# Self-Supervised Low-Rank Representation (SSLRR) for Hyperspectral Image Classification

Yuebin Wang, Jie Mei, Liqiang Zhang, Bing Zhang, Anjian Li, Yibo Zheng, and Panpan Zhu

***Abstract*— Low-rank representation (LRR) can construct the relationships among pixels for hyperspectral image (HSI) classification with a given dictionary and a noise term. However, the accuracy of HSI classification based on LRR methods is degraded with the redundant and noise information existed in pixels. The neglect of semantic information around pixels in the LRR methods may cause "salt-and-pepper" problem in HSI classification. To avoid the aforementioned problems, a novel self-supervised low-rank representation method called SSLRR is developed. In SSLRR, the LRR and spectral–spatial graph regularization are developed as the pixel-level constraints to remove the redundant and noise information in HSIs. Superpixel constraints including data structure and relationship construction are further utilized to provide supervised feedback information to the subspace learning to avoid the "salt-and-pepper" problem generated in the pixel-based classification methods, and simultaneously enhance the performance of LRR. The pixel-level and superpixel-level regularizations are explicitly integrated into a unified objective function for LRR. By means of the linearized alternating direction method with adaptive penalty, the solution to the objective function is achieved by employing a customized iterative algorithm. We perform comprehensive evaluation of the proposed method on three challenging public HSI data sets. We obtain new state-of-the-art performance on these data sets, and achieve improvements of 44.3%, 13.4%, and 30.1% in overall accuracy compared to the best LRR method.**

***Index Terms*— Hyperspectral image (HSI) classification, low-rank representation (LRR), manifold learning, pixel and superpixel, self-supervised.**

## I. Introduction

LOW-RANK learning plays an important role in recent computer vision works, such as data clustering and image classification [1]–[12]. Among these methods, low-rank representation (LRR) [10] is a typical approach which seeks the LRR of all data jointly. Each data point is represented by a linear combination of the bases in a given dictionary; typically, the data matrix itself is chosen as the dictionary. In real applications, the data are often noisy and even grossly corrupted. The noise term is also added to the objective function.

LRR can provide a powerful tool for constructing data relationships in hyperspectral images (HSIs), which also called the reconstruction matrix. The $ij$th element of the matrix reflects the "similarity" between the pixel pair $i$ and $j$. The relationships between image pixels are used for data clustering and HSI classification. LRR cannot accurately describe the pixel relations due to the noise and mixed spectral pixels existed in HSIs. Thus, the classification accuracy is degraded with the redundant and noise information in the pixels. As shown in [13], good data representation quality can help to enhance the classification accuracy. Removal of the redundant and noise information from the high-dimensional features of HSIs is the key toward a successful classification [14].

Subspace learning [17], [21]–[23] and sparse representation [18]–[20] are effective solutions in reducing redundant information among the pixels of HSIs. To a certain extent, dimensionality reduction is equal to subspace learning, that is, projecting the original high-dimensional feature space to a low-dimensional subspace where the statistical properties like independent component analysis (ICA) [15] and principal component analysis [16] can be well preserved. Sparse representation has also been proven to be a powerful tool for extracting features from HSIs [18]–[20]. A similar framework for dictionary training and feature extraction is also found in [19] and [20]. Different from [18] and [19], a multiscale adaptive sparse representation model is proposed in [20]. In the regions of different scales, the complementary yet correlated information is incorporated for HSI classification.

In HSIs, the performance of LRR is affected not only by redundant and noise information, but also by the semantic information around one pixel. With data samples, traditional LRR algorithms can obtain the relationships among image pixels. They inevitably produce the "salt-and-pepper" problem in HSI classification [24]–[28] since the semantic information around one pixel is overlooked or the techniques are not well-established. Thus, the HSI classification accuracy is degraded. Pixel-based HSI classification methods have been studied in [24] and [27]. In order to preserve the semantic structural information in HSIs, superpixel is introduced to avoid the "salt-and-pepper" problem [33]–[36]. Superpixels are usually generated using the graph-based algorithms such as normalized cuts [29], entropy rate superpixel segmentation [30], and the gradient-descent-based algorithms such as SLIC [31] and SEEDS [32].

Just as Li *et al.* [37] stated, it is difficult to obtain an accurate over-segmentation superpixel map for HSI classification. Once

Manuscript received December 26, 2017; revised March 19, 2018; accepted March 23, 2018. Date of publication May 1, 2018; date of current version September 25, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 41371324 and in part by the China Postdoctoral Science Foundation under Grant 2016M600953. *(Yuebin Wang, Jie Mei, and Anjian Li contributed equally to this work.) (Corresponding author: Liqiang Zhang.)*

Y. Wang, J. Mei, L. Zhang, A. Li, Y. Zheng, and P. Zhu are with the State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China (e-mail: xxgcdxwyb@163.com; meijie@mail.bnu.edu.cn; zhanglq@bnu.edu.cn; anjianli94@163.com; bnuzyb@163.com; zlyxbmsl@163.com).

B. Zhang is with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China (e-mail: zb@ceode.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2018.2823750

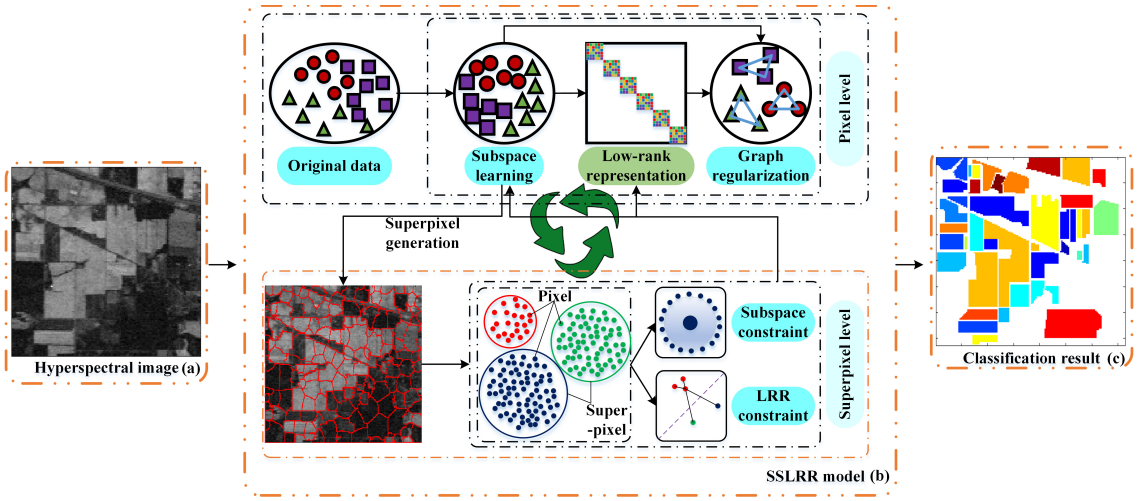



Fig. 1. Workflow of the SSLRR for HSI classification. (a) Training HSI. (b) Learning process of the SSLRR. (c) Classification results.

spectral pixels within a superpixel belong to different classes, a wrong classification cannot be avoided because of the "one label for one superpixel" manner. Under this condition, it is better to integrate the advantages of the pixel and superpixel into LRR for obtaining the relationship. LRR can further provide a better solution for describing the relationships through pixel-level and superpixel-level constraints. In the pixel level, with the manifold assumption introduced in [1], [38], and [39], if two pixels in HSIs have a close latent relationship, they have similar feature structures in subspace. We can embed the LRR into the function to remove the redundant and noise information (also called subspace learning in this paper). In the superpixel level, the superpixel generation depends on the clustering results of pixels, where the data representations of pixels play an important role for pixel clustering. The quality of subspace learning and the constructed relationships need to be evaluated for better superpixel generation. The rule of "one label for one superpixel" manner can be adopted to validate the pixel clustering purity, which provides the supervised feedback information to subspace learning of pixels and LRR. Since the LRR and subspace learning are simultaneously optimized with the manifold regularization, the feedback information from superpixel is also enhanced for LRR. Thus, LRR and subspace learning in pixel level, combined with the purity control in superpixel level, form a circulation. In this circulation, the relationships between image pixels are learned by LRR. This method is termed as self-supervised low-rank representation (SSLRR). The overview of the SSLRR for HSI classification is shown in Fig. 1. In the SSLRR, the LRR and the constraints defined in the pixel level and superpixel level form a unified objective function. The objective function is solved by means of a customized iterative algorithm. We test our method on three widely used HSI classification data sets. The experimental results show that our method can outperform the related HSI classification methods.

The main contributions of this paper are summarized as follows:

1) The pixel-level and superpixel-level regularizations are explicitly integrated into a unified objective function for LRR.
2) The LRR and spectral–spatial graph regularization are developed as the pixel-level constraints for removing the redundant and noise information in HSIs. Superpixel constraints are further utilized to provide feedback information to the subspace learning to avoid the "salt-and-pepper" problem generated in the pixel-based classification methods and enhance the accuracy of LRR.
3) The solution to the objective function is achieved by employing a customized iterative algorithm, and it converges very fast. Thus, the proposed method is very effective and efficient for unsupervised HSI classification. It far outperforms many recently proposed methods [i.e., LRR and Laplacian regularized LRR (LapLRR)] in terms of classification accuracy.

For clarity, we illustrate important notations and definitions used throughout this paper in Table I.

## II. Related Work

In this section, we briefly review the related studies including LRR, reconstruction independent component analysis (RICA), and manifold learning.

### A. LRR

Given a data matrix $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n] \in \mathbb{R}^{d\times n}$, whose columns are $n$ data samples drawn from independent subspaces.

Then, each column is represented by a linear combination of bases in a given data set $\mathbf{B}$. This problem can be formulated as

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_*, \quad \text{s.t. } \mathbf{X} = \mathbf{BZ} \tag{1}$$

where $\mathbf{B}$ is the base matrix and $\|\cdot\|_*$ represents the nuclear norm of a matrix. In real applications, we often choose the data matrix as the base matrix; thus, (1) can be rewritten as

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_*, \quad \text{s.t. } \mathbf{X} = \mathbf{XZ}. \tag{2}$$

Considering the noise in the data, a noise term $\mathbf{E}$ is always added to (2)

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_* + \beta\|\mathbf{E}\|_{2,1}, \quad \text{s.t. } \mathbf{X} = \mathbf{XZ} + \mathbf{E} \tag{3}$$

TABLE I
NOTATIONS AND DEFINITIONS

| Notation | Definition |
|---|---|
| $\mathbf{A}$ | Database images. $\mathbf{A}\in\mathbb{R}^{d\times n1\times n2}$ |
| $\mathbf{X}$ | The matrix of the original feature. $\mathbf{X}\in\mathbb{R}^{d\times n}$. |
| $\mathbf{W}$ | The dimension reduction matrix. $\mathbf{W}\in\mathbb{R}^{r\times d}$ . |
| $\mathbf{Z}$ | The low-rank representation matrix. $\mathbf{Z}\in\mathbb{R}^{n\times n}$ |
| $\mathbf{E}$ | The error matrix. $\mathbf{E}\in\mathbb{R}^{r\times n}$ |
| $\mathbf{Y}$ | The label matrix. $\mathbf{Y}\in\mathbb{R}^{n\times c}$ |
| $\mathbf{S}$ | The reconstruction coefficient matrix. |
| $\mathbf{F}$ | The classifier matrix. $\mathbf{F}\in\mathbb{R}^{n\times c}$ |
| $\mathbf{G}$ | The adaptive graph. |
| $\mathbf{H}$ | The weight matrix of $\mathbf{G}$. $\mathbf{H}\in\mathbb{R}^{n\times n}$ |
| $\mathbf{L}$ | The Laplacian matrix. |
| $\mathbf{I}$ | The identity matrix. |
| $\alpha,\beta,\gamma,\delta_s$ | They are used to balance the importance of the corresponding term. |
| $\lambda_1,\lambda_2,\lambda_3,\lambda_4$ | They are used to balance the importance of the corresponding term. |
| $D$ | The original number of bands. |
| $R$ | The updated number of bands after dimension reduction |
| $N$ | The number of image pixel. |
| $C$ | The number of classes. |
| $p_i$ | The location of image pixel $x_i$. |
| $\mathcal{N}_k(x_j)$ | The k-nearest neighbors of image pixel $x_j$. |
| $K$ | The number of chosen neighbors. |
| $\|\bullet\|_F$ | Frobenius norm. |
| $tr(\bullet)$ | The trace of the matrix. |

where $\|\mathbf{E}\|_{2,1} = \sum_{j=1}^{n}\sqrt{\sum_{i=1}^{m}([\mathbf{E}]_{ij})^2}$ is called $l_{2,1}$-norm [40] and $\beta$ is used to balance the effect of noise. $l_{2,1}$-norm encourages the columns of $\mathbf{E}$ to be 0, which assumes that the corruptions are "sample specific," i.e., some data vectors are corrupted and the others are clean [3].

