
Woods	98.28 (0.61)	97.72 (0.85)	91.17 (1.4)	98.19 (0.12)	96.82 (1.86)	98.40 (0.63)	92..02 (1.64)	**100** (0)	99.45 (0.4)	99.92 (0.08)
Buildings-Grass -Trees-Drives	76.18 (4.71)	91.27 (3.23)	87.82 (2.15)	50.34 (0.95)	92.88 (3.42)	92.70 (6.12)	73.39 (5.50)	95.69 (0.23)	**99.43** (0.41)	99.34 (0.70)
Stone-Steel-Towers	98.21 (1.96)	91.19 (4.58)	**98.58** (0.03)	88.69 (1)	98.11 (0.61)	50.83 (22.90)	97.74 (2.01)	97.92 (0.60)	96..73 (4.05)	95.24 (4.31)
OA	86.32 (3.61)	91.31 (0.61)	83.97 (0.61)	79.23 (0.13)	95.08 (1.23)	95.75 (0.55)	81.18 (1.00)	97.14 (0.03)	97.10 (0.49)	**98.86** (0.20)
KC	84.24 (4.25)	90.09 (0.70)	81.54 (0.72)	76.03 (0.15)	94.36 (1.42)	95.15 (0.63)	78.44 (1.14)	96.74 (0.03)	96.70 (0.56)	**98.70** (0.22)

TABLE IV
OVERALL ACCURACY (%) OF DIFFERENT CLASSIFICATION METHODS FOR KSC DATASET

Class	SVMCK	JSR	cdSRC	PCSSR	EPF	LSLRR	SLGDA	KLRRLSCK	SSBDFCP	SLRC
Scrub	97.80 (1.24)	92.64 (2.05)	73.81 (4.89)	95.55 (1.14)	91.41 (7.46)	93.18 (3.94)	84.79 (2,85)	100 (0)	**100** (0)	**100** (0)
Willow swamp	96.41 (2.47)	91.11 (4.21)	79.61 (1.91)	87.1 (8.67)	96.34 (5.18)	94.02 (12.04)	76.76 (8.94)	97.26 (4.75)	87.77 (14.31)	**99.35** (1.50)
Cabbage palm hammock	86.64 (5.73)	78.61 (10.51)	73.25 (7.58)	83.07 (6.28)	91.17 (12.88)	**98.74** (0.80)	65.61 (9.03)	94.81 (6.91)	96.82 (5.56)	98.52 (2.18)
Cabbage palm/oak	60.13 (5.62)	77.76 (2.34)	55.69 (5.70)	68.58 (8.52)	91.15 (5.62)	**98.9** (3.03)	47.77 (4.85)	99.17 (0.72)	93.85 (5.04)	97.54 (3.63)
Slash pine	62.48 (12.17)	87.35 (4.37)	90.71 (8.48)	69.35 (6.40)	94.47 (8.00)	82.48 (4.77)	49.29 (8.94)	92.16 (9.63)	90.52 (10.94)	**96.80** (4.73)
Oak/broadleaf hammock	73.35 (7.66)	85.13 (6.06)	79.09 (12.41)	65.78 (6.27)	88.69 (31.36)	**100** (0)	50.16 (7.67)	95.41 (3.18)	**100** (0)	**100** (0)
Hardwood swamp	70.10 (16.31)	88.74 (6.20)	88.85 (11.33)	63.9 (17.47)	97.25 (5.99)	84.11 (28.93)	58.48 (13.66)	98.67 (1.15)	**100** (0)	**100** (0)
Graminoid marsh	93.44 (2.21)	89.64 (2.86)	87.35 (2.18)	90.68 (2.81)	93.57 (4.72)	97.42 (1.39)	73.40 (5.25)	100 (0)	98.41 (1.33)	**99.78** (0.35)
Spartina marsh	98.36 (1.62)	92.60 (1.83)	88.57 (1.69)	96.92 (1.22)	98.07 (2.47)	**100** (0)	88.77 (6.36)	95.28 (8.18)	99.29 (1.16)	98.36 (5.19)
Cattail marsh	99.35 (0.69)	94.68 (1.95)	99.15 (0.97)	99.22 (0.97)	99.77 (0.49)	99.95 (0.17)	88.40 (15.93)	100 (0)	99.48 (0.74)	**100** (0)
Salt marsh	95.69 (2.51)	85.19 (3.23)	97.80 (1.93)	96.84 (1.03)	**100** (0)	99.79 (0.35)	90.22 (7.90)	97.74 (0)	97.93 (2.07)	99.55 (0.87)
Mud flats	96.99 (1.79)	95.44 (1.51)	94.14 (2.78)	97.05 (1.16)	97.60 (2.65)	94.66 (4.22)	58.15 (30.11)	100 (0)	94.72 (6.19)	**98.83** (1.16)
Water	**100** (0)	98.73 (0.58)	99.68 (0.17)	**100** (0)	**100** (0)	**100** (0)	99.41 (1.33)	100 (0)	99.89 (0.23)	99.98 (0.07)
OA	92.59 (1.16)	90.88 (1.24)	86.26 (0.99)	91.54 (0.62)	96.06 (2.06)	97.01 (0.85)	68.96 (13.49)	98.45 (0.82)	97.74 (0.33)	**99.34** (0.62)
KC	91.75 (1.30)	89.85 (1.38)	84.59 (1.12)	90.56 (0.69)	95.60 (2.31)	96.68 (0.94)	65.66 (14.80)	98.28 (0.92)	97.49 (0.37)	**99.27** (0.69)



