Unsupervised Dimensionality Reduction With Multifeature Structure Joint Preserving Embedding for Hyperspectral Imagery

Kai Chen, Guoguo Yang, Jing Wang, Qian Du, *Fellow, IEEE*, and Hongjun Su, *Senior Member, IEEE*

***Abstract*—Graph embedding is an effective method that has shown superiority in dimensionality reduction (DR) for hyperspectral imagery (HSI) due to its ability to characterize the intrinsic geometric structure of the data. However, it may ignore some feature information, and the performance of the single model may result in poor classification after DR. In this article, a novel unsupervised DR method, termed multifeature structure joint preserving embedding (MFS-PE), is proposed for hyperspectral image classification. At first, a spatial–spectral model is designed based on the cooperative representation theory, which exploits the potential spatial and spectral features. Then, a neighborhood-constrained model is constructed by implementing sample augmentation through superpixel segmentation, and superpixel labels are used in local enhancement for the spatial–spectral model. Next, a k-nearest neighbor selection method is devised in the local neighborhood-constrained model to select the most suitable neighbors. Finally, both models that can maximize the total scatter of the hyperspectral data to exploit global features are combined to produce an optimal projection for DR. MFS-PE combining multiple feature information can effectively reveal the intrinsic structure of HSIs, and experiments on three publicly available HSI datasets demonstrate that it can offer better classification results compared to the state-of-the-art DR methods.**

***Index Terms*—Dimensionality reduction (DR), graph embedding (GE), hyperspectral image, multifeature.**

I. Introduction

HYPERSPECTRAL imagery (HSI) is a type of high-dimensional data, which is usually captured by satellites and airborne sensors. Each of its pixels is related to solar reflectance in spectral bands, which includes rich spectral features [1], [2], [3], [4]. Accordingly, this characteristic makes it more valuable in applications, such as land-cover species mapping, mineral exploration, food engineering, precision agriculture, and atmospheric research [5], [6], [7], [8]. However, HSI data have high spectral dimensionality, and adjacent bands have high spectral correlation, leading to problems, such as redundancy of information, high computational effort, and difficulties in data storage [9], [10], [11]. In addition, the Hughes phenomenon further limits the accuracy in image classification [12], [13]. To alleviate these problems, dimensionality reduction (DR) of HSI data is an important process to exploit the potential of HSI in real applications [14], [15], [16].

Hyperspectral DR approaches are usually divided into linear and nonlinear methods. Nonlinear DR can uncover the hidden nonlinear data structure and, thus, extract features more effectively than linear DR. Manifold learning as a nonlinear DR method restores the structure of low-dimensional manifolds from high-dimensional sampled data, and it can reveal the essence in the observed phenomena, i.e., the inherent geometric structure of the data [17], [18], [19], [20]. Algorithms, such as Laplacian eigenmaps (LE) [21], isometric feature mapping (ISOMAP) [22], and local linear embedding (LLE) [23], have been widely used for DR. LE constructs an adjacency matrix based on graph theory, which can reconstruct the local structural features of the data manifold [24]. ISOMAP is designed to preserve the essential geometry of the nonlinear data and to approximate the geodesic distance by measuring the shortest path between data points [25]. LLE assumes that the data are linear in small localities and uses the overall nonlinearly distributed data to learn locally linear low-dimensional manifolds of arbitrary dimension [26]. To overcome the out-of-sample problem caused by the inability of the aforementioned algorithms to obtain an explicit projection matrix, researchers have proposed a number of linearized flow learning methods [27], [28], such as locality preserving projections [29] and neighborhood preserving embedding (NPE) [30]. However, the independence of these algorithms limits the subsequent development of manifold learning, so finding commonalities between algorithms becomes the focus of future research.

A GE framework based on statistical or geometric theory has been proposed to redefine many existing DR methods [31], [32]. Based on this framework, representation learning has a better scope in GE-based DR for HSI [33], [34], [35]. Ly et al.

Manuscript received 14 May 2023; revised 28 July 2023; accepted 5 August 2023. Date of publication 10 August 2023; date of current version 24 August 2023. This work was supported by the National Natural Science Foundation of China under Grant 42122008 and Grant 41871220. *(Corresponding author: Hongjun Su.)*

Kai Chen and Hongjun Su are with the School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China, and also with the Jiangsu Province Engineering Research Center of Water Resources and Environment Assessment Using Remote Sensing, Hohai University, Nanjing 211100, China (e-mail: kaichen_hhu@163.com; hjsu@hhu.edu.cn).

Guoguo Yang and Jing Wang are with the School of Geographic Information and Tourism, Chuzhou University, Chuzhou 239000, China (e-mail: yangguoguo@chzu.edu.cn; wangjing@chzu.edu.cn).

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

Digital Object Identifier 10.1109/JSTARS.2023.3304119

[36] applied the sparse representation to the GE framework and proposed sparse graph discriminant analysis (SGDA), which used the l_1-norm as a constraint to query the low-dimensional feature space. While the idea clarified the importance of sparsity in the DR of GE, the geometric structure between the data may not be well preserved. Cao et al. [37] proposed local and global DR sparse representation, which calculated the distance between training and testing samples through a unified metric matrix to extract local and global features of the manifold, which can effectively preserve the geometric structure of the data. Luo et al. [38] proposed semisupervised sparse manifold discriminative analysis using manifold sparse representations to construct intra- and interclass graphs, as well as unsupervised graph, revealing the advantages of sparse manifold properties of data. In addition, Luo et al. [39] proposed multistructure unified discriminative embedding using SR to construct intra- and interclass tangent graphs with tangential information, and incorporate neighborhood and statistical properties into the model to achieve complementarity of different features. However, solving the problem of the l_1-norm-based sparse representation was an iterative process, which results in a significant amount of computational time and operational costs [40], [41]. In [42], collaborative representation (CR) was shown to have better superiority compared to sparse representation. Ly et al. [43] used l_2-norm to construct a collaborative graph based on SGDA and showed that the CR is simple and effective. Liu et al. [44] added a competition term to the CR model, and proposed collaboration-competition preserving GE demonstrated the importance of local information in feature extraction. Yang et al. [45] proposed CR-based projections (CRPs) to fully extract the global features of the data by maximizing the overall scattering matrix. Inspired by the preceding approaches, Jiang et al. [46] proposed Laplacian regularized CR projection (LRCRP) by further research, which combined Laplacian local regular term with maximized total scattering to fully utilize local–global features. However, all the aforementioned studies are conducted in the spectral domain and ignore the spatial factor. Research is still needed to fully exploit the manifold information in spatial and spectral domains.