### B. RICA

Given the unlabeled data set X, the optimization problem of the standard ICA for estimating independent components [41], [23] is generally defined as

$$\min_{\mathbf{W}} \frac{1}{d}\sum_{i=1}^{d} g(\mathbf{WX}) \quad \text{s.t. } \mathbf{WW}^T = \mathbf{I} \tag{4}$$

where $g$ is the nonlinear convex penalty function, $\mathbf{W}$ is the data transformation matrix, and $\mathbf{I}$ is the identity matrix. In addition, the orthonormality constraint is traditionally utilized to prevent the row vectors in $\mathbf{W}$ from becoming degeneration. A widely used smooth penalty function is $g(\cdot) = \log(\cosh(\cdot))$ [42].

RICA [21] uses a soft reconstruction cost to replace the orthonormality constraint in ICA. By applying this replacement, RICA can be formulated as the following unconstrained problem:

$$\min_{\mathbf{W}} \|\mathbf{W}^T\mathbf{WX} - \mathbf{X}\|_F^2 + \alpha g(\mathbf{WX}) \tag{5}$$

where $\alpha$ is the tradeoff between the reconstruction error and sparsity. By swapping the orthonormality constraint with a reconstruction penalty, RICA learns sparse representations even on unwhitened data when $\mathbf{W}$ is overcomplete [23].

### C. Manifold Learning

Many existing subspace clustering methods fail to discover the intrinsic geometry structure of the data manifold [1]. From the manifold assumption [43], we know that if two data points such as $\mathbf{X}_i$ and $\mathbf{X}_j$ are close in the intrinsic geometry of the data manifold, the representations of the two data points are also close to each other. Lots of efforts on manifold learning [43], [44] have shown that the local geometric structure of the data manifold is effectively modeled through a nearest-neighbor graph on the sampled data points. A nearest-neighbor graph is usually used to characterize the local geometry of the data manifold [1]. For data set $\mathbf{X}$, we build a nearest-neighbor graph $\mathbf{G}$ with its node corresponding to the data point. The nearest neighbors of each vertex are selected according to weight matrix $\mathbf{Z}$ between one sample and other samples. Under the manifold assumption, i.e., if two data samples have a close relationship, their representations are close to each other, the corresponding objective function is expressed as

$$\Omega = \frac{1}{2}\sum_{i,j=1}^{n}\mathbf{Z}_{ij}\|\mathbf{X}_i - \mathbf{X}_j\|_2^2 = \text{tr}(\mathbf{XLX}^T) \tag{6}$$

where $\mathbf{L}$ is the Laplacian matrix. The manifold learning has been applied to improve various kinds of algorithms [1], [45]–[49].

## III. SELF-SUPERVISED LOW-RANK REPRESENTATION

In this section, we discuss the details of the proposed method, i.e., SSLRR. We first present the motivation of SSLRR, and subsequently describe the objective function of SSLRR.

### A. Problem Formulation

Given an HSI, the 3-D HSI data set $\mathbf{A} \in \mathbb{R}^{d\times n1\times n2}$ can be built, where $d$ denotes the number of bands and $n_1 \times n_2$ denotes the number of pixels in each band. Then, we put the 3-D $\mathbf{A}$ into a matrix $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n] \in \mathbb{R}^{d\times n}$, where $n = n_1 \times n_2$.

To construct better relationships between image pixels for HSI classification, the SSLRR is proposed in this paper under the framework of LRR. LRR unifies the constraints of pixel level and superpixel level to construct the latent relationships for image pixels. In the pixel level, a subspace learning method based on RICA is introduced to remove the redundant and noise information. By means of the spectral information and spatial correlation among pixels in the HSI, the spectral–spatial graph is set up to embed the latent relationships into subspace learning. Thus, the LRR and subspace learning can be simultaneously optimized. In superpixel level, an adaptive mean shift-clustering algorithm based on the obtained subspace of pixels is employed to generate superpixels, which provides feedback information to the subspace learning and LRR in the

TABLE II
LAND COVER CLASSES WITH SAMPLES' NUMBER FOR THE INDIAN PINES DATA SET

| Class | Land Cover Type | Labeled samples | Testing |
|---|---|---|---|
| 1 | Corn-notill | 20 | 1408 |
| 2 | Corn-mintill | 20 | 810 |
| 3 | Grass-pasture | 20 | 463 |
| 4 | Grass-trees | 20 | 710 |
| 5 | Hay-windrowed | 20 | 458 |
| 6 | Soybean-notill | 20 | 952 |
| 7 | Soybean-mintill | 20 | 2435 |
| 8 | Soybean-clean | 20 | 573 |
| 9 | Woods | 20 | 1245 |
| 10 | Buildings-Grass-Trees-Drives | 20 | 366 |

pixel level. The feedback information is also considered as the constraints of subspace learning and LRR. Thus, the LRR and the constraints defined in the pixel level and superpixel level form a unified objective function.

### *B. Objective Function of SSLRR*

Considering the conditions defined in the pixel level and superpixel level, the objective function of SSLRR is constructed as follows:

$$\begin{aligned}\min_{\mathbf{W},\mathbf{Z},\mathbf{E}} \ & \Theta(\mathbf{W},\mathbf{Z},\mathbf{E}) = \min_{\mathbf{W},\mathbf{Z},\mathbf{E}} \ \|\mathbf{W}^T\mathbf{W}\mathbf{X} - \mathbf{X}\|_F^2 + \alpha g(\mathbf{W}\mathbf{X}) \\ & + \lambda_1(\|\mathbf{Z}\|_* + \beta\|\mathbf{E}\|_{2,1}) \\ & + \lambda_2 \mathrm{tr}((\mathbf{W}\mathbf{X})(\mathbf{L}_\mathbf{Z} + \gamma\,\mathbf{L}_2)(\mathbf{W}\mathbf{X})^T) \\ & + \lambda_3\left(\sum_{i=1}^{n}\exp\left(\frac{\|(\mathbf{W}\mathbf{X})_i - \theta_t\|_2^2}{\delta_s}\right) + \eta\|\mathbf{Z} - \mathbf{Z}_s\|_F^2\right) \\ \text{s.t. } & \mathbf{W}\mathbf{X} = \mathbf{W}\mathbf{X}\mathbf{Z} + \mathbf{E} \end{aligned} \tag{7}$$

where $\mathbf{W}$ is the data transformation matrix of subspace learning, $\mathbf{Z}$ is the relationships constructed by LRR, and $\mathbf{E}$ is the data noise term. If the $i$th pixel belongs to the $t$th superpixel, $\theta_t$ represents the feature of this superpixel. $(\mathbf{Z}_s)_i$ is statistical LRR information of superpixel $i$. $\lambda_1$, $\lambda_2$, and $\lambda_3$ are the tradeoff factors. $\mathbf{L\_Z}$ and $\mathbf{L}_2$ are the Laplacian matrices constructed by the spectral and spatial information, respectively. $\delta_s$ is the heat kernel.

In (7), with LRR, the novel latent relationships between image pixels can be constructed. $\|\mathbf{Z}\|_*$ ensures that the learned relationship matrix has a low rank. Simultaneously, $l_{2,1}$-norm encourages the columns of $\mathbf{E}$ to be 0. While the redundant and noise information exists in the original data, thus, the first term in (6) is introduced to learn the new suitable subspace for HSI classification. The corresponding term of $\lambda_2$ is used to simultaneously optimize the latent relationships and the data subspace with the spectral–spatial graph, which is under the assumptions of manifold learning.

1) For pixels in HSI, if $x_i$ and $x_j$ have a close relationship in terms of spectral information, they have similar data structures in a low-dimensional subspace [1].

TABLE III
LAND COVER CLASSES WITH SAMPLES' NUMBER FOR THE SALINAS DATA SET

| Class | Land Cover Type | Labeled samples | Testing |
|---|---|---|---|
| 1 | Brocoli_green_weeds_1 | 20 | 181 |
| 2 | Brocoli_green_weeds_2 | 20 | 353 |
| 3 | Fallow | 20 | 178 |
| 4 | Fallow_rough_plow | 20 | 119 |
| 5 | Fallow_smooth | 20 | 248 |
| 6 | Stubble | 20 | 376 |
| 7 | Celery | 20 | 338 |
| 8 | Grapes_untrained | 20 | 1107 |
| 9 | Soil_vinyard_develop | 20 | 600 |
| 10 | Corn_senesced_green_weeds | 20 | 308 |
| 11 | Lettuce_romaine_4wk | 20 | 87 |
| 12 | Lettuce_romaine_5wk | 20 | 173 |
| 13 | Lettuce_romaine_6wk | 20 | 72 |
| 14 | Lettuce_romaine_7wk | 20 | 87 |
| 15 | Vinyard_untrained | 20 | 707 |
| 16 | Vinyard_vertical_trellis | 20 | 161 |

TABLE IV
LAND COVER CLASSES WITH SAMPLES' NUMBER FOR THE PAVIAU DATA SET

| Class | Land Cover Type | Labeled samples | Testing |
|---|---|---|---|
| 1 | Asphalt | 20 | 643 |
| 2 | Meadows | 20 | 1845 |
| 3 | Gravel | 20 | 190 |
| 4 | Trees | 20 | 286 |
| 5 | Painted metal sheets | 20 | 115 |
| 6 | Bare Soil | 20 | 483 |
| 7 | Bitumen | 20 | 113 |
| 8 | Self-Blocking Bricks | 20 | 348 |
| 9 | Shadows | 20 | 75 |

2) For pixels in HSI, if $x_i$ and $x_j$ are close spatial distance in terms of spatial information, they have similar data structures in a low-dimensional subspace.

To spectral–spatial graph construction, the information of spectral and spatial are employed to compute the corresponding relationships. Adaptive graphs $\mathbf{G}_1$ and $\mathbf{G}_2$ are constructed to represent the spectral and spatial graphs, respectively.