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Fig. 5. Classification maps of different methods for Indian pines. (a) SVMCK. (b) JSR. (c) cdSRC. (d) PCSSR. (e) EPF. (f) LSLRR. (g) SLGDA. (h) KLRRLSCK. (i) SSBDFCP. (j) SLRC.

TABLE V
OVERALL ACCURACY (%) OF DIFFERENT CLASSIFICATION METHODS FOR PAVIAU DATASET

Class	SVMCK	JSR	cdSRC	PCSSR	EPF	LSLRR	SLGDA	KLRRLSCK	SSBDFCP	SLRC
Asphalt	83.27 (2.64)	91.85 (4.62)	90.02 (1.55)	78.18 (2.08)	96.21 (0.81)	93.54 (3.73)	81.38 (1.34)	94.70 (3.70)	98.32 (0.41)	**98.51** (0.79)
Meadows	97.78 (0.39)	96.92 (0.46)	91.82 (0.98)	93.35 (0.56)	98.92 (0.37)	96.41 (0.36)	93.21 (1.37)	**99.93** (0.05)	99.26 (0.34)	99.24 (0.23)
Gravel	74.25 (1.82)	93.78 (1.44)	79.04 (3.45)	74.65 (4.03)	**99.73** (0.59)	97.96 (0.74)	80.89 (3.47)	85.95 (8.88)	95.58 (1.14)	98.84 (0.48)
Trees	96.89 (0.85)	97.40 (2.41)	97.33 (1.20)	90.29 (2.07)	**99.85** (0.23)	90.10 (1.28)	93.39 (2.10)	96.56 (0.54)	91.86 (1.48)	99.58 (0.17)
Painted metal sheets	99.97 (0.04)	88.89 (0.71)	**100** (0)	99.98 (0.04)	99.80 (0.31)	90.35 (0.70)	99.82 (0.21)	99.92 (0)	96.52 (1.79)	99.12 (0.57)
Bare Soil	98.10 (0.57)	98.88 (0.54)	93.86 (0.57)	90.52 (0.50)	96.35 (1.25)	99.74 (0.17)	90.70 (1.17)	**100** (0)	98.56 (0.59)	99.88 (0.13)
Bitumen	85.60 (3.04)	95.73 (6.70)	82.56 (3.25)	70.53 (7.39)	**100** (0)	93.99 (4.02)	86.19 (3.98)	97.78 (1.59)	96.22 (1.70)	98.08 (1.02)
Self-Blocking Bricks	77.44 (3.12)	92.02 (1.13)	77.22 (0.77)	65.93 (1.34)	89.54 (2.13)	87.73 (2.57)	70.32 (1.62)	**99.93** (0)	97.45 (0.88)	96.70 (1.47)
Shadows	99.87 (0.13)	99.17 (0.27)	**99.98** (0.05)	99.30 (0.37)	99.26 (0.82)	99.64 (0.83)	99.61 (0.41)	99.44 (0.65)	93.57 (1.23)	99.97 (0.05)
OA	91.28 (1.04)	95.29 (0.67)	89.66 (0.41)	85.09 (0.59)	97.04 (0.51)	95.03 (0.59)	87.51 (0.60)	97.85 (0.88)	97.59 (0.23)	**98.91** (0.17)
KC	89.56 (1.25)	94.37 (0.80)	87.60 (0.49)	82.15 (0.69)	96.45 (0.61)	94.05 (0.71)	85.04 (0.72)	97.43 (1.05)	97.11 (0.28)	**98.70** (0.21)