Spatial attention mechanism plays an increasingly significant role in the field of hyperspectral DR [47], [48], [49]. Jiang et al. [50] embedded spatial features into a CR model and constructed a simple and effective classifier termed joint spatial-aware CR. In fact, the classifier is constructed based on the GE framework, so the model can be applied to DR to exploit spatial–spectral features. Joint spatial-aware collaboration-competitive preserving GE with Tikhonov (JSaCCPGT) proposed by Shah and Du [51] considered collaborative-competitive relationships where multiple feature information was fully extracted to reveal the inherent manifolds of the data. Zhang et al. [52] used homogeneous regions after superpixel segmentation to select suitable nearest neighbors of each sample for local reconstruction, and the proposed spectral–spatial and SuperPCA (S^3-PCA) method can effectively extract spatial and spectral, local and global information. Huang et al. [53] used weighted mean filters for noise removal while a new spatial–spectral combination distance was designed to explore in depth the spatial relationships between the samples. The aforementioned studies demonstrate that the combination of spatial and spectral features can effectively improve the classification accuracy (CA) after DR. However, there may be an upper performance limit on the single model itself, so that even after combining multiple feature information, the accuracy cannot be further improved. Therefore, it is necessary to take into account the strengths and weaknesses of the models and to investigate the models with different mechanisms in depth.

In this article, a new unsupervised GE-based DR model termed multifeature structure joint preserving embedding (MFS-PE) is proposed to make full use of multiple feature information and to tackle the problem of underutilization of feature information by a single model. MFS-PE improves the DR performance by fusing two different GE models. Multiple feature information is then combined in the new model to reveal the full essential structure of the manifold. The main contributions of this article include the following.

1) MFS-PE incorporates two GE models with different mechanisms, which not only solves the performance upper limit of a single model, but also efficiently extracts local and global features without adding regular terms. In addition, the criterion of maximizing global scattering and minimizing local scattering is applied to the new model to enhance the local–global features.
2) In order to reduce the impact on the model when the number of samples is small, superpixel segmentation is added to amplify the samples without labels, and use the labels after superpixel segmentation for local enhancement to further improve the model performance.
3) A new k-nearest neighbor (KNN) selection method is devised to select more suitable neighbors. And three public datasets are used to show the superiority of the proposed method.

The rest of this article is organized as follows. Related work is reviewed, including GE, CR, and superpixel segmentation, in Section II. Section III illustrates in detail the principle of the proposed MFS-PE method. The results of the experiment are discussed in Section IV. Section V analyzes parameter selection. Finally, Section VI concludes this article.

II. Related Work

Let HSI data with M samples and N bands be denoted as $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_M] \in \boldsymbol{R}^{N\times M}$. The objective of DR is to map $\boldsymbol{X} \in \boldsymbol{R}^{N\times M}$ into $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \ldots, \boldsymbol{y}_M] \in \boldsymbol{R}^{d\times M}$, where $d \ll N$ is the low-dimensional feature space. When dealing with linear DR methods, the low-dimensional embedding feature matrix $\boldsymbol{Y} \in \boldsymbol{R}^{d\times M}$ is computed via explicit projection $\boldsymbol{Y} = \boldsymbol{P}^T \boldsymbol{X}$, where $\boldsymbol{P} \in \boldsymbol{R}^{N\times d}$ denotes the projection matrix.

A. GE Framework

The GE framework is a method for capturing the geometric properties of high-dimensional data for DR, which unifies most classical feature extraction methods [31], [38], [54]. The objective of GE is to describe statistical or geometric features of the data by constructing an intrinsic graph $\boldsymbol{G} = \{\boldsymbol{X}, \boldsymbol{W}\}$ and a penalty graph $\boldsymbol{G}^P = \{\boldsymbol{X}, \boldsymbol{W}^P\}$, where $\boldsymbol{X}$ is the vertex

set, and $\boldsymbol{W} \in \boldsymbol{R}^{M\times M}$ and $\boldsymbol{W}^P \in \boldsymbol{R}^{M\times M}$ are the two weight matrices [55], [56].

In GE, the similarity between samples in a high-dimensional space can be represented by a graph in a low-dimensional space. The optimal objective function of the GE framework can be given by

$$\min_{tr\left(\boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y}\right)=h} \frac{1}{2}\sum_{i\neq j}\|\boldsymbol{y}_i-\boldsymbol{y}_j\|^2 w_{ij} = \min_{tr\left(\boldsymbol{Y}^T\boldsymbol{H}\boldsymbol{Y}\right)=h} tr\left(\boldsymbol{Y}^T\boldsymbol{L}\boldsymbol{Y}\right) \quad (1)$$

where h is a constant, $\boldsymbol{L}$ is the Laplacian matrix, and $\boldsymbol{H}$ denotes the constraint matrix that prevents trivial solutions from occurring. Typically, $\boldsymbol{H}$ can be set as a Laplace matrix of the penalty graph $\boldsymbol{G}^P$, i.e., $\boldsymbol{H} = \boldsymbol{L}^P$, then the Laplace matrices $\boldsymbol{L}$ and $\boldsymbol{L}^P$ are given by

$$\boldsymbol{L} = \boldsymbol{D}-\boldsymbol{W},\ \boldsymbol{D}_{ii}=\sum_{j\neq i}^{M} w_{ij},\ \boldsymbol{W} = [w_{ij}]_{i,j=1}^{M} \quad (2)$$

and

$$\boldsymbol{L}^P = \boldsymbol{D}^P - \boldsymbol{W}^P,\ \boldsymbol{D}_{ii}^P=\sum_{j\neq i}^{M} w_{ij}^P,\ \boldsymbol{W}^P = \left[w_{ij}^P\right]_{i,j=1}^{M} \quad (3)$$

respectively, where $\boldsymbol{D}$ is a diagonal matrix, and the rest of the position elements are zero.