For the spectral graph, weight matrix $\mathbf{H}_1 \in \mathbb{R}^{n\times n}$ in graph $\mathbf{G}_1$ is constructed by the LRR matrix $\mathbf{Z}$

$$\mathbf{H}_1 = |\mathbf{Z} + \mathbf{Z}^T|/2. \tag{8}$$

In order to construct the Laplacian matrix of spectral graph,$\mathbf{D}_1$ is first computed. It is a diagonal matrix, in which the $(i,i)$th element is equal to the sum of the $i$th row of $\mathbf{H}_1$. Then, the Laplacian matrix of the spectral graph $\mathbf{L\_Z} = \mathbf{D}_1 - \mathbf{H}_1$.

For the spatial graph, weight matrix $\mathbf{H}_2 \in \mathbb{R}^{n\times n}$ in graph $\mathbf{G}_2$ is defined using the following function:

$$\mathbf{H}_{2ij} = \begin{cases} \exp\left(-\dfrac{\|P_i - P_j\|_2^2}{\omega}\right), & x_i \in \mathcal{N}_k(x_j) \text{ or } x_j \in \mathcal{N}_k(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

TABLE V
UNSUPERVISED HSI CLASSIFICATION RESULTS (%) BY SELECTING 20 LABELED SAMPLES FOR EACH CLASS ON THE INDIAN PINES DATA SET

| Class | KNN | LLE | NNLRS | LRR | LapLRR | SSLRR |
|---|---|---|---|---|---|---|
| Corn-notill | 47.59 | 56.11 | 47.02 | 17.63 | 42.26 | 75.99 |
| Corn-mintill | 22.72 | 34.20 | 45.06 | 21.22 | 45.19 | 93.21 |
| Grass-pasture | 76.67 | 55.94 | 57.88 | 77.80 | 76.67 | 93.09 |
| Grass-trees | 88.87 | 93.66 | 89.58 | 83.06 | 98.45 | 99.44 |
| Hay-windrowed | 99.86 | 99.78 | 99.78 | 77.78 | 99.78 | 99.96 |
| Soybean-notill | 43.70 | 49.37 | 60.19 | 30.87 | 50.42 | 93.80 |
| Soybean-mintill | 59.06 | 50.76 | 43.20 | 66.22 | 63.53 | 87.72 |
| Soybean-clean | 23.73 | 31.59 | 36.82 | 12.69 | 29.32 | 97.73 |
| Woods | 95.26 | 87.39 | 97.35 | 66.37 | 93.33 | 93.98 |
| Buildings-Grass-Trees-Drives | 20.77 | 22.13 | 34.97 | 11.44 | 20.77 | 90.71 |
| OA | 58.92 | 58.43 | 59.07 | 48.53 | 62.69 | **90.49** |
| AA | 57.82 | 58.09 | 61.19 | 46.51 | 61.97 | **92.56** |
| Kappa | 0.52 | 0.52 | 0.53 | 0.40 | 0.56 | **0.89** |

*The best results are highlighted in bold.*

where $\mathcal{N}_k(x_j)$ denotes the $k$-nearest neighbors (KNNs) of image pixel $x_j$ according to the spatial distance $\|P_i - P_j\|$. $\omega$ is an alternative parameter. Then, diagonal matrix $\mathbf{D}_2$ is introduced, where the $(i, i)$th element is equal to the sum of the $i$th row of $\mathbf{H}_2$. The Laplacian matrix of the spatial graph $\mathbf{L}_2 = \mathbf{D}_2 - \mathbf{H}_2$.

Under the manifold assumptions, the spectral and spatial graphs are combined into the following equation:

$$\begin{aligned}\mathcal{T} &= \frac{1}{2}\sum_{i,j=1}^{n}(\mathbf{H}_{1ij} + \gamma\,\mathbf{H}_{2ij})\|(\mathbf{WX})_i - (\mathbf{WX})_j\|_2^2 \\ &= \mathrm{tr}((\mathbf{WX})(\mathbf{L\_Z} + \gamma\,\mathbf{L}_2)(\mathbf{WX})^T) \end{aligned} \tag{10}$$

where $\gamma$ is a tradeoff factor.

The corresponding term of $\lambda_3$ is developed to provide the supervised feedback information in the superpixel level, which can also be used for enhancing the qualities of subspace learning and LRR. From an HSI, we generate the superpixels using the adaptive mean shift-clustering algorithm [55]. Each pixel in a superpixel has similar spectral information represented by the cluster center of the pixels. Since the relationships between image pixels are constructed by LRR, this corresponding objective function in (7) is to minimize the average "impurity" of the class distribution of the pixels in each superpixel with LRR optimization. The choice of the superpixel thus attempts to find a consistent overall segmentation in which each segment contains pixels only belonging to one category. The class "impurity" is determined by (7), which represents the class distribution satisfying the following assumptions.

1) Under the same superpixel, pixels should have similar data structures.
2) Under the same superpixel, the constructed relationships between image pixels should be consistent with each other.

The constraint 1) corresponds to subspace learning while the constraint 2) is to regularize the LRR. This method allows us to consider full families of segmentation components rather than a unique, predetermined segmentation. Once trained, the superpixel generation procedure is a parameter free and requires no adjustment of thresholds.

At the beginning, the purity of the superpixels is low due to the insufficient learning of LRR. As the number of the iterations increases, the feedback information from the superpixels to the LRR learning is more accurate. In the optimal stage, more accurate data relationships are achieved in the procedure of LRR. As the superpixels with class purity are derived, the performance of HSI classification is thus greatly enhanced.

Owing to the close dependence of $\mathbf{L\_Z}$ on $\mathbf{Z}$, auxiliary variable $\mathbf{J} \in \mathbb{R}^{n\times n}$ is introduced to separate (7). Then, the whole objective function is transformed into

$$\begin{aligned}\min_{\mathbf{W},\mathbf{Z},\mathbf{E},\mathbf{J}}\ & \Theta(\mathbf{W},\mathbf{Z},\mathbf{J},\mathbf{E}) = \min_{\mathbf{W},\mathbf{Z},\mathbf{E}}\ \|\mathbf{W}^T\mathbf{WX} - \mathbf{X}\|_F^2 + \alpha g(\mathbf{WX}) \\ &+ \lambda_1(\|\mathbf{J}\|_* + \beta\|\mathbf{E}\|_{2,1}) \\ &+ \lambda_2\mathrm{tr}((\mathbf{WX})(\mathbf{L\_Z} + \gamma\,\mathbf{L}_2)(\mathbf{WX})^T) \\ &+ \lambda_3\left(\sum_{i=1}^{n}\exp\left(\frac{\|(\mathbf{WX})_i - \theta_t\|_2^2}{\delta_s}\right) + \eta\|\mathbf{Z} - \mathbf{Z}_s\|_F^2\right) \\ \text{s.t. }& \mathbf{WX} = \mathbf{WXZ} + \mathbf{E},\ \mathbf{Z} = \mathbf{J}\end{aligned} \tag{11}$$

where $\mathbf{L\_Z}$ is constructed by $\mathbf{Z}$.

## IV. OPTIMIZATION OF SSLRR

We adopt the linearized alternating direction method with adaptive penalty (LADMAP) [56] to solve (11). The augmented Lagrangian function of (11) is

$$\begin{aligned}&L(\mathbf{W},\mathbf{Z},\mathbf{J},\mathbf{E},\mathbf{\Psi}_1,\mathbf{\Psi}_2,\mu) \\ &= \|\mathbf{W}^T\mathbf{WX} - \mathbf{X}\|_F^2 + \alpha g(\mathbf{WX}) \\ &\quad + \lambda_1(\|\mathbf{J}\|_* + \beta\|\mathbf{E}\|_{2,1}) \\ &\quad + \lambda_2\mathrm{tr}((\mathbf{WX})(\mathbf{L\_Z} + \gamma\,\mathbf{L}_2)(\mathbf{WX})^T) \\ &\quad + \lambda_3\left(\sum_{i=1}^{n}\exp\left(\frac{\|(\mathbf{WX})_i - \theta_t\|_2^2}{\delta_s}\right) + \eta\|\mathbf{Z} - \mathbf{Z}_s\|_F^2\right) \\ &\quad + \langle\mathbf{\Psi}_1, \mathbf{WX} - \mathbf{WXZ} - \mathbf{E}\rangle + \langle\mathbf{\Psi}_2, \mathbf{Z} - \mathbf{J}\rangle\end{aligned}$$

TABLE VI
UNSUPERVISED HSI CLASSIFICATION RESULTS (%) BY SELECTING 20 LABELED SAMPLES FOR EACH CLASS ON THE SALINAS DATA SET

| Class | KNN | LLE | NNLRS | LRR | LapLRR | SSLRR |
|---|---|---|---|---|---|---|
| Brocoli_green_ weeds_1 | 98.90 | 99.43 | 99.33 | 99.45 | 98.34 | 99.63 |
| Brocoli_green_ weeds_2 | 98.87 | 93.11 | 99.72 | 94.90 | 96.32 | 99.76 |
| Fallow | 84.27 | 67.02 | 70.21 | 80.34 | 84.83 | 99.31 |
| Fallow_rough_ plow | 99.31 | 99.19 | 99.23 | 99.16 | 99.06 | 90.76 |
| Fallow_smooth | 95.56 | 74.42 | 75.19 | 95.56 | 98.39 | 93.55 |
| Stubble | 98.94 | 98.45 | 98.19 | 98.96 | 98.67 | 97.87 |
| Celery | 98.82 | 89.37 | 92.53 | 99.65 | 99.41 | 98.22 |
| Grapes_untrained | 63.69 | 59.71 | 61.06 | 79.49 | 63.14 | 95.48 |
| Soil_vinyard_ develop | 96.33 | 99.35 | 99.51 | 99.71 | 97.50 | 99.69 |
| Corn_senesced_ green_weeds | 83.44 | 89.31 | 88.05 | 90.91 | 82.79 | 97.08 |
| Lettuce_ romaine_4wk | 99.26 | 95.88 | 99.63 | 98.85 | 98.85 | 90.80 |
| Lettuce_ romaine_5wk | 99.57 | 96.17 | 96.17 | 82.08 | 98.95 | 93.64 |
| Lettuce_ romaine_6wk | 98.61 | 87.80 | 89.02 | 94.44 | 95.83 | 81.94 |
| Lettuce_ romaine_7wk | 90.80 | 72.16 | 72.16 | 83.91 | 93.10 | 96.55 |
| Vinyard_ untrained | 66.90 | 93.86 | 93.86 | 56.58 | 69.31 | 93.49 |
| Vinyard_ vertical_trellis | 92.55 | 99.68 | 99.16 | 95.65 | 96.89 | 99.89 |
| OA | 84.63 | 85.36 | 86.42 | 86.56 | 85.04 | **96.49** |
| AA | 91.61 | 88.43 | 89.56 | 90.60 | 91.96 | **95.48** |
| Kappa | 0.83 | 0.84 | 0.85 | 0.85 | 0.83 | **0.96** |

*The best results are highlighted in bold.*

$$
\begin{aligned}
&+\frac{\mu}{2}\left(\|\mathbf{WX}-\mathbf{WXZ}-\mathbf{E}\|_F^2+\|\mathbf{Z}-\mathbf{J}\|_F^2\right)\\
&=\|\mathbf{W}^T\mathbf{WX}-\mathbf{X}\|_F^2+\alpha g(\mathbf{WX})\\
&\quad+\lambda_1(\|\mathbf{J}\|_*+\beta\|\mathbf{E}\|_{2,1})\\
&\quad+\lambda_2\mathrm{tr}((\mathbf{WX})(\mathbf{L\_Z}+\gamma\mathbf{L}_2)(\mathbf{WX})^T)\\
&\quad+\lambda_3\left(\sum_{i=1}^{n}\exp\left(\frac{\|(\mathbf{WX})_i-\theta_t\|_2^2}{\delta_s}\right)+\eta\|\mathbf{Z}-\mathbf{Z}_s\|_F^2\right)\\
&\quad+\frac{\mu}{2}\left(\|\mathbf{WX}-\mathbf{WXZ}-\mathbf{E}+\frac{\mathbf{\Psi}_1}{\mu}\|_F^2+\|\mathbf{Z}-\mathbf{J}+\frac{\mathbf{\Psi}_2}{\mu}\|_F^2\right)\\
&\quad-\frac{1}{2\mu}\left(\|\mathbf{\Psi}_1\|_F^2+\|\mathbf{\Psi}_2\|_F^2\right) \qquad (12)
\end{aligned}
$$

where $\mathbf{\Psi}_1 \in \mathbb{R}^{r\times n}$ and $\mathbf{\Psi}_2 \in \mathbb{R}^{n\times n}$ are the Lagrange multipliers and $\mu > 0$ is the penalty parameter.