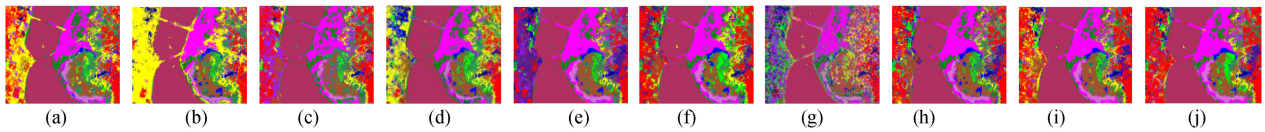


Fig. 6. Classification maps of different methods for KSC. (a) SVMCK. (b) JSR. (c) cdSRC. (d) PCSSR. (e) EPF. (f) LSLRR. (g) SLGDA. (h) KLRRLSCK. (i) SSBDFCP. (j) SLRC.

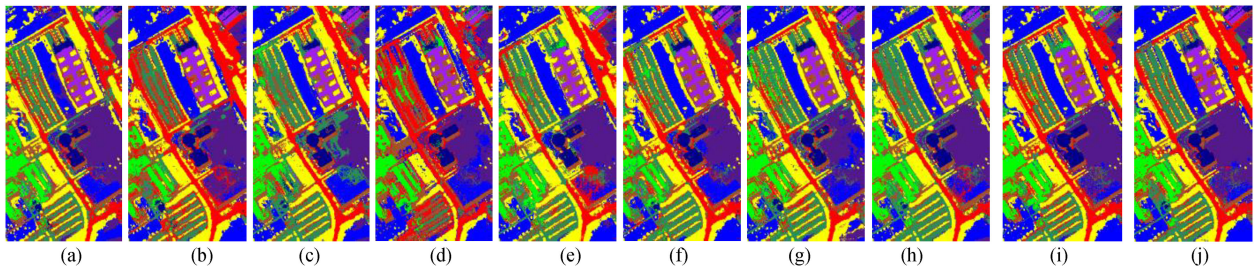


Fig. 7. Classification maps of different methods for PaviaU. (a) SVMCK. (b) JSR. (c) cdSRC. (d) PCSSR. (e) EPF. (f) LSLRR. (g) SLGDA. (h) KLRRLSCK. (i) SSBDFCP. (j) SLRC.

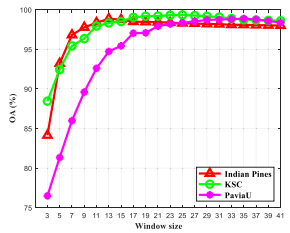


Fig. 8. OAs of the SLRC method with different spatial window sizes on the three datasets.

for the effectiveness of the SLRC method is that the SLRC incorporates the adaptive probability graph structure information to adaptively capture the probabilistic connectivity of intraclass representations by using the valuable representations vectors, which can efficiently preserve latent subspace constructions and near-optimal block-diagonal representations.

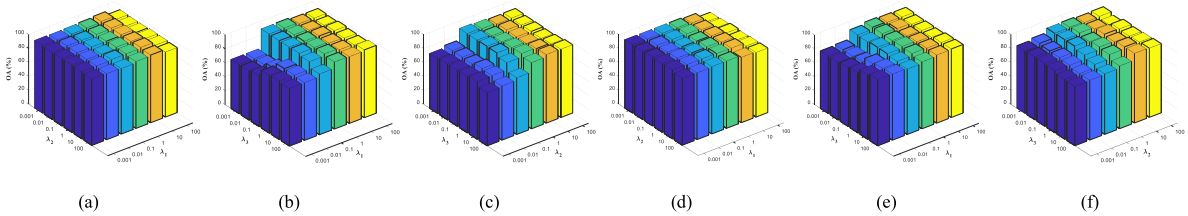
D. Analysis of Spatial Window Size

Although the SLRC model inherits the spatial constraint from the SSBDFCP model, the spatial contextual information has to be analyzed in HSI classification. The classification results of the SLRC method with different spatial window sizes can be seen in Fig. 8, which is tested based on three datasets. From Fig. 8, with the increase in the window size, the OAs display an inclination to ascend and then stabilize near the peak value. From Fig. 8, the spatial window size corresponding to the best OA for the KSC data is 25×25, which is larger than the best window size for the Indian Pines data of 13×13.