B. Collaboration Graph Representation

The cooperative representation (CR) consisting of l_2-norm makes full use of the relationship between all training samples to obtain a better weight graph [57]. For each training sample $\boldsymbol{x}_i$ in HSI, a typical corepresentation optimization objective function is

$$\min_{\boldsymbol{w}_i}\|\boldsymbol{w}_i\|_2 \text{ such that } \|\boldsymbol{x}_i-\boldsymbol{X}_i\boldsymbol{w}_i\|_2^2 \leq \varepsilon \quad (4)$$

where $\boldsymbol{X}_i \in \boldsymbol{R}^{N\times M-1}$ is the original training sample and serves as the dictionary without the current reconstructed sample $\boldsymbol{x}_i$ itself, $\varepsilon > 0$ is a tolerance, and $\boldsymbol{w}_i = [\boldsymbol{w}_{i,1}, \boldsymbol{w}_{i,1}, \ldots, \boldsymbol{w}_{i,M-1}]^T$ is the coefficient of CR for the ith sample with size of $(M-1)$. By regularizing the constraints on the CR coefficients, the optimal objective function can be reformulated as

$$\arg\min_{\boldsymbol{w}_i}\|\boldsymbol{x}_i-\boldsymbol{X}_i\boldsymbol{w}_i\|_2^2+\lambda\|\boldsymbol{w}_i\|_2^2 \quad (5)$$

in which λ is a hyperparameters to balance the contributions of the regularization items.

By equating the first-order derivative to zero, the CR coefficient $\boldsymbol{w}_i$ is calculated as

$$\boldsymbol{w}_i = \left(\boldsymbol{X}_i^T\boldsymbol{X}_i+\lambda\boldsymbol{I}\right)^{-1}\boldsymbol{X}_i^T\boldsymbol{x}_i \quad (6)$$

where $\boldsymbol{I}$ represents an identity matrix of size $(M-1)\times(M-1)$.

C. Superpixel Segmentation

Image segmentation is an important preprocessing process in the field of computer vision and is widely used in many fields due to its effectiveness and efficiency [57], [58], [59]. Superpixel segmentation stands out in image segmentation for its lower computational component and less data redundancy [60], [61].

The HSI, with its contiguous spectrum, provides a detailed description of the richness of features but also presents problems of data redundancy and high computational costs. Superpixels form a homogeneous subgraph by aggregating a number of pixels with similar characteristics in terms of texture, brightness, etc., it can therefore be applied to HSI.

In [52], entropy rate superpixel segmentation provides superior performance in terms of both segmentation efficiency and accuracy. It consists of two parts. The entropy rate of a random walk on a graph contributes to the formation of homogeneous and compact clusters, and the balance term drives the formation of similarly sized subgraphs. An input image can be regarded as a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$, in which $\boldsymbol{V}$ and $\boldsymbol{E}$ denote the sets of vertices and edges, respectively. Graphical construction can be achieved by

$$\max_{A}\mathcal{H}(\boldsymbol{A})+\lambda\mathcal{B}(\boldsymbol{A}) \text{ s.t. } \boldsymbol{A}\subseteq\boldsymbol{E} \quad (7)$$

where $\lambda \geq 0$ is a scale parameter for adjusting the balancing term, and A denotes the set of selected edges.

III. Proposed Methods

In this article, MFS-PE method for unsupervised DR is proposed, which can fully extract the various feature information contained in the HSI and to overcome the problem of upper limit on the performance of individual representation model. MFS-PE first applies mean filtering to the original HSI data for the purpose of removing random noise, and it designs a novel spatial—spectral CR model. This model makes full utilization of spatial–spectral and local–global feature information of samples to obtain better low-dimensional manifold collaboration coefficients. Then, sample augmentation and local enhancement for the model are achieved by superpixel segmentation of the original HSI data, and the augmented samples are applied to a local neighborhood model capable of extracting local geometric structure, which can effectively exploit the neighborhood information. After that, the criteria of global scatter maximization and local scatter minimization are applied to the aforementioned two models to enhance the local–global features. Finally, the spatial–spectral CR model and the local neighborhood model are combined to learn an effective projection matrix. The flowchart of the proposed MFS-PE method is shown in Fig. 1.

A. Spatial–Spectral CRP Model

Hyperspectral image sensors suffer from a lack of brightness, uniformity of brightness, and interaction between components of the circuit, which can lead to noise in the HSI data. In [62], joint CR explains that mean filtering smooths HSI data and removes random noise. Inspired by this, mean filtering is applied as a preprocess to denoise the HSI data, which resulted in new marker