**W** is solved when **Z**, **E,** and **J** are fixed. The optimization problem defined in (12) is written as follows:

$$
\begin{aligned}
\min_{\mathbf{W}} L(\mathbf{W}) &= \min_{\mathbf{W}} \|\mathbf{W}^T\mathbf{WX}-\mathbf{X}\|_F^2+\alpha g(\mathbf{WX})\\
&\quad+\lambda_2\mathrm{tr}((\mathbf{WX})(\mathbf{L\_Z}+\gamma\mathbf{L}_2)(\mathbf{WX})^T)\\
&\quad+\lambda_3\sum_{i=1}^{n}\exp\left(\frac{\|(\mathbf{WX})_i-\theta_t\|_2^2}{\delta_s}\right)\\
&\quad+\frac{\mu}{2}\left\|\mathbf{WX}-\mathbf{WXZ}-\mathbf{E}+\frac{\mathbf{\Psi}_1}{\mu}\right\|_F^2. \qquad (13)
\end{aligned}
$$

**J** is solved when **W**, **Z**, and **E** are fixed. The optimization problem defined in (12) is written as follows:

$$
\begin{aligned}
\min_{\mathbf{J}} L(\mathbf{J}) &= \min_{\mathbf{J}} \lambda_1\|\mathbf{J}\|_*+\frac{\mu}{2}\|\mathbf{J}-\left(\mathbf{Z}+\frac{\mathbf{\Psi}_2}{\mu}\right)\|_F^2\\
&\Leftrightarrow \min_{\mathbf{J}} \frac{\lambda_1}{\mu}\|\mathbf{J}\|_*+\frac{1}{2}\left\|\mathbf{J}-\left(\mathbf{Z}+\frac{\mathbf{\Psi}_2}{\mu}\right)\right\|_F^2. \qquad (14)
\end{aligned}
$$

**Z** is solved when **W**, **J**, and**E** are fixed. The optimization problem defined in (12) is written as follows:

$$
\begin{aligned}
\min_{\mathbf{Z}} L(\mathbf{Z}) &= \min_{\mathbf{Z}} \lambda_2\mathrm{tr}((\mathbf{WX})\mathbf{L\_Z}(\mathbf{WX})^T)+\lambda_3\eta\|\mathbf{Z}-\mathbf{Z}_s\|_F^2\\
&\quad+\frac{\mu}{2}\left(\|\mathbf{WX}-\mathbf{WXZ}-\mathbf{E}\|_F^2+\|\mathbf{Z}-\mathbf{J}\|_F^2\right)\\
&= \min_{\mathbf{Z}} \frac{\lambda_2}{2}\sum_{i,j=1}^{n}\mathbf{Z}_{ij}\|(\mathbf{WX})_i-(\mathbf{WX})_j\|_2^2\\
&\quad+\lambda_3\eta\|\mathbf{Z}-\mathbf{Z}_s\|_F^2\\
&\quad+\frac{\mu}{2}\left(\|\mathbf{WX}-\mathbf{WXZ}-\mathbf{E}\|_F^2+\|\mathbf{Z}-\mathbf{J}\|_F^2\right). \qquad (15)
\end{aligned}
$$

**E** is solved when **W**, **J**, and**Z** are fixed. The optimization problem defined in (12) is written as follows:

$$
\begin{aligned}
\min_{\mathbf{E}} L(\mathbf{E}) &= \min_{\mathbf{E}} \lambda_1\beta\|\mathbf{E}\|_{2,1}+\frac{\mu}{2}\left\|\mathbf{E}-\left(\mathbf{WX}-\mathbf{WXZ}+\frac{\mathbf{\Psi}_1}{\mu}\right)\right\|_F^2\\
&\Leftrightarrow \min_{\mathbf{E}} \frac{\lambda_1\beta}{\mu}\|\mathbf{E}\|_{2,1}+\frac{1}{2}\left\|\mathbf{E}-\left(\mathbf{WX}-\mathbf{WXZ}+\frac{\mathbf{\Psi}_1}{\mu}\right)\right\|_F^2. \qquad (16)
\end{aligned}
$$

TABLE VII
UNSUPERVISED HSI CLASSIFICATION RESULTS (%) BY SELECTING 20 LABELED SAMPLES FOR EACH CLASS ON THE PAVIAU DATA SET

| Class | KNN | LLE | NNLRS | LRR | LapLRR | SSLRR |
|---|---|---|---|---|---|---|
| Asphalt | 18.20 | 14.93 | 35.83 | 4.04 | 18.82 | 49.92 |
| Meadows | 82.87 | 79.46 | 65.44 | 49.49 | 82.76 | 89.16 |
| Gravel | 50.53 | 46.32 | 81.00 | 81.05 | 77.89 | 87.89 |
| Trees | 92.66 | 97.90 | 89.86 | 93.36 | 90.91 | 88.11 |
| Painted metal sheets | 99.56 | 99.66 | 94.40 | 99.13 | 99.13 | 99.89 |
| Bare Soil | 23.40 | 4.97 | 57.20 | 12.22 | 26.71 | 97.93 |
| Bitumen | 69.03 | 25.66 | 99.32 | 34.51 | 50.44 | 95.58 |
| Self-Blocking Bricks | 79.31 | 79.89 | 62.57 | 16.67 | 57.76 | 85.63 |
| Shadows | 93.33 | 97.33 | 47.06 | 88.00 | 98.67 | 53.33 |
| OA | 64.89 | 59.76 | 63.66 | 41.39 | 64.20 | **83.49** |
| AA | 67.65 | 60.68 | 70.30 | 53.16 | 67.01 | **83.05** |
| Kappa | 0.53 | 0.46 | 0.55 | 0.30 | 0.53 | **0.78** |

*The best results are highlighted in bold.*

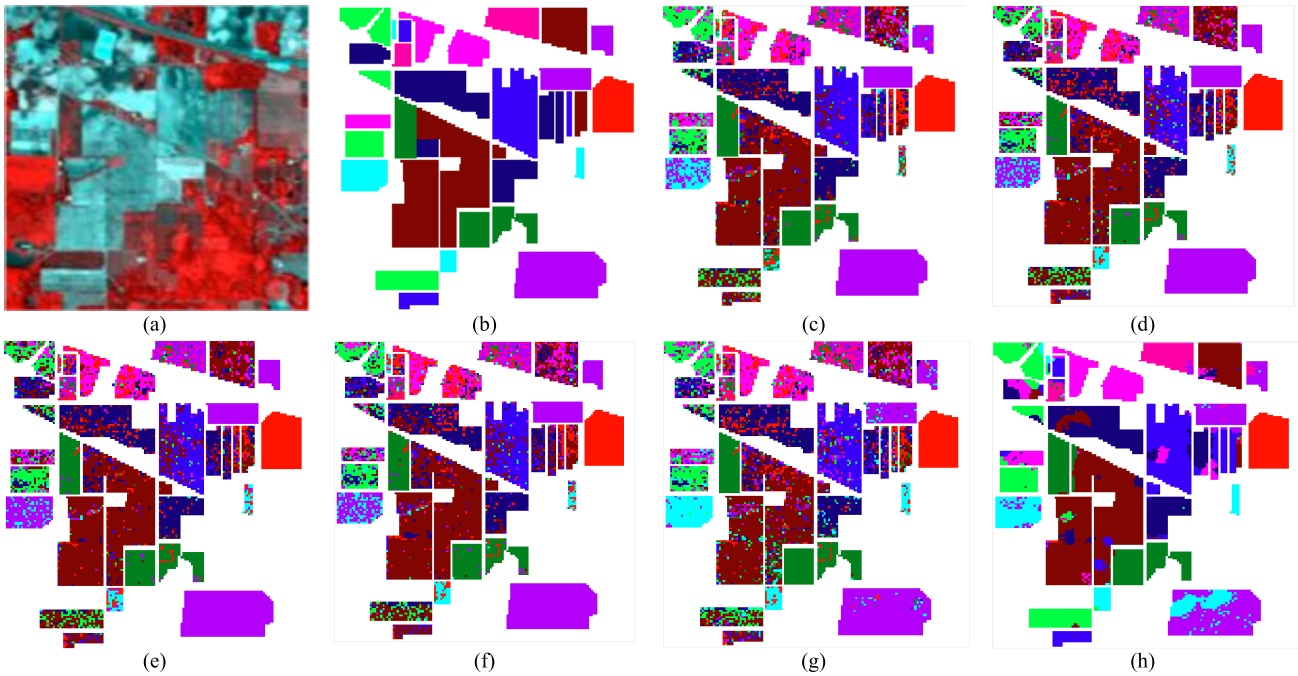



Fig. 2. Classification maps of the Indian Pines data set by selecting three labeled samples for each class. (a) False-color image. (b) Ground truth. (c) Classification map obtained by KNN. (d) Classification map obtained by LLE. (e) Classification map obtained by NNLRS. (f) Classification map obtained by LRR. (g) Classification map obtained by LapLRR. (h) Classification map obtained by our method.

The Lagrangian multipliers are updated as follows:

$$(\mathbf{\Psi}_1)_{\text{new}} = (\mathbf{\Psi}_1)_{\text{old}} + \mu(\mathbf{WX} - \mathbf{WXZ} - \mathbf{E})$$
$$(\mathbf{\Psi}_2)_{\text{new}} = (\mathbf{\Psi}_2)_{\text{old}} + \mu(\mathbf{Z} - \mathbf{J}). \quad (17)$$

So far, the solutions of all variables are obtained. We develop Algorithm 1 to summarize the procedure.

From Algorithm 1, we know that the computational cost of the proposed method mainly lies in updating the variables: **W**, **J**, **Z,** and **E**. **W** is computed using the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) in each iteration. There is a need to compute the objective function cost and grad, whose complexities are $O\ (drd + rdn + dn^2)$ and $O\ (rdn + dn^2 + dnd + rdr + drd)$. **J** is computed with the cost of $O\ (nm^2)$, whereas **Z** is computed with the cost of $O\ (rdn + drd + dn^2)$. The complexity for computing **E** is $O\ (rdn + dn^2)$. As $c \ll n$ and $r < d$, and the linearized method is adopted, our proposed method can converge quickly.

More details of SSLRR optimization can be referred to the Appendixes.

## V. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed method SSLRR for HSI classification. We first briefly describe the used HSI data. Afterward, we compare the classification results of the SSLRR with those of the related approaches.

### A. Experimental Data Sets

Three HSI data sets are used to evaluate the performance of the proposed method.