The main reason for this difference is that the distribution of different classes in KSC is dispersed, and a larger window can contain more spatial neighbors from the same class. In terms of PaviaU data, the best window size for SLRC method is 35×35 because of the intrinsic mechanism of the spatial neighborhoods that the distribution of intraclass samples in PaviaU land cover type is relatively concentrated on homogenous area with large regions

E. Analysis of Tradeoff Parameters

In the proposed SLRC problem (4), there are four parameters, namely, λ_1, λ_2, λ_3, and λ_4, that must be adjusted to balance the contribution of each constraint. First, in our experiment, the SLRC method is not sensitive to noise, and then λ_4 can be empirically large based on the analysis of the SSBDFCP method. Meanwhile, for clarity, the other three parameters will be divided into different groups for analysis, then one parameter will be changed while fixing the others. Finally, the optimal parameters were obtained with a grid search algorithm. Fig. 9 demonstrates the variations of OA with different tradeoff parameters in Indian Pines and KSC datasets. From Fig. 9, specifically,



(a) (b) (c) (d) (e) (f)

Fig. 9. OAs of different tradeoff parameters for Indian Pines are shown in (a)–(c) and the OAs of different tradeoff parameters for KSC are shown in (d)–(f).

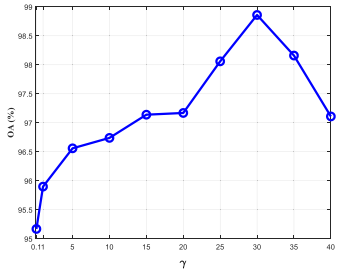


Fig. 10. OA variation with the different γ on Indian Pines.

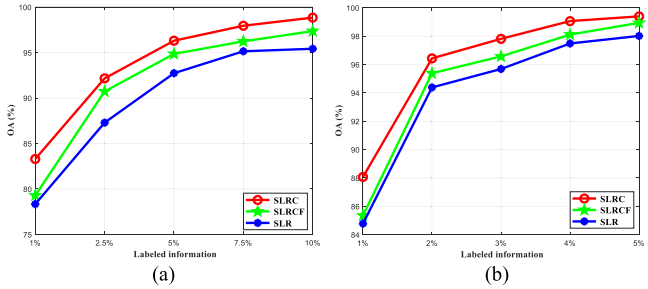


Fig. 11. OAs of three classification methods on the two data sets. (a) Indian Pines. (b) KSC.

the classification performance of the SLRC method is promising when λ_1 varies near 0.1 for the two datasets. Parameter λ_2 can be varied around 10 for Indian Pines and 1 for KSC because the self-expressiveness of the data and adaptive probability graph model are most consistent for intraclass representations for these values. The main role of the parameter λ_3 is to balance spatial information and spectral information, then it is not overly large or small. The analysis of the parameter γ in (4) has been discussed in details as follows. For convenience, the analysis of the parameter γ can be conducted on a small HSI dataset such as Indian Pines for optimal parameter. We only finetune the parameters γ when applying to another dataset, and then the optimal parameter γ is exploited for classification results. The analysis of the parameter γ on Indian Pines dataset is shown Fig. 10, which indicates the OA variation with different γ. From Fig. 10, we can see that the OA is highest when the value of the parameter γ is setting to 30. The possible reason for this phenomenon is that the adaptive key connectivity of intraclass representation can be more preserved and be beneficial to achieve the separation of interclass representation.

In addition, the parameters settings on PaviaU dataset for SLRC method are basically consistent with those on Indian Pines dataset by comprehensive experiments. Consequently, it is also proved that the SLRC method is robust to change of datasets.

F. Analysis of Adaptive Probability Graph

The adaptive probability graph structure is crucial for capturing the key connectivity of the complex HSI. Thus, in this section, we mainly explore what the role of the adaptive probability graph structure is to improve the connectivity of the intraclass representations in the two datasets. First, in the proposed SLRC problem, we assume that the adaptive probability graph structure does not exist, that is, $\lambda_2 = 0$, which can be denoted as the SLR model. Then, we assume that the adaptive probability graph structure is transformed into a fixed probability graph structure for application, which can be denoted as the SLRCF model. Finally, the classification results of three methods, the SLRC, SLRCF, and SLR methods, are shown in Fig. 11. From Fig. 11, the classification results of the SLRC model is better than those of the SLR and SLRCF models, which indicate that the adaptive probability connectivity graph structure in SLRC exploits the key connectivity of intraclass representations and is more able to deal with the complex HSI.