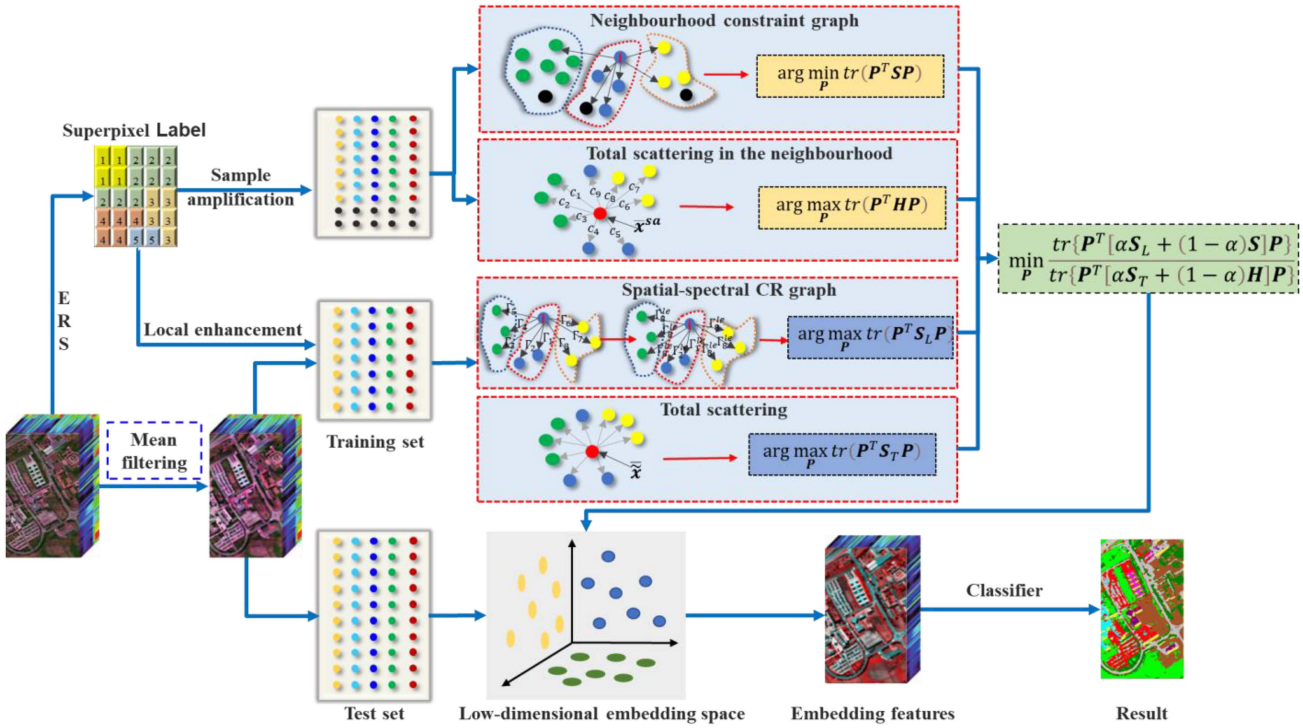

Fig. 1. Flowchart of the proposed MFS-PE method.

symbols, such as $\tilde{\boldsymbol{x}}_i$ and $\tilde{\boldsymbol{x}}$, denoting the training samples and all training data after preprocessing, respectively, as represented by

$$\arg\min_{\boldsymbol{w}_i} \|\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{x}}_i \boldsymbol{w}_i\|_2^2 + \lambda \|\boldsymbol{\Gamma} \boldsymbol{w}_i\|_2^2 + \beta \|\mathrm{diag}\,(\boldsymbol{s})\, \boldsymbol{w}_i\|_2^2 \quad (8)$$

where λ and β are two regularization hyperparameters used to balance the contribution of minimization between the two regularization terms and the residual component. Here, $\boldsymbol{\Gamma}$ is the Tikhonov matrix that represents the spectral distance between the reconstructed sample $\tilde{\boldsymbol{x}}_i$ and the dictionary $\tilde{\boldsymbol{x}}$ given by

$$\begin{aligned} \boldsymbol{\Gamma} &= \begin{pmatrix} \Gamma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Gamma_{(M-1)} \end{pmatrix} \\ &= \begin{pmatrix} \|\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{x}}_{i1}\|_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{x}}_{i(M-1)}\|_2 \end{pmatrix} \end{aligned} \quad (9)$$

in which $\tilde{\boldsymbol{x}}_{i1}, \ldots, \tilde{\boldsymbol{x}}_{i(M-1)}$ are the columns of matrix $\tilde{\boldsymbol{x}}_i$, and $\boldsymbol{s} = [s_1, s_2, \ldots, s_{M-1}]$ is spatial distance between points $\tilde{\boldsymbol{x}}_{i1}$ and $\tilde{\boldsymbol{x}}_{ij} (j = 1, 2, \ldots, M-1)$ with pixel coordinates (a_i, b_i) and (a_{ij}, b_{ij}) calculated as $\boldsymbol{s} = [\mathrm{dist}((a_i, b_i), (a_{ij}, b_{ij}))]^t$ with t being the parameter of smoothing.

However, the spatial–spectral CR model is globally scoped to reconstruct the training sample $\tilde{\boldsymbol{x}}_i$, which allows samples far away from $\tilde{\boldsymbol{x}}_i$ to also play a significant role, which may lead to decrease in model performance. To minimize the negative impact from these samples, we introduce superpixel labels (mentioned later), which can reconstruct $\tilde{\boldsymbol{x}}_i$ much better if other samples have the same superpixel label as the reconstructed sample

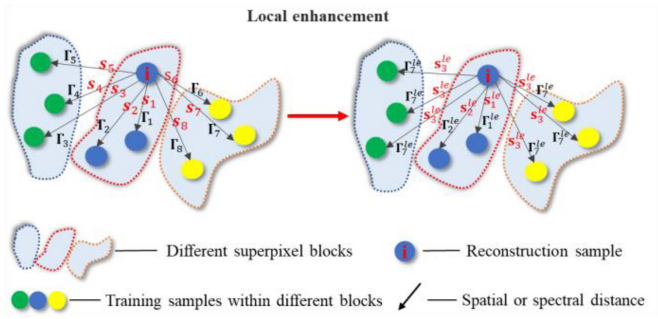

Fig. 2. Local enhancement.

(i.e., they are within the same superpixel block), and conversely, if these sample has a different label from $\tilde{\boldsymbol{x}}_i$, then by setting its distance to the maximum of the distance of all samples from $\tilde{\boldsymbol{x}}_i$ to perform local enhancement. Fig. 2 represents the principle of local enhancement, where Γ_7 and s_3 denote the farthest spectral and spatial distances from sample i, respectively. The new Γ_j^{le} is defined as

$$\Gamma_j^{le} = \begin{cases} \|\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{x}}_{ij}\|_2, \ if\ \tilde{\boldsymbol{x}}_i \text{ has the same label as } \tilde{\boldsymbol{x}}_{ij} \\ \max\{\|\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{x}}_i\|_2\}, \ \text{otherwise} \end{cases} \quad (10)$$

$$\boldsymbol{\Gamma}^{le} = \begin{bmatrix} \Gamma_1^{le} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Gamma_{M-1}^{le} \end{bmatrix} \quad (11)$$

$\boldsymbol{s}^{le}$ is also based on this principle: if $\tilde{\boldsymbol{x}}_i$ has the same label as $\tilde{\boldsymbol{x}}_{ij}$, then $\boldsymbol{s}^{le} = [\mathrm{dist}((a_i, b_i), (a_{ij}, b_{ij}))]^t$; otherwise, $\boldsymbol{s}^{le}$ is the maximum value of the distance between the reconstructed

sample $\tilde{\boldsymbol{x}}_i$ and all samples. Local enhancement based on superpixel labels shows strong performance when the KNNs are not explicitly selected in the data of the Euclidean distance metric.