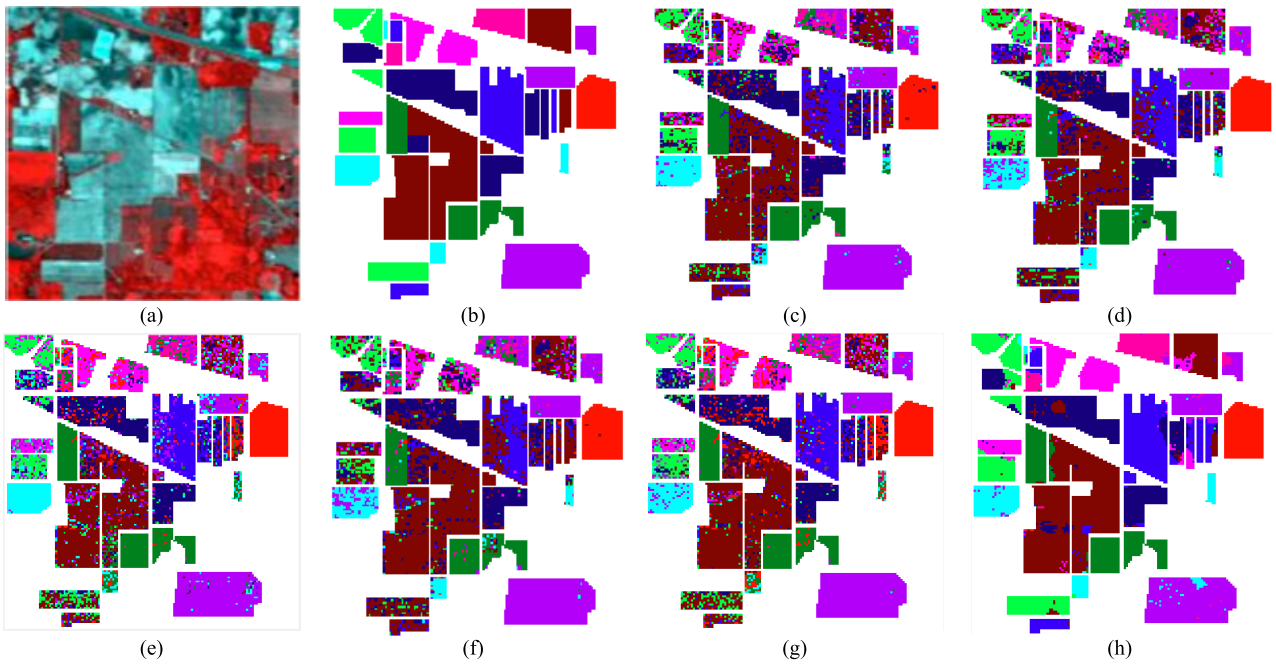



Fig. 3. Classification maps of the Indian Pines data set by selecting 20 labeled samples for each class. (a) False-color image. (b) Ground truth. (c) Classification map obtained by KNN. (d) Classification map obtained by LLE. (e) Classification map obtained by NNLRS. (f) Classification map obtained by LRR. (g) Classification map obtained by LapLRR. (h) Classification map obtained by our method.

**Algorithm 1** SSLRR

**Input:** Training set $\mathbf{X}$, parameters$\alpha, \beta, \gamma, \eta, \lambda_1, \lambda_2, \lambda_3, \lambda_4$;
**Initialization**: $\mathbf{W}, \mathbf{Z}, \mathbf{J}, \mathbf{E}, \mathbf{L}_2, \mathbf{\Psi}_1, \mathbf{\Psi}_2$; $\mu = 10^4$, $\mu_{max} = 10^{10}$, $\rho = 1.1$, $\varepsilon_1 = 10^{-6}$, $\varepsilon_2 = 10^{-2}$;
**Procedure:**
1: **while** not converged do
2: Fix the others and update $\mathbf{W}$, $\mathbf{J}$, $\mathbf{Z}$ and $\mathbf{E}$ respectively by solving Eq. (13), (14), (15) and (16);
3: Update Lagrange multipliers by Eq. (17);
4: Update $\mu$ by $\mu = \min(\rho\mu, \max \mu)$;
5: Check the convergence conditions:

$$\frac{\|\mathbf{WX} - \mathbf{WXZ} - \mathbf{E}\|_F}{\|\mathbf{X}\|_F} < \varepsilon_1$$
$$\|\mathbf{Z} - \mathbf{J}\|_F < \varepsilon_2$$

6: **end while**
**Output:** The matrix $\mathbf{W}$, $\mathbf{Z}$ and $\mathbf{E}$.

The first data set is the Indian Pines data set, which was gathered by AVIRIS sensor over the Indian Pines test site of North-western Indiana in 1992. It consists of $145 \times 145$ pixels and 224 spectral reflectance bands in the wavelength range of 0.4–2.5 $\mu$ m with a spatial resolution of 20 m. The bands covering the region of water absorption (104–108, 150–163, and 220) are removed, and hence 200 out of the 224 bands are preserved. The data set contains 10 classes and 9620 labeled pixels. The detailed information is listed in Table II.

The second data set is the Salinas data set, which was collected by the 224-band AVIRIS sensor over Salinas Valley, CA, USA. Each image size is $512 \times 217$ pixels and is characterized by high spatial resolution (3.7-m pixels). As with Indian Pines data set, 20 water absorption bands (108–112, 154–167, and 224) out of 224 bands are discarded; thus, 204 bands are used in our experiment. The Salinas data set contains 16 classes and 54 129 labeled pixels, as shown in Table III.

The third data set is the University of Pavia (PaviaU) data set, which was acquired by the ROSIS-03 sensor over an urban area, Northern Italy. The spatial size is $610 \times 340$ and the geometric resolution is 1.3 m. The 12 noisy bands are removed, and 103 out of the 115 bands are used in our experiment. There are nine classes and 42 776 labeled pixels in the PaviaU data set. The details are shown in Table IV.

In Tables II–IV, only 20 labeled samples are listed. In the evaluation of our proposed method, different numbers of labeled pixels, such as 3, 5, 10, and 20, are utilized to classify HSIs.

### *B. Alternative Approaches*

Since our proposed method is unsupervised for the relationship construction in HSIs, we compared our method with the following related approaches in terms of HSI classification accuracy.

1) *KNNs:* It uses Euclidean distance as the similarity measure and adopts a Gaussian kernel to reweight the edges.
2) *Local Linear Embedding (LLE) [52]:* In LLE, the linear coefficients that best reconstruct each data point from its neighbors are used to represent the local properties of each neighborhood.
3) *Constructing a Nonnegative Low-Rank and Sparse (NNLRS) Graph With Data-Adaptive Features [3]:* It builds an NNLRS graph to represent the given data, which seeks an NNLRS reconstruction coefficient matrix

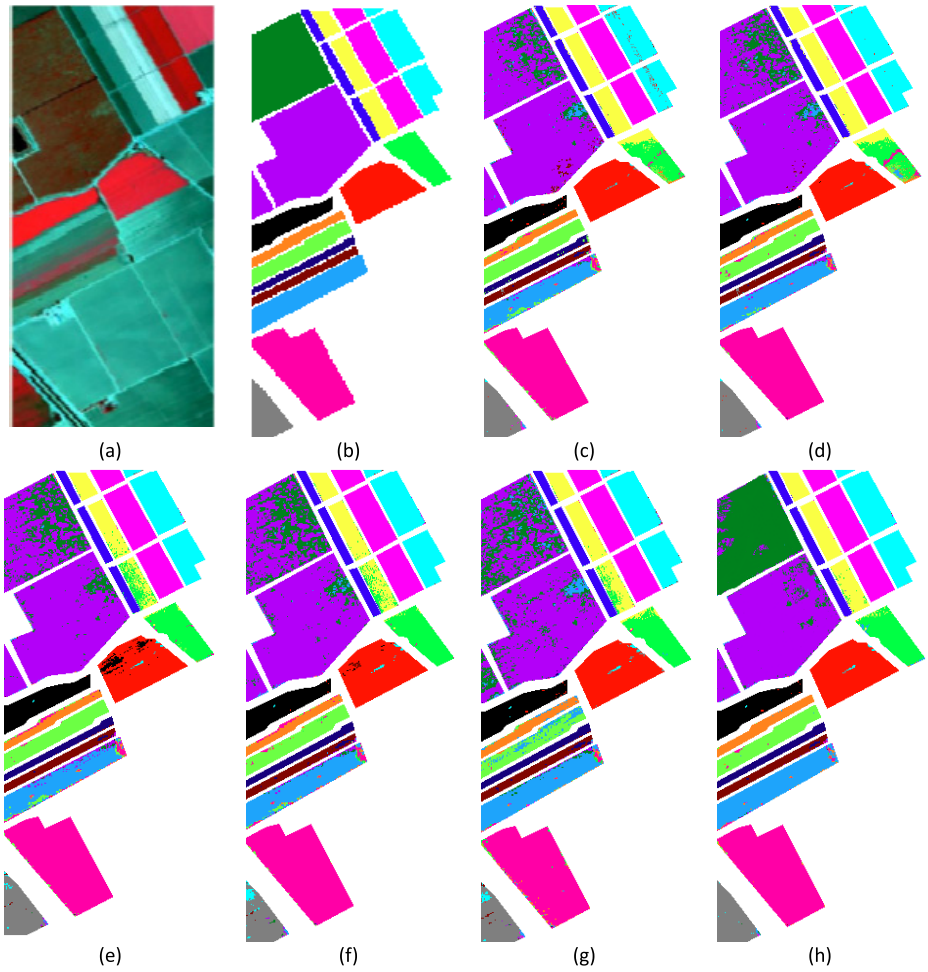



Fig. 4. Classification maps of the Salinas data set by selecting 20 labeled samples for each class. (a) False-color image. (b) Ground truth. (c) Classification map obtained by KNN. (d) Classification map obtained by LLE. (e) Classification map obtained by NNLRS. (f) Classification map obtained by LRR. (g) Classification map obtained by LapLRR. (h) Classification map obtained by our method.

that represents each data sample as a linear combination of others to obtain the weights of edges in the graph.

4) *Robust Subspace Segmentation by LRR [54]:* It finds the LRR of all data jointly, which is used to define the affinities of an undirected graph.
5) *Enhancing Low-Rank Subspace Clustering by Manifold Regularization (LapLRR) [1]:* In the LapLRR, a manifold regularization characterized by a Laplacian graph has been incorporated into LRR to exploit the local manifold structure of the data, resulting in the proposed Laplacian regularized LRR.

For HSI classification, we use the local and global consistency (LGC) [53] to compare the effectiveness of different methods. In LGC, the labeled data and unlabeled data need to be specified for HSI classification. Thus, we follow the following data settings.

In the KNN and LLE, the number of nearest neighbors is selected from the set {3, 4, 5, 6, 7, 8, 9, 10}. The distances of the pixel features are calculated using $\exp(-(\|\mathbf{X}_i - \mathbf{X}_j\|^2)/\sigma))$ in the KNN and LLE, where $\sigma$ is the heat kernel, and it is selected from the set $\{10^{-9}, 10^{-8}, \ldots, 10^{8}, 10^{9}\}$. The defined graph learning functions give the number of nearest neighbors and compute the pixel feature distances in the NNLRS, LRR, and LapLRR. In the SSLRR, the spatial graph needs to be constructed, in which the numbers of $k$ are chosen from the set {3, 4, 5, 6, 7, 8, 9, 10}, while the spectral graph does not need to specify the number of nearest neighbors and compute the pixel feature distances. $\omega$ is also selected from the set $\{10^{-9}, 10^{-8}, \ldots, 10^{8}, 10^{9}\}$. Among these methods, the balancing parameters such as $\lambda_1$, $\lambda_2$, and $\lambda_3$ are tuned from the set $\{10^{-9}, 10^{-8}, \ldots, 10^{8}, 10^{9}\}$, respectively.