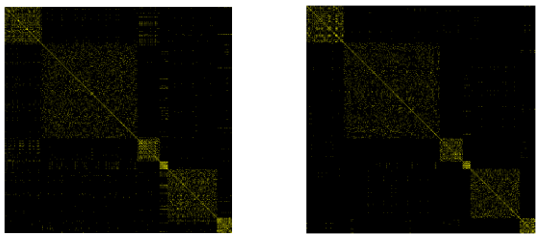
G. Analysis of the Class-Wise Block-Diagonal Structure

The discriminative capability of SLRC model is analyzed in detail. Thus, the block-diagonal structure of data representation via our SLRC model is analyzed as follows. In other words, the discriminative capability of SLRC can be interpreted as the block-diagonal structure of feature representation in hyperspectral image classification, which aims at achieving the separation of interclass representation coefficients and preserving the key connectivity of intraclass class representation coefficients.

To further analyze the block-diagonal structure of data, the connectivity of the intraclass representations of the SLRC and the related SLR methods can be obtained in Fig. 12. They were tested for six classes including soybean-notill, soybean-mintill, soybean-clean, wheat, woods, and bulidings-grass-trees-drives, from the Indian Pines data. From Fig. 12, the representation coefficients of the intraclass representation of the SLRC method is superior to the SLR method, mainly because the SLRC method focuses on capturing the underlying intraclass semantic representations by enforcing the constraint of the adaptive probability graph structure.

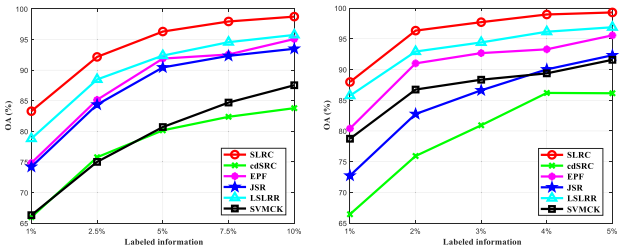
H. Analysis of Different Training Samples

In order to demonstrate the efficiency of the SLRC method, some related methods including the SVMCK, JSR, cdSRC, EPF,



(a) (b)

Fig. 12. Block-diagonal structure of two classification methods in terms of six land-covers. (a) SLR. (b) SLRC.



(a) (b)

Fig. 13. OAs of different methods with different labeled samples for the two datasets. (a) Indian Pines. (b) KSC.

TABLE VI
COMPUTATIONAL TIMES OF DIFFERENT METHODS FOR THE HSI DATASETS (SECONDS)

Method	SVMCK	JSR	cdSRC	PCSSR	EPF	LSLRR	SLRC
Indian Pines	81.34	92.33	248.19	11450.95	38.14	336.25	739.93
KSC	18.02	32.41	109.35	1483.22	13.05	39.65	62.54
PaviaU	786.39	800.31	129.41	75468.36	73.59	985.64	3210.54

and LSLRR methods are tested on two datasets with different percentages of training samples. As shown in Fig. 13, the OA values of the SLRC method are always superior to the other methods. The main reason for this phenomenon is mainly attributed to that the SLRC method is not only considering the separability of the interclass representations but also preserving the key connectivity of the intraclass representations.

I. Run Time Analysis

Table VI displays the running time of different classification methods based on the three HSI datasets. The experiments were conducted in MATLAB R2017b on a PC with an Intel Xeon E3-1240 3.50 GHz CPU and 48 GB RAM. According to the table, the computational times of EPF and SVMCK are the lowest and that of the PCSSR is the highest because the class probability obtained by sparse representations is time consuming in estimating the class distribution of unlabeled samples. The time consumed by SLRC is greater than that of LSLRR, JSR, and cdSRC. The main reason for this difference is that the operations in the SLRC method including those for SVD, matrix inversion, and the unified framework between self-expressiveness and the adaptive graph structure, are time consuming. Although the SLRC model still requires some time to run, the classification precision is significantly improved compared to that of other methods. Meanwhile, the time problem will no longer be a major issue owing to the improvements in the computing capacity [22].