After local enhancement, by taking the derivative of the coefficient vector $\boldsymbol{w}_i$ by (8) and setting the inverse equation to zero, we can obtain

$$-\tilde{\boldsymbol{x}}_i^T\tilde{\boldsymbol{x}}_i+\tilde{\boldsymbol{x}}_i^T\tilde{\boldsymbol{x}}_i\boldsymbol{w}_i+\lambda\boldsymbol{\Gamma}^{le}+\beta\text{diag}\left(\boldsymbol{s}^{le}\right)=0. \tag{12}$$

The closed-form solution is

$$\boldsymbol{w}_i=\left(\tilde{\boldsymbol{x}}_i^T\tilde{\boldsymbol{x}}_i+\lambda\boldsymbol{\Gamma}^{le}+\beta\text{diag}\left(\boldsymbol{s}^{le}\right)\right)^{-1}\tilde{\boldsymbol{x}}_i^T\tilde{\boldsymbol{x}}_i. \tag{13}$$

Inspired by the work in [14], after deriving the CR graph $\boldsymbol{W}=(\boldsymbol{w}_1,\ldots,\boldsymbol{w}_M)$, the low-dimensional embedding projection can be obtained by simultaneously minimizing local tightness given by

$$\begin{aligned}\boldsymbol{J}_L&=\sum_{i=1}^{M}\left\|\boldsymbol{P}^T\tilde{\boldsymbol{x}}_i-\sum_{j=1}^{M}\boldsymbol{w}_{ij}\boldsymbol{P}^T\tilde{\boldsymbol{x}}_j\right\|^2\\&=tr\left(\boldsymbol{P}^T\left(\sum_{i=1}^{M}\left(\tilde{\boldsymbol{x}}_i-\sum_{j=1}^{M}\boldsymbol{w}_{ij}\tilde{\boldsymbol{x}}_j\right)\right.\right.\\&\quad\left.\left.\times\left(\tilde{\boldsymbol{x}}_i-\sum_{j=1}^{M}\boldsymbol{w}_{ij}\tilde{\boldsymbol{x}}_j\right)^T\right)\boldsymbol{P}\right)\\&=tr\left(\boldsymbol{P}^T\left(\tilde{\boldsymbol{x}}\left(\boldsymbol{I}-\boldsymbol{W}\right)\left(\boldsymbol{I}-\boldsymbol{W}\right)^T\tilde{\boldsymbol{x}}^T\right)\boldsymbol{P}\right)\\&=tr\left(\boldsymbol{P}^T\boldsymbol{S}_L\boldsymbol{P}\right)\end{aligned} \tag{14}$$

where $\boldsymbol{S}_L=\tilde{\boldsymbol{x}}\ (\boldsymbol{I}-\boldsymbol{W}-\boldsymbol{W}^T+\boldsymbol{W}\boldsymbol{W}^T)\tilde{\boldsymbol{x}}^T$ denotes local scattering matrix.

The total separability can be defined by

$$\begin{aligned}\boldsymbol{J}_T&=\sum_{i=1}^{M}\|\boldsymbol{P}^T\tilde{\boldsymbol{x}}_i-\boldsymbol{P}^T\overline{\tilde{\boldsymbol{x}}}\|^2\\&=\sum_{i=1}^{M}\boldsymbol{P}^T\left(\tilde{\boldsymbol{x}}_i-\overline{\tilde{\boldsymbol{x}}}\right)\left(\tilde{\boldsymbol{x}}_i-\overline{\tilde{\boldsymbol{x}}}\right)^T\boldsymbol{P}\\&=tr\left(\boldsymbol{P}^T\left(\sum_{i=1}^{M}\left(\tilde{\boldsymbol{x}}_i-\overline{\tilde{\boldsymbol{x}}}\right)\left(\tilde{\boldsymbol{x}}_i-\overline{\tilde{\boldsymbol{x}}}\right)^T\right)\boldsymbol{P}\right)\\&=tr\left(\boldsymbol{P}^T\boldsymbol{S}_T\boldsymbol{P}\right)\end{aligned} \tag{15}$$

where $\boldsymbol{S}_T=\sum_{i=1}^{M}(\tilde{\boldsymbol{x}}_i-\overline{\tilde{\boldsymbol{x}}})(\tilde{\boldsymbol{x}}_i-\overline{\tilde{\boldsymbol{x}}})^T$, and $\overline{\tilde{\boldsymbol{x}}}$ is the mean of training samples.

Similar to CRP, we minimize local tension and maximize total differentiability for classification purposes, so the low-dimensional embedding projection matrix $\boldsymbol{P}$ can be determined as

$$\begin{cases}\arg\min\limits_{\boldsymbol{P}}tr\left(\boldsymbol{P}^T\boldsymbol{S}_L\boldsymbol{P}\right)\\\arg\max\limits_{\boldsymbol{P}}tr\left(\boldsymbol{P}^T\boldsymbol{S}_T\boldsymbol{P}\right)\end{cases}. \tag{16}$$

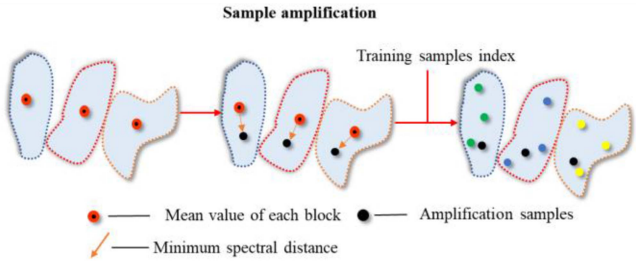

Fig. 3. Sample amplification.