In the Indian Pines data set, we select 100% samples as the classification data. Since there are too many pixels in the Salinas and PaviaU data sets, we randomly select 10% samples as the classification data. We further randomly select $s$ samples of the training data as the labeled data; $s$ is set to 3, 5, 10, and 20, respectively. The remaining data are unlabeled data.

### C. Experimental Results

Three metrics, i.e., overall accuracy (OA), average accuracy, and Kappa coefficient, are used to evaluate the classification results. We report the HSI classification performance of

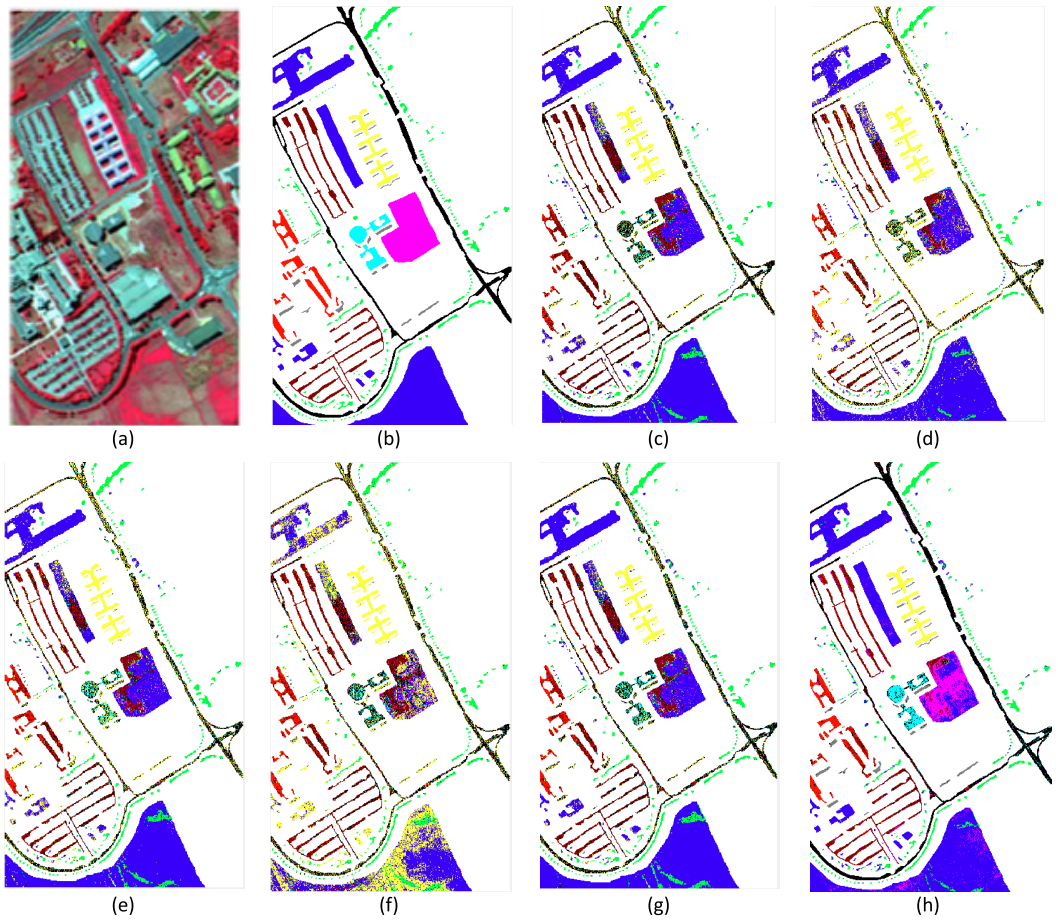



Fig. 5. Classification maps of the PaviaU data set by selecting 20 labeled samples for each class. (a) False-color image. (b) Ground truth. (c) Classification map obtained by KNN. (d) Classification map obtained by LLE. (e) Classification map obtained by NNLRS. (f) Classification map obtained by LRR. (g) Classification map obtained by LapLRR. (h) Classification map obtained by our method.

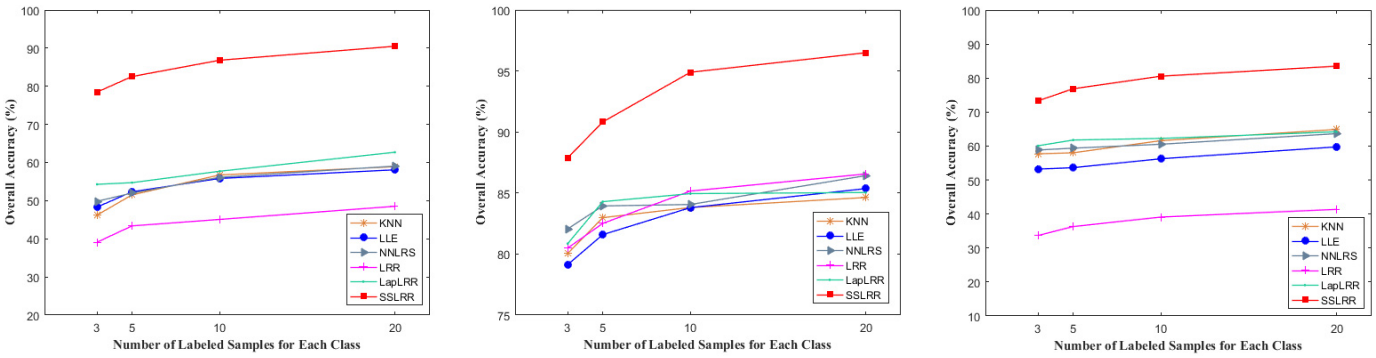



Fig. 6. Classification results with varying numbers of labeled samples by KNN, LLE, NNLRS, LRR, LapLRR, and the proposed method. (a) Indian Pines data set. (b) Salinas data set. (c) PaviaU data set.

different methods (see Tables V–VII, and Figs. 2–6) over 20 random splits on the labeled data set and the unlabeled data set.

From the results listed in Tables V–VII and illustrated in Figs. 2–6, we have the following observations.

1) In comparison with the recently proposed methods, the SSLRR achieves the best classification results for all testing samples. The classification accuracies on the three data sets obtained by the SSLRR are much higher than those obtained by other methods. It indicates that SSLRR can effectively construct the relationships among image pixels, which are very useful for HSI classification.
2) From Figs. 2–5, the SSLRR has more compact HSI classification results on the HSI data sets. It also validates that the superpixel generation procedure can provide useful feedback information for the subspace learning in the pixel level.

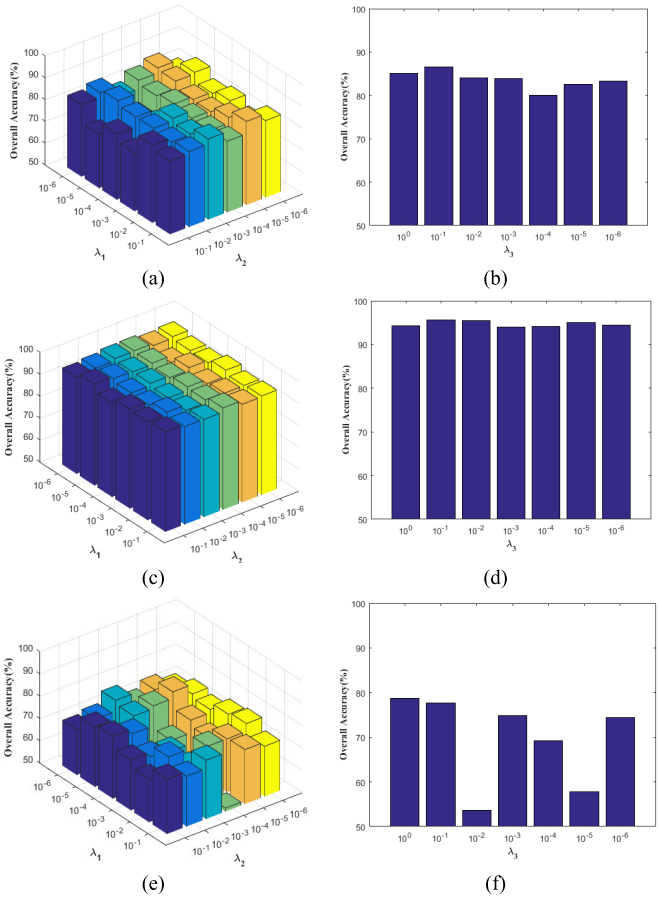



Fig. 7. Influences of different values of the parameters on the HSI classification results. (a) and (b) Influences of different values of $\lambda_1$, $\lambda_2$, and $\lambda_3$ on the classification results for Indian Pines data set. (c) and (d) Influences of different values of $\lambda_1$, $\lambda_2$, and $\lambda_3$ on the classification results for Salinas data set. (e) and (f) Influences of different values of $\lambda_1$, $\lambda_2$, and $\lambda_3$ on the classification results for PaviaU data set.

3) With the number of labeled pixels increasing as shown in Fig. 6, the classification accuracies for all the compared classifiers are increased. In addition, higher overall classification accuracies are obtained by the SSLRR with varying numbers of training samples, demonstrating that the SSLRR outperforms the other compared methods. In the SSLRR, the LRR and the constraints defined in the pixel level and superpixel level form a unified objective function. The feedback information from superpixel level can also offer the evaluation results for LRR. Thus, the constructed spectral–spatial graph well reflects the relationships between image pixels in HSIs. Then, the SSLRR achieves much better HSI classification results.

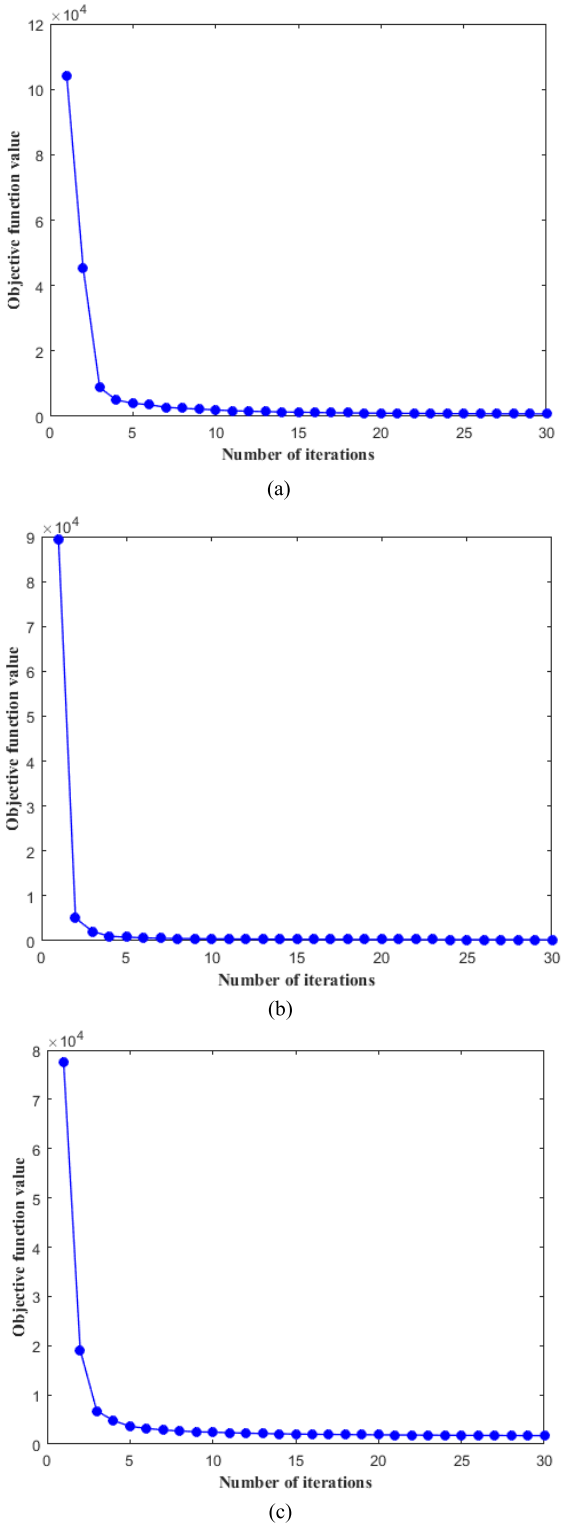



Fig. 8. Convergence processes of different data sets. (a) Indian Pines data set. (b) Salinas data set. (c) PaviaU data set.