TABLE VII
OVERALL ACCURACY (STD) (%) OF DIFFERENT PREPROCESSING METHODS

Data	Indian Pines		KSC		PaviaU	
Methods	SuperPCA	MNF	SuperPCA	MNF	SuperPCA	MNF
OA	98.17 (0.36)	**98.86** (0.20)	98.87 (0.85)	**99.34** (0.62)	97.68 (0.30)	**98.91** (0.17)
KC	97.91 (0.42)	**98.70** (0.22)	98.57 (0.94)	**99.27** (0.69)	96.42 (0.36)	**98.70** (0.21)

J. Analysis of Preprocessing Methods

To alleviate the computational burden of the SLRC model, the MNF is first applied to reduce the dimensionality of HSI. Furthermore, for analyzing the influence of different preprocessing methods in SLRC model, the commonly used MNF and the attractive SuperPCA [56] are applied to HSI reduction before performing SLRC model. The classification results on three datasets are shown in Table VII, which indicates the MNF is more suitable for the SLRC model.

K. Superiority of Adaptive Probability Graph

For demonstrating the advantage of the proposed adaptive probability graph preserving key connectivity, we provide the LSLRR model with the adaptive key connectivity and the corresponding objective function can be defined as follows:

$$\min_{\mathbf{Z},\mathbf{A},\mathbf{E}} \| \mathbf{Z}\|_* + \lambda \| \mathbf{E}\|_{2,1} + \alpha \| \bar{\mathrm{M}} \odot \mathbf{Z}\|_1 + \beta \left(\| \mathbf{Z} - \mathbf{A} \|_F^2 + \gamma \sum_{i=1}^{l} \sum_{j=1}^{N} \| \mathbf{z}_i - \mathbf{z}_j \|_2^2 \, \boldsymbol{a}_{ij} \right)$$
$$\text{s.t. } \mathbf{X} = \mathbf{X}_{tr}\mathbf{Z} + \mathbf{E}, \mathbf{1Z} = \mathbf{1}, \mathbf{Z} \geq 0 \quad (26)$$

where the $\| \mathbf{Z} - \mathbf{R} \|_F^2$ in the LSLRR model has been replaced with $\| \mathbf{Z} - \mathbf{A} \|_F^2 + \gamma \sum_{i=1}^{l} \sum_{j=1}^{N} \| \mathbf{z}_i - \mathbf{z}_j \|_2^2 \, \boldsymbol{a}_{ij}$ adaptively preserving the key connectivity by using richer information $\mathbf{z}_i \in \mathbf{Z}$ than original feature space due to sparse and low-rank constraints on representation coefficients $\mathbf{Z}$. The problem (26) is defined as the LSLRRA model. The classification performance of the LSLRR and LSLRRA methods on three datasets is shown in Table VIII. From Table VIII, the classification results of the LSLRRA model are better than the LSLRR model because the LSLRRA model adaptively preserves the key connectivity of intraclass representation by using richer information.

L. Analysis of Classifier

For evaluating the classification performance of different classifier, the popular support vector machine (SVM) with radial basis function and K-Nearest-Neighbor (KNN) classifiers with five neighbors are compared with linear classifier in HSI classification. Fig. 14 shows that OAs of different classifiers are obtained via the proposed SLRC method with increasing labeled information on Indian Pines dataset. The OAs obtained by linear classifier is best compared with SVM and KNN classifiers. Moreover, the performance of linear classifier performs better

TABLE VIII
OVERALL ACCURACY (OA) AND KAPPA COEFFICIENT (KC) (STD) % ON THREE PUBLIC HSI DATA SET

Data	Indian Pines		KSC		PaviaU	
Methods	LSLRRA	LSLRR	LSLRRA	LSLRR	LSLRRA	LSLRR
OA	**96.79** (0.38)	95.75 (0.55)	**98.14** (0.45)	97.01 (0.85)	**96.52** (0.31)	95.03 (0.59)
KC	**96.48** (0.43)	95.15 (0.63)	**97.93** (0.51)	96.68 (0.94)	**96.03** (0.36)	94.05 (0.71)

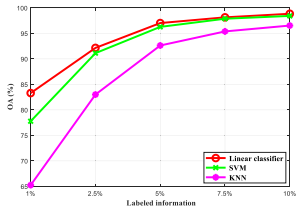


Fig. 14. OAs of different classifiers with different labeled samples for the Indian Pines datasets.

especially in the small percentage of labeled samples compared with SVM and KNN, and linear classifier is more suitable for the feature obtained by the SLRC method. And with the increasing of labeled information, the classification performance of three classifiers has little difference, which attributes to the sufficient class information.