B. Sample Augmented Neighborhood-Constrained Model

The susceptibility of the spectral profiles of HSI pixels to the effects of instrumentation and external conditions resulting in different spectral profiles for pixels of the same feature will limit the subsequent application of HSI [63], [64], [65]. To overcome this problem, superpixel segmentation is used to increase the number of training samples for constructing more robust neighborhood-constrained graph and then designs a new KNN selection method, which is able to select better neighbors.

1) Superpixel-Based Sample Amplification: As a preprocessing step, the most informative first principal component was obtained by principal component analysis (PCA) on the HSI data to reduce segmentation costs. At first, the HSI data are split into $\mathcal{K}$ superpixel subblocks by ERS. Then, the superpixel labels resulted from image segmentation are used with local enhancement of the spatial–spectral CR model, and calculate the mean value of each subblock to represent the overall characteristics of the block. After that, find the pixel with the smallest spectral domain Euclidean distance between the mean point and all points within each subblock. Finally, this pixel replaces the mean value as the sample amplification point, which means that the neighborhood-constrained graph has more robust stability. Since the training samples are selected from the mean filtering process, the index of the training samples is used to query the samples corresponding to the original data, and the queried samples are combined with the augmented samples to form the new training samples. Therefore, the number of amplified samples becomes $M+\mathcal{K}$, which resulted in new marker symbols, such as $\boldsymbol{x}_i^{sa}$ and $\boldsymbol{X}^{sa}$, denoting the training samples and all training data, respectively. The sample amplification process is illustrated in Fig. 3.

2) Novel KNN Selection Method: Traditional KNN method is generally based on spectral domain selection of neighbors. To fully exploit the spatial features, a novel KNN selection method is proposed by using superpixels. First, for each reconstructed sample i $(i=1,2,\ldots,M+\mathcal{K})$, calculate its spatial Euclidean distance from other training samples $\boldsymbol{D}_i=[D_1^{\text{spa}},\ldots,D_{M+\mathcal{K}}^{\text{spa}}]$. Then, the spectral Euclidean distance between the block where the reconstructed sample i is situated and other blocks is calculated based on the mean value of the superpixel block defined as $\boldsymbol{S}_i=[S_1^{\text{spe}},\ldots S_{i-1}^{\text{spe}},0,S_{i+1}^{\text{spe}},S_{\mathcal{K}}^{\text{spe}}]$. Since the values of the HSI data after normalization are very small, the spectral distance between each block will be much smaller than the spatial distance between the training samples, also the spectral distance between blocks indicates to some

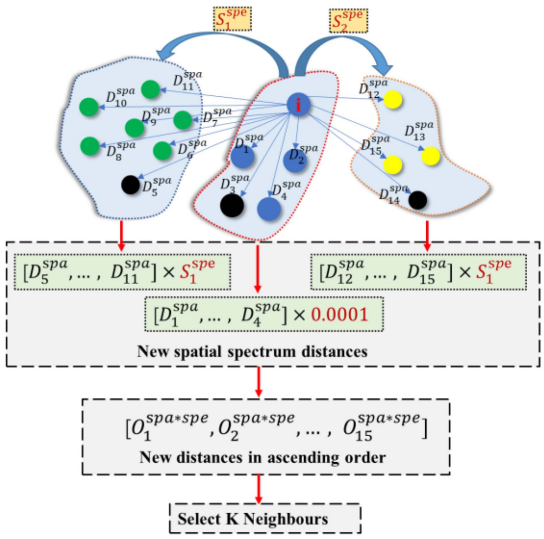

Fig. 4. KNN selection method.

extent the similarity between blocks. So, the spectral distance between blocks is elegantly set as a scaling factor, which physical meaning is that if the training sample in a superpixel block is spatially distant from the reconstructed sample i, but the Euclidean distance between that superpixel block and the superpixel block in which the reconstructed sample is located is close, then the spatial distance between the training sample and the reconstruction sample is multiplied by the Euclidean distance between the two superpixel blocks to obtain a scaled distance. The significance of this method is that it improves the probability that training samples, which are further away from the position of the reconstructed sample, will be selected. It is worth noting that if the number of training samples within the same superpixel block as the reconstructed sample exceeds the value of k, then the k neighbors are selected with high probability within that block, to prevent the block from being unable to select nearest neighbors. This is because the distances after multiplication within the block are all zero, a very small value of 0.0001 is set as the distance of the superpixel block to itself for query purposes, so $\boldsymbol{S}_i$ is redefined as $\boldsymbol{S}_i = [\, S_1^{\text{spe}}, \ldots\, S_{i-1}^{\text{spe}}, 0.0001, S_{i+1}^{\text{spe}}, \ldots\, S_{\mathcal{K}}^{\text{spe}}]$. Finally, the new spatial–spectral distances are sorted in ascending order to obtain a new order $\boldsymbol{O}_i = [O_1^{\text{spa*spe}}, \ldots, O_{M+\mathcal{K}}^{\text{spa*spe}}]$, and select the k neighbors according to that order, which is interpreted in Fig. 4.

3) Neighborhood-Constrained Graph: In the neighborhood-constrained graph $\boldsymbol{G}_N$, the novel KNN selection method is applied to select the most suitable neighborhood for each training sample. Then, the similarity weight w_{ij} between $\boldsymbol{x}_i^{sa}$ and $\boldsymbol{x}_j^{sa}$ can be determined according to the following Gaussian kernel function:

$$w_{ij} = \begin{cases} \exp\left(-\frac{\|(\boldsymbol{x}_i^{sa} - \boldsymbol{x}_j^{sa})\|^2}{2t^2}\right), & \boldsymbol{x}_i^{sa} \text{ and } \boldsymbol{x}_j^{sa} \text{ are neighbors} \\ 0, \text{ otherwise} \end{cases} \tag{17}$$

where $\frac{1}{k}\sum_{i=1}^{k} \|\boldsymbol{x}_i^{sa} - \boldsymbol{x}_j^{sa}\|$.