### D. Parameter Analysis

In our method, seven parameters including $\alpha$, $\beta$, $\gamma$, $\eta$, $\lambda_1$, $\lambda_2$, and $\lambda_3$ need to be tuned in each data set. We set $\alpha = 0.01$, $\beta = 10$, $\gamma = 1$, and $\eta = 0.01$ in the experiment. We mainly discuss the influences of the main parameters $\lambda_1$, $\lambda_2$, and $\lambda_3$ on the HSI classification results.

$\lambda_1$, $\lambda_2$, and $\lambda_3$ correspond to the terms of LRR, the spectral–spatial graph regularization, and superpixel-level feedback information, respectively. As shown in Fig. 7, we tune the parameters as we classify each data set. The parameters that make the HSI classification results the best are different for

each data set since each data set has its own distinctive data structure. It is found that the value of $\lambda_3$ is larger than those of $\lambda_1$ and $\lambda_2$ among three data sets, which proves that the superpixel-level constraint plays a more important role in HSI classification. From Fig. 7, it is noted that $\lambda_1$ and $\lambda_2$ are nearly the same. This indicates that the constraints of the LRR and the spectral–spatial graph regularization in pixel level are equally important in HSI classification.

### *E. Algorithmic Convergence*

Solving **W**, **Z**, and **E** in (7) simultaneously is very difficult due to the highly nonlinear nature of (7). With the least-squares quantization, we adopt a customized iterative algorithm to optimize the variables. The objective function can converge to a local optimum by using Algorithm 1. With the auxiliary variable added, four variables **W**, **Z**, **E**, and **J** need to be optimized in (11). In each iteration, with the help of the LADMAP, the process for optimizing **W** makes the objective function achieve a local minimum as other variables are fixed. The function of optimizing **Z** is convex, and thus it is convergent. **E** and **J** are optimized with the singular value thresholding (SVT) operator and $l_{2,1}$ minimization operator, respectively. With the optimized variables, the objective function can converge to a local optimum.

The convergence processes under different data sets are shown in Fig. 8. It is noted that our proposed objective function can converge to a local optimum (or even a global minimum) and converge very fast. The objective function defined in (11) usually reaches the convergence within about five iterations for each HSI data set. Therefore, the proposed solution in Algorithm 1 is very effective.

## VI. Conclusion

In this paper, a novel LRR method called SSLRR is developed for HSI classification. The main contribution of the SSLRR lies in explicitly integrating the pixel-level and superpixel-level regularizations into an objective function for LRR. Simultaneously, the criterion defined in superpixel level can provide the feedback information to subspace learning of pixels and LRR. Thus, relationship expression between image pixels is enhanced for HSI classification. The solution to the objective function is achieved by employing the LADMAP algorithm, and it converges very fast. Then, the classification efficiency is high. Experimental results on three HSI data sets show the effectiveness of the SSLRR. We obtain the state-of-the-art performances on these data sets and achieve absolute boosts in OA compared to the best LRR method.

In future work, we will combine the SSLRR with deep learning techniques to automatically learn more representative features of the pixels and superpixels for further enhancing performance of HSI classification.

## Appendix

### *A. Optimization for W*

**W** is solved when **Z**, **E**, and **J** are fixed. The optimization problem defined in (12) is written as (13)

$$\begin{aligned}\min_{\mathbf{W}} L(\mathbf{W}) = \min_{\mathbf{W}} \; & \|\mathbf{W}^T\mathbf{W}\mathbf{X} - \mathbf{X}\|_F^2 + \alpha g(\mathbf{W}\mathbf{X}) \\ & + \lambda_2 \mathrm{tr}((\mathbf{W}\mathbf{X})(\mathbf{L\_Z} + \gamma \mathbf{L}_2)(\mathbf{W}\mathbf{X})^T) \\ & + \lambda_3 \sum_{i=1}^{n} \exp\left(\frac{\|(\mathbf{W}\mathbf{X})_i - \theta_t\|_2^2}{\delta_s}\right) \\ & + \frac{\mu}{2}\left\|\mathbf{W}\mathbf{X} - \mathbf{W}\mathbf{X}\mathbf{Z} - \mathbf{E} + \frac{\mathbf{\Psi}_1}{\mu}\right\|_F^2.\end{aligned}$$

The derivative of (13) with respect to **W** is computed, and then we have the following result:

$$\begin{aligned}\frac{\partial L(\mathbf{W})}{\partial \mathbf{W}} = \; & 2\mathbf{W}(\mathbf{W}^T\mathbf{W}\mathbf{X}\mathbf{X}^T + \mathbf{X}\mathbf{X}^T\mathbf{W}^T\mathbf{W} - 2\mathbf{X}\mathbf{X}^T) \\ & + \alpha \frac{\partial g(\mathbf{W}\mathbf{X})}{\partial \mathbf{W}} \\ & + 2\lambda_2 \mathbf{W}\mathbf{X}(\mathbf{L\_Z} + \gamma \mathbf{L}_2)^T\mathbf{X}^T \\ & + 2\lambda_3 \sum_{i=1}^{n} \exp\left(\frac{\|(\mathbf{W}\mathbf{X})_{i\cdot} - \theta_t\|_2^2}{\delta_s}\right) \\ & \times (\mathbf{W}(\mathbf{X}_{i\cdot}\mathbf{X}_{i\cdot}^T) - \theta_t \mathbf{X}_{i\cdot}^T) \\ & + \mu\left(\mathbf{W}\mathbf{X}(\mathbf{I} - \mathbf{Z}) + \frac{\mathbf{\Psi}_1}{\mu} - \mathbf{E}\right) \times (\mathbf{X}(\mathbf{I} - \mathbf{Z}))^T \quad (18)\end{aligned}$$

where

$$\frac{\partial g(\mathbf{W}\mathbf{X})}{\partial \mathbf{W}_{ij}} = \sum_{k=1}^{n} \tanh(\mathbf{W}_{i\cdot}\mathbf{X}_{\cdot k})\mathbf{X}_{jk}. \quad (19)$$

Given a training data matrix **X**, we compute the function cost of (13) and the gradient using (18). Then, the objective function defined in (13) is minimized through the unconstrained optimizer (e.g., L-BFGS) to update **W**. $\mathbf{W}_{i\cdot}$ denotes the $i$th row of matrix **W**.

### *B. Optimization for J*

**J** is solved when **W**, **Z**, and**E** are fixed. The optimization problem defined in 12) is written as (14)

$$\begin{aligned}\min_{\mathbf{J}} L(\mathbf{J}) \; &= \; \min_{\mathbf{J}} \; \lambda_1\|\mathbf{J}\|_* + \frac{\mu}{2}\left\|\mathbf{J} - \left(\mathbf{Z} + \frac{\mathbf{\Psi}_2}{\mu}\right)\right\|_F^2 \\ &\Leftrightarrow \min_{\mathbf{J}} \; \frac{\lambda_1}{\mu}\|\mathbf{J}\|_* + \frac{1}{2}\left\|\mathbf{J} - \left(\mathbf{Z} + \frac{\mathbf{\Psi}_2}{\mu}\right)\right\|_F^2.\end{aligned}$$

We adopt SVT operator [50] to compute the optimal **J**. First, we obtain the singular value decomposition of the matrix $\mathbf{Z} + ((\mathbf{\Psi}_2)/\mu)$ as follows:

$$\mathbf{Z} + \frac{\mathbf{\Psi}_2}{\mu} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (20)$$

where $\mathbf{\Sigma} = \mathrm{diag}(\{\sigma_i\}_{1 \le i \le m})$ ($m$ is the rank) and $\mathbf{U} \in \mathbb{R}^{n \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times m}$ are the orthogonal matrices. Then, we can obtain the optimal solution of **J** with singular value shrinkage

$$\mathbf{J} = \mathbf{U}\mathbf{\Omega}_{\frac{\lambda_1}{\mu}}\mathbf{\Sigma}\mathbf{V}^T \quad (21)$$

where $\mathbf{\Omega}_{(\lambda_1/\mu)}\mathbf{\Sigma} = \mathrm{diag}(\{\sigma_i - (\lambda_1/\mu)\}_+)$.

### C. Optimization for Z

**Z** is solved when **W**, **J**, and **E** are fixed. The optimization problem defined in (12) is written as (15)

$$\begin{aligned}\min_{\mathbf{Z}} L(\mathbf{Z}) &= \min_{\mathbf{Z}} \lambda_2 tr((\mathbf{WX})\mathbf{L\_Z}(\mathbf{WX})^T) + \lambda_3\eta\|\mathbf{Z}-\mathbf{Z}_s\|_F^2 \\ &\quad + \frac{\mu}{2}\left(\|\mathbf{WX}-\mathbf{WXZ}-\mathbf{E}\|_F^2 + \|\mathbf{Z}-\mathbf{J}\|_F^2\right) \\ &= \min_{\mathbf{Z}} \frac{\lambda_2}{2}\sum_{i,j=1}^{n}\mathbf{Z}_{ij}\|(\mathbf{WX})_{i\cdot}-(\mathbf{WX})_{j\cdot}\|_2^2 \\ &\quad + \lambda_3\eta\|\mathbf{Z}-\mathbf{Z}_s\|_F^2 \\ &\quad + \frac{\mu}{2}\left(\|\mathbf{WX}-\mathbf{WXZ}-\mathbf{E}\|_F^2 + \|\mathbf{Z}-\mathbf{J}\|_F^2\right).\end{aligned}$$

The derivative of (15) with respect to **Z** is computed, and then we have the following result:

$$\begin{aligned}\frac{\partial L(\mathbf{Z})}{\partial \mathbf{Z}} &= \frac{\lambda_2}{2}\sum_{i,j=1}^{n}\|(\mathbf{WX})_{i\cdot}-(\mathbf{WX})_{j\cdot}\|_2^2 + 2\lambda_3\eta(\mathbf{Z}-\mathbf{Z}_s) \\ &\quad + \mu((\mathbf{WX})^T(\mathbf{WXZ}+\mathbf{E}-\mathbf{WX})+\mathbf{Z}-\mathbf{J}). \quad (22)\end{aligned}$$