VI. CONCLUSION

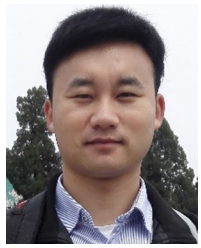
In this article, we proposed the useful SLRC method for HSI classification. To overcome the drawback, that is, the absence of adaptive key connectivity of CSRLRR-based methods, the SLRC method introduces an adaptive probability graph structure to improve the HSI classification results based on the SSBDFCP model. The adaptive probability graph structure is incorporated to flexibly disclose latent subspace constructions and preserve the probabilistic connectivity of the intraclass representations. Finally, some related experiments based on three public HSI datasets are performed to evaluate the SLRC approach, which demonstrates that the classification performance of the SLRC is superior to the other popular methods.

However, the SLRC method needs further improvement. For example, achieving the good postprocessing strategy in representation coefficients of intraclass and interclass samples will be good for the subspace-preserving property. In the future, graph-based convolutional networks [57] with block-diagonal structure will be explored to analyze the complex nonlinear structure of HSIs for achieving improved classification.

REFERENCES

[1] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1259–1270, Mar. 2018.
[2] Z. Shao, L. Zhang, X. Zhou, and L. Ding, "A novel hierarchical semisupervised SVM for classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 9, pp. 1609–1613, Sep. 2014.
[3] Z. Shao, L. Wang, Z. Wang, and J. Deng, "Remote sensing image superresolution using sparse representation and coupled sparse autoencoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2663–2674, Aug. 2019.
[4] P. Yi, Z. Wang, K. Jiang, Z. Shao, and J. Ma, "Multi-temporal ultra dense memory network for video super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2503–2516, Aug. 2020.
[5] Y. Guan, G. Shan, X. Yong, J. Liu, and Z. Xia, "Application of airborne hyperspectral data for precise agriculture," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2004, pp. 4195–4198.
[6] H. Xin, T. Hu, J. Li, Q. Wang, and J. A. Benediktsson, "Mapping urban areas in China using multisource data with a novel ensemble SVM method," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4258–4273, Aug. 2018.
[7] A. Subramanya and P. P. Talukdar, *Graph-Based Semi-Supervised Learning*, vol. 8, San Rafael, CA, USA: Morgan Claypool, 2014, pp. 1–25.
[8] T. Xie, S. Li, and B. Sun, "Hyperspectral images denoising via nonconvex regularized low-rank and sparse matrix decomposition," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 44–56, 2020.
[9] 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.
[10] Z. Ren, Q. Sun, B. Wu, X. Zhang, and W. Yan, "Learning latent low-rank and sparse embedding for robust image feature extraction," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 2094–2107, 2019.
[11] J. Li, H. Zhang, Y. Huang, and L. Zhang, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3707–3719, Jun. 2014.
[12] H. Su, Y. Yu, Q. Du, and P. Du, "Ensemble learning for hyperspectral image classification using tangent collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3778–3790, Jun. 2020.
[13] Z. Hongjun, Q. Bo, and Peijun, "Kernel collaborative representation with local correlation features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1230–1241, Feb. 2019.
[14] H. Su, B. Zhao, Q. Du, P. Du, and Z. Xue, "Multifeature dictionary learning for collaborative representation classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2467–2484, Apr. 2018.
[15] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.
[16] R. Hang and Q. Liu, "Dimensionality reduction of hyperspectral image using spatial regularized local graph discriminant embedding," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3262–3271, Sep. 2018.
[17] L. Wei and D. Qian, "A survey on representation-based classification and detection in hyperspectral remote sensing imagery ☆," *Pattern Recognit. Lett.*, vol. 83, pp. 115–123, 2015.
[18] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
[19] M. Cui and S. Prasad, "Class-dependent sparse representation classifier for robust hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2683–2695, May 2015.
[20] Y. Y. Tang, H. Yuan, and L. Li, "Manifold-based sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7606–7618, Dec. 2014.
[21] B. N. Soomro, L. Xiao, M. Molaei, L. Huang, Z. Lian, and S. H. Soomro, "Local and nonlocal context-aware elastic net representation-based classification for hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2922–2939, Jun. 2017.
[22] Y. Shao, N. Sang, C. Gao, and L. Ma, "Probabilistic class structure regularized sparse representation graph for semi-supervised hyperspectral image classification," *Pattern Recognit.*, vol. 63, pp. 102–114, 2017.
[23] Y. Shao, N. Sang, C. Gao, and L. Ma, "Spatial and class structure regularized sparse representation graph for semi-supervised hyperspectral image classification," *Pattern Recognit.*, vol. 81, pp. 81–94, 2018.