To fully capture the local feature information of the HSI, the relationship between each point and its KNN in the high-dimensional space should remain constant in the low-dimensional space. The neighborhood optimization problem via the similarity weight matrix can be formulated as

$$\begin{aligned} \boldsymbol{J}(l) &= \sum_{i=1}^{M+\mathcal{K}} \sum_{j=1}^{k} \|\boldsymbol{y}_i^{sa} - \boldsymbol{y}_j^{sa}\|^2 w_{ij} \\ &= \sum_{i=1}^{M+\mathcal{K}} \sum_{j=1}^{k} \|\boldsymbol{P}^T\boldsymbol{x}_i^{sa} - \boldsymbol{P}^T\boldsymbol{x}_j^{sa}\|^2 w_{ij} \\ &= tr\left(\boldsymbol{P}^T\left(\sum_{i=1}^{M+\mathcal{K}} \sum_{j=1}^{k} (\boldsymbol{x}_i^{sa} - \boldsymbol{x}_j^{sa})(\boldsymbol{x}_i^{sa} - \boldsymbol{x}_j^{sa})^T w_{ij}\right)\boldsymbol{P}\right) \\ &= tr\left(\boldsymbol{P}^T\boldsymbol{S}\boldsymbol{P}\right) \end{aligned} \tag{18}$$

where $\boldsymbol{S} = \sum_{i=1}^{M+\mathcal{K}} \sum_{j=1}^{k} (\boldsymbol{x}_i^{sa} - \boldsymbol{x}_j^{sa})(\boldsymbol{x}_i^{sa} - \boldsymbol{x}_j^{sa})^T w_{ij}$, and $\boldsymbol{Y}^{sa} = [\boldsymbol{y}_1^{sa}, \ldots, \boldsymbol{y}_{M+\mathcal{K}}^{sa}]$ is the local neighborhood scattering matrix with a low-dimensional representation of $\boldsymbol{X}^{sa}$ after sample augmentation.

In addition, total scattering matrix maximization is used to improve separability using global features, which can be given by

$$\begin{aligned} \boldsymbol{J}(t) &= \sum_{i=1}^{M+\mathcal{K}} \|\boldsymbol{P}^T\boldsymbol{x}_i^{sa} - \boldsymbol{P}^T\bar{\boldsymbol{x}}^{sa}\|^2 c_i \\ &= \sum_{i=1}^{M+\mathcal{K}} \boldsymbol{P}^T(\boldsymbol{x}_i^{sa} - \bar{\boldsymbol{x}}^{sa})\, c_i (\boldsymbol{x}_i^{sa} - \bar{\boldsymbol{x}}^{sa})^T\boldsymbol{P} \\ &= tr\left(\boldsymbol{P}^T\left(\sum_{i=1}^{M+\mathcal{K}} (\boldsymbol{x}_i^{sa} - \bar{\boldsymbol{x}}^{sa})\, c_i (\boldsymbol{x}_i^{sa} - \bar{\boldsymbol{x}}^{sa})^T\right)\boldsymbol{P}\right) \\ &= tr\left(\boldsymbol{P}^T\boldsymbol{H}\boldsymbol{P}\right) \end{aligned} \tag{19}$$

where $\boldsymbol{H} = \sum_{i=1}^{M+\mathcal{K}} (\boldsymbol{x}_i^{sa} - \bar{\boldsymbol{x}}^{sa}) c_i (\boldsymbol{x}_i^{sa} - \bar{\boldsymbol{x}}^{sa})^T$ is the total scatter matrix, $\bar{\boldsymbol{x}}^{sa}$ denotes the mean value of the amplified sample, and $c_i = \exp(-\|\boldsymbol{x}_i^{sa} - \bar{\boldsymbol{x}}^{sa}\|^2 / 2t_{ik})$ is the weight between $\boldsymbol{x}_i^{sa}$ and $\bar{\boldsymbol{x}}^{sa}$.

By exploring the local neighborhood information in the low-dimensional space and fully considering the global information, the optimal projection matrix can be obtained as

$$\begin{cases} \arg\min\limits_{\boldsymbol{P}} tr\left(\boldsymbol{P}^T\boldsymbol{S}\boldsymbol{P}\right) \\ \arg\max\limits_{\boldsymbol{P}} tr\left(\boldsymbol{P}^T\boldsymbol{H}\boldsymbol{P}\right) \end{cases}. \tag{20}$$

The aforementioned two models can uncover the collaboration and neighborhood structure relationships between pixels in high-dimensional data, respectively. To solve the single model performance upper limit problem, based on (16) and (20), a MFS-PE method is proposed for HSI data DR. The method not only extracts multiple feature information, but also combines the collaborative model with the neighborhood-constrained model to learn more efficient projections, and the optimal objective function of MFS-PE can be designed as

$$J = \min_{\boldsymbol{P}} \frac{tr\left\{\boldsymbol{P}^T\left[\alpha\boldsymbol{S}_L + (1-\alpha)\,\boldsymbol{S}\right]\boldsymbol{P}\right\}}{tr\left\{\boldsymbol{P}^T\left[\alpha\boldsymbol{S}_T + (1-\alpha)\,\boldsymbol{H}\right]\boldsymbol{P}\right\}} \tag{21}$$

Algorithm 1: The Proposed MFS-PE.

Input: High-dimensional training samples $\boldsymbol{X} \in \boldsymbol{R}^{N\times M}$, training label $t \subset \boldsymbol{R}^{M}$, pre-set dimension of the low-dimensional subspace d, the number of superpixel blocks $\mathcal{K}$, hyperparameters λ and β, trade-off parameters α, testing samples $\boldsymbol{X}^{*} \in \boldsymbol{R}^{N\times S}$, testing label $\boldsymbol{t}^{*} \subset \boldsymbol{R}^{S}$.