We can set the derivative equation (22) to 0 and directly obtain $Z$

$$\begin{aligned}\mathbf{Z} &= \left((\mathbf{WX})^T\mathbf{WX}+\mathbf{I}+\frac{2\lambda_3\eta}{\mu}\mathbf{I}\right)^{-1} \\ &\quad \times\left(\mathbf{J}-\frac{\lambda_2}{2\mu}\sum_{i,j=1}^{n}\|(\mathbf{WX})_{i\cdot}-(\mathbf{WX})_{j\cdot}\|_2^2\right. \\ &\quad \left. + (\mathbf{WX})^T(\mathbf{WX}-\mathbf{E}) + \frac{2\lambda_3\eta}{\mu}\mathbf{Z}_s\right). \quad (23)\end{aligned}$$

### D. Optimization for E

**E** is solved when **W**, **J**, and**Z** are fixed. The optimization problem defined in (12) is written as (16)

$$\begin{aligned}&\min_{\mathbf{E}} L(\mathbf{E}) \\ &= \min_{\mathbf{E}} \lambda_1\beta\|\mathbf{E}\|_{2,1} + \frac{\mu}{2}\left\|\mathbf{E}-\left(\mathbf{WX}-\mathbf{WXZ}+\frac{\mathbf{\Psi}_1}{\mu}\right)\right\|_F^2 \\ &\Leftrightarrow \min_{\mathbf{E}} \frac{\lambda_1\beta}{\mu}\|\mathbf{E}\|_{2,1} + \frac{1}{2}\left\|\mathbf{E}-(\mathbf{WX}-\mathbf{WXZ}+\frac{\mathbf{\Psi}_1}{\mu})\right\|_F^2. \quad (24)\end{aligned}$$

This problem has been solved by Yang *et al.* [51], and its optimal solution is given by

$$\mathbf{E}_{\cdot i} = \begin{cases}\dfrac{\|\mathbf{Q}_{\cdot i}\|_2 - \frac{\lambda_1\beta}{\mu}}{\|\mathbf{Q}_{\cdot i}\|_2}\mathbf{Q}_{\cdot i}, & \text{if } \|\mathbf{Q}_{\cdot i}\|_2 > \dfrac{\lambda_1\beta}{\mu} \\ 0, & \text{otherwise.}\end{cases} \quad (25)$$

where $\mathbf{Q} = \mathbf{WX} - \mathbf{WXZ} + ((\mathbf{\Psi}_1)/\mu)$ and $\mathbf{E}_{\cdot i}$ denotes the $i$th column of matrix **E**.

## REFERENCES


[1] J. Liu, Y. Chen, J. Zhang, and Z. Xu, "Enhancing low-rank subspace clustering by manifold regularization," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4022–4030, Sep. 2014.

[2] L. Jing, L. Yang, Y. Jian, and M. K. Ng, "Semi-supervised low-rank mapping learning for multi-label classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1483–1491.

[3] L. Zhuang *et al.*, "Constructing a nonnegative low-rank and sparse graph with data-adaptive features," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3717–3728, Nov. 2015.

[4] M. Yin, J. Gao, Z. Lin, Q. Shi, and Y. Guo, "Dual graph regularized latent low-rank representation for subspace clustering," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4918–4933, Dec. 2015.

[5] W. Yang, Y. Gao, Y. Shi, and L. Cao, "MRM-lasso: A sparse multiview feature selection method via low-rank analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 11, pp. 2801–2815, Nov. 2015.

[6] X. Fang, Y. Xu, X. Li, Z. Lai, and W. K. Wong, "Robust semi-supervised subspace clustering via non-negative low-rank representation," *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1828–1838, Aug. 2016.

[7] Y. Lu, Z. Lai, Y. Xu, X. Li, D. Zhang, and C. Yuan, "Low-rank preserving projections," *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1900–1913, Aug. 2016.

[8] G.-F. Lu, Y. Wang, and J. Zou, "Low-rank matrix factorization with adaptive graph regularizer," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2196–2205, May 2016.

[9] P. Li, J. Yu, M. Wang, L. Zhang, D. Cai, and X. Li, "Constrained low-rank learning using least squares-based regularization," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4250–4262, Dec. 2017.

[10] 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.

[11] C.-F. Chen, C.-P. Wei, and Y.-C. F. Wang, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *Proc. CVPR*, Jun. 2012, pp. 2618–2625.

[12] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. ICCV*, Nov. 2011, pp. 1615–1622.

[13] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[14] Y. Gao, R. Ji, P. Cui, Q. Dai, and G. Hua, "Hyperspectral image classification through bilayer graph-based learning," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2769–2778, Jul. 2014.

[15] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.

[16] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.

[17] Q. Shi, L. Zhang, and B. Du, "Semisupervised discriminative locally enhanced alignment for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4800–4815, Sep. 2013.

[18] J. Li, H. Zhang, L. Zhang, X. Huang, and L. Zhang, "Joint collaborative representation with multitask learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5923–5936, Sep. 2014.

[19] 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.

[20] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.

[21] Q. V. Le, A. Karpenko, J. Ngiam, and A. Y. Ng, "ICA with reconstruction cost for efficient overcomplete feature learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 1017–1025.

[22] J. Yang, X. Gao, D. Zhang, and J.-Y. Yang, "Kernel ICA: An alternative formulation and its application to face recognition," *Pattern Recognit.*, vol. 38, no. 10, pp. 1784–1787, Oct. 2005.

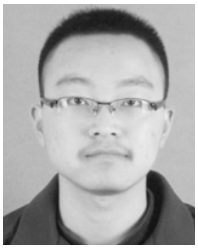
[23] Y. Xiao, Z. Zhu, Y. Zhao, Y. Wei, and S. Wei, "Kernel reconstruction ICA for sparse representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1222–1232, Jun. 2015.

[24] P. Zhong and R. Wang, "Learning conditional random fields for classification of hyperspectral images," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1890–1907, Jul. 2010.

[25] L. Ma, M. M. Crawford, and J. Tian, "Local manifold learning-based $k$-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4099–4109, Nov. 2010.
[26] G. Camps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
[27] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.
[28] Y. Zhou and Y. Wei, "Learning hierarchical spectral–spatial features for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 46, no. 7, pp. 1667–1678, Jul. 2016.
[29] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
[30] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2011, pp. 2097–2104.
[31] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
[32] M. Van den Bergh, X. Boix, G. Roig, and L. Van Gool, "SEEDS: Superpixels extracted via energy-driven sampling," *Int. J. Comput. Vis.*, vol. 111, no. 3, pp. 298–314, 2015.
[33] G. Zhang, X. Jia, and J. Hu, "Superpixel-based graphical model for remote sensing image mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 5861–5871, Nov. 2015.
[34] L. Y. Fang, S. T. Li, X. D. Kang, and J. A. Benediktsson, "Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.
[35] J. Li, H. Zhang, and L. Zhang, "Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338–5351, Oct. 2015.
[36] W. Fu, S. Li, L. Fang, and J. A. Benediktsson, "Adaptive spectral–spatial compression of hyperspectral image with sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 671–682, Feb. 2017.
[37] S. Li, T. Lu, L. Fang, X. Jia, and J. A. Benediktsson, "Probabilistic fusion of pixel-level and superpixel-level hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7416–7430, Dec. 2016.
[38] Y. Wang *et al.*, "Learning a discriminative distance metric with label consistency for scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4427–4440, Aug. 2017, doi: 10.1109/TGRS.2017.2692280.
[39] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1178–1188, Jul. 2008.
[40] J. Liu, S. Ji, and J. Ye, "Multi-task feature learning via efficient $l_{2,1}$ norm minimization," in *Proc. 25th Conf. Uncertainty Artif. Intell.*, 2009, pp. 339–348.
[41] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, NY, USA: Wiley, 2001.
[42] A. Hyvärinen, J. Hurri, and P. Hoyer, *Natural Image Statistics*. Berlin, Germany: Springer-Verlag, 2009.
[43] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16. 2004, pp. 153–160.
[44] H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science*, vol. 290, no. 5500, pp. 2268–2269, 2000.
[45] D. Cai, X. He, and J. Han, "Document clustering using locality preserving indexing," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 12, pp. 1624–1637, Dec. 2005.
[46] X. He, D. Cai, Y. Shao, H. Bao, and J. Han, "Laplacian regularized Gaussian mixture model for data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 9, pp. 1406–1418, Sep. 2011.
[47] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.
[48] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
[49] C. Wang, X. He, J. Bu, Z. Chen, C. Chen, and Z. Guan, "Image representation using Laplacian regularized nonnegative tensor factorization," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2516–2526, 2011.
[50] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[51] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 569–592, 2009.
[52] J. Wang, F. Wang, C. Zhang, H. C. Shen, and L. Quan, "Linear neighborhood propagation and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1600–1615, Sep. 2009.
[53] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Advances in Neural Information Processing Systems 16*. Cambridge, MA, USA: MIT Press, Dec. 2003, pp. 595–602.
[54] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 663–670.
[55] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
[56] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 612–620.

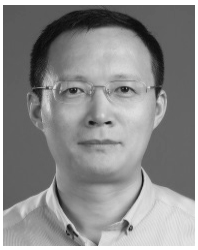
**Yuebin Wang** received the Ph.D. degree from the School of Geography, Beijing Normal University, Beijing, China, in 2016.

He is currently a Post-Doctoral Researcher with the School of Mathematical Sciences, Beijing Normal University. His research interests include remote sensing imagery processing and 3-D urban modeling.



**Jie Mei** is currently pursuing the master's degree with the Faculty of Geographical Science, Beijing Normal University, Beijing, China.

His research interests include remote sensing image processing, and image-based and LiDAR-based segmentation and



**Liqiang Zhang** received the Ph.D. degree in geoinformatics from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2004.

He is currently a Professor with the Faculty of Geographical Science, Beijing Normal University, Beijing. His research interests include remote sensing image processing, 3-D urban reconstruction, and spatial object recognition.

**Bing Zhang** is a Full Professor and the Deputy Director of the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, where he has been leading lots of key scientific projects in the area of hyperspectral remote sensing for more than 20 years. He has authored over 300 publications, including over 190 journal papers. He has edited six books/contributed book chapters on hyperspectral image processing and subsequent applications. He has developed five software systems in the image processing and applications. His research interests include the development of mathematical and physical models and image processing software for the analysis of hyperspectral remote sensing data in many different areas.

Dr. Zhang was a recipient of the National Science Foundation for Distinguished Young Scholars of China in 2013 and the 2016 Outstanding Science and Technology Achievement Prize of the Chinese Academy of Sciences for his special achievements in hyperspectral remote sensing. He is currently serving as the Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He is also the Guest Editor for series of special issues of the IEEE JSTARS, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the PROCEEDINGS OF IEEE, and the *Pattern Recognition Letters*. He has been serving as a Technical Committee Member of the IEEE Workshop on Hyperspectral Image and Signal Processing since 2011, and the President of the Hyperspectral Remote Sensing Committee of the China National Committee of International Society for Digital Earth since 2012.

**Anjian Li** received the B.S. degree in applied mathematics from Beijing Normal University, Beijing, China, in 2017.

His research interests include machine learning and computer vision.

**Yibo Zheng** is currently pursuing the bachelor's degree with the Faculty of Geographical Science, Beijing Normal University, Beijing, China.

Her research interests include remote sensing image processing and classification.

**Panpan Zhu** is currently pursuing the Ph.D. degree with the Faculty of Geographical Science, Beijing Normal University, Beijing, China.

Her research interests include remote sensing image processing, classification, and retrieval.