[24] J. Yang, J. Liang, K. Wang, P. Rosin, and M.-H. Yang, "Subspace clustering via good neighbors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 6, pp. 1537–1544, Jun. 2020.
[25] 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.
[26] 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.
[27] Q. Liu, Y. Sun, R. Hang, and H. Song, "Spatial–spectral locality-constrained low-rank representation with semi-supervised hypergraph learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4171–4182, Sep. 2017.
[28] Q. Liu, Z. Wu, L. Sun, Y. Xu, L. Du, and Z. Wei, "Kernel low-rank representation based on local similarity for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1920–1932, Jun. 2019.
[29] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral–spatial hyperspectral image classification with weighted Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1490–1503, Mar. 2015.
[30] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[31] X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, and J. Paisley, "Hyperspectral image classification with Markov random fields and a convolutional neural network," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2354–2367, May 2018.
[32] W. Li, J. Liu, and Q. Du, "Sparse and low-rank graph for discriminant analysis of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4094–4105, Jul. 2016.
[33] Z. Zhang, Y. Xu, L. Shao, and J. Yang, "Discriminative block-diagonal representation learning for image recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 3111–3125, Jul. 2018.
[34] Y. Li, J. Liu, Z. Li, Y. Zhang, H. Lu, and S. Ma, "Learning low-rank representations with classwise block-diagonal structure for robust face recognition," in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 2810–2816.
[35] Y.-X. Wang, H. Xu, and C. Leng, "Provable subspace clustering: When LRR meets SSC," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5406–5432, Sep. 2019.
[36] Z. Xue, P. Du, J. Li, and H. Su, "Simultaneous sparse graph embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6114–6133, Nov. 2015.
[37] Y. Ding, S. Pan, and Y. Chong, "Robust spatial-spectral block-diagonal structure representation with fuzzy class probability for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1747–1762, Mar. 2020.
[38] A. Singh, "Graph connectivity in noisy sparse subspace clustering," *Comput. Sci.*, vol. 17, pp. 689–708, 2015.
[39] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 977–986.
[40] B. Nasihatkon and R. Hartley, "Graph connectivity in sparse subspace clustering," in *Proc. CVPR*, 2011, pp. 2137–2144.
[41] X. Peng, Z. Yu, Z. Yi, and H. Tang, "Constructing the L2-graph for robust subspace learning and subspace clustering," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1053–1066, Apr. 2017.
[42] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
[43] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
[44] L. Ma, M. M. Crawford, X. Yang, and Y. Guo, "Local-manifold-learning-based graph construction for semisupervised hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2832–2844, May 2015.
[45] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations Trends Mach. Learn.*, vol. 3, pp. 1–122, 2011.
[46] F. Nie, X. Wang, M. I. Jordan, and H. Huang, "The constrained laplacian rank algorithm for graph-based clustering," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 1969–1976.
[47] J. F. Cai, S. Cand, J. Emmanuel, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, pp. 1956–1982, 2008.
[48] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2010, pp. 1813–1821.
[49] Y. Zhang, Z. Jiang, and L. S. Davis, "Learning structured low-rank representations for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 676–683.
[50] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, pp. 293–318, 1992.
[51] D. P. Bertsekas, *Convex Optim. Theory*. Nashua, NH, USA: Athena Scientific, 2009.
[52] V. Haertel and D. A. Langrebe, "On the classification of classes with nearly equal spectral response in remote sensing hyperspectral image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2374–2386, Sep. 1999.
[53] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
[54] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
[55] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, pp. 65–74, 1988.
[56] J. Junjun, M. Jiayi, C. Chen, W. Zhongyuan, C. Zhihua, and W. Lizhe, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.
[57] B. Jiang, Z. Zhang, D. Lin, J. Tang, and B. Luo, "Semi-supervised learning with graph learning-convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11313–11320.



Yun Ding received the master's degree from the Anhui University, Auhui, China, in 2018. He is currently working toward the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China.

His research interests include remote sensing image processing and pattern recognition.

Yanwen Chong received the B.S. degree from the Qufu Normal University, Jining, China, in 1995, and the M.S. and Ph.D. degrees from Wuhan University, Wuhan, China, in 1998 and 2001, respectively.

He is a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. His research interests include image processing, and computer vision and pattern recognition.

Shaoming Pan received the B.S. and M.S. degrees from Huazhong University of Science and Technology, Wuhan, China, in 1998, and the Ph.D. degree from Wuhan University, Wuhan.

He is an Associate Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is mainly engaged in the research on image processing, multimedia communication, and spatial information storage.