Output: $\boldsymbol{s} = \{Acc, \boldsymbol{P}\}$

1. Pre-processing of all the training and testing samples by average filtering;
2. Superpixel segmentation of the original HSI;
3. Calculation of the local scattering matrix and the overall scattering matrix of the spatial-spectral CR model by (14) and (15) and local enhancement;
4. Sample unlabeled amplification and selection of spatial spectral k -nearest neighbours;
5. Calculation of the local scattering matrix and the overall scattering matrix of the neighbourhood-constrained model by (18) and (19);
6. Calculation of the projection matrix $\boldsymbol{P}$ by solving the generalized eigenvalues (23);
7. Feature reduction for training and testing data by $\boldsymbol{Y} = \boldsymbol{P}^{T}\boldsymbol{X}$ and $\boldsymbol{Y}^{*} = \boldsymbol{P}^{T}\boldsymbol{X}^{*}$;
8. SVM-based classification in the resulting low-dimensional feature space and return classification accuracy Acc;

Return: $\boldsymbol{s}$

where $\alpha \in [0, 1]$ is a tradeoff parameter that balances the contributions of spatial–spectral CR model and sample augmented neighborhood-constrained model. Then, the optimization problem of (21) can be transformed as

$$\begin{cases} \min\limits_{\boldsymbol{P}} \boldsymbol{P}^{T}\left[\alpha \boldsymbol{S}_{L} + (1-\alpha)\,\boldsymbol{S}\right]\boldsymbol{P} \\ s.t.\ \boldsymbol{P}^{T}\left[\alpha \boldsymbol{S}_{T} + (1-\alpha)\,\boldsymbol{H}\right]P = B \end{cases} \tag{22}$$

where **B** is a constant matrix.

The aforementioned problem can be solved by the Lagrange multiplier method, and the solution can be obtained through the following generalized eigenvalue problem:

$$\left[\alpha \boldsymbol{S}_{L} + (1-\alpha)\,\boldsymbol{S}\right]\boldsymbol{p}_{i} = \lambda_{i}\left[\alpha \boldsymbol{S}_{T} + (1-\alpha)\,\boldsymbol{H}\right]\boldsymbol{p}_{i} \tag{23}$$

in which λ_i and $\boldsymbol{p}_i$ are the ith eigenvalue and eigenvector, respectively. With the eigenvectors $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_d$ corresponding to the first d eigenvalues, the optimal projection matrix can be represented as $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_d] \in \boldsymbol{R}^{N\times d}$. Then, the low-dimensional feature $\boldsymbol{Y}$ can be represented as $\boldsymbol{Y} = \boldsymbol{P}^{T}\boldsymbol{X}$. Algorithm 1 summarizes the pseudocode.

IV. Experimental Results

In this section, eight sets of comparative experiments were set up to validate MFS-PE on three mainstream HSI datasets (i.e., Houston, Dafeng Natural Reserve, and WHU-Hi-LongKou).

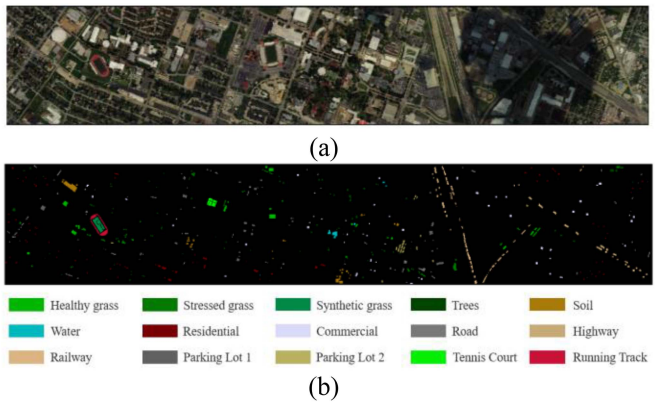

Fig. 5. Houston dataset. (a) False color image. (b) Ground truth.

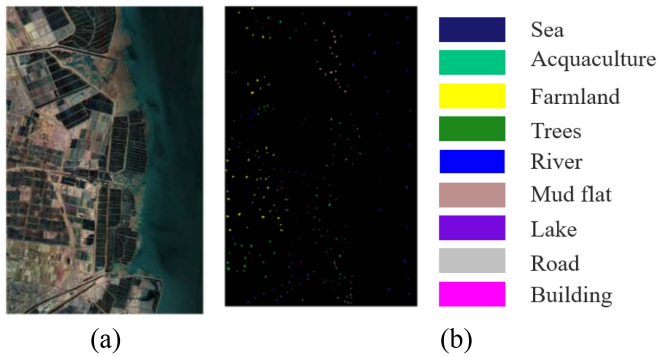

Fig. 6. Dafeng Natural Reserve dataset. (a) False color image. (b) Ground truth.

A. Dataset Description

1) Houston Dataset: This dataset is acquired by the Compact Airborne Spectrographic Imager over the Houston university campus and surroundings in June 2012. The Houston data are introduced in GRSS Data Fusion Contest 2013 [66], and it consists of 3449 × 1905 pixels with 144 bands. The amount of surface material available of this dataset is 15, and has a spatial resolution of 2.5 m. The scene in false color and corresponding ground truth are shown in Fig. 5.

2) Dafeng Natural Reserve: This dataset is located in Yancheng, a coastal city in Jiangsu Province, China [67]. It is a wetland dataset geographically located between latitudes 32° 56′–33° 36′ N and longitudes 120° 42′–120° 51′ E. The Dafeng Natural Reserve dataset is acquired by GF-5, it has a pixel size of 986 × 632 and contains a total of 9 types of features. After removing the 75 bad bands, the remaining 256 bands are used for subsequent experiments. The scene in false color and corresponding ground truth are shown in Fig. 6.

3) WHU-Hi-LongKou: The WHU-Hi-LongKou dataset is captured by the Headwall Nano-Hyperspec imaging sensor in LongKou Town, Hubei Province, China, in July 2018 [68]. The dataset has a spatial resolution of 0.463 m and contains 270 bands in the wavelength range of 0.4 to 1 μm. This dataset is an agricultural scene with a total of 9 surface materials. The scene in false color and corresponding ground truth are shown in Fig. 7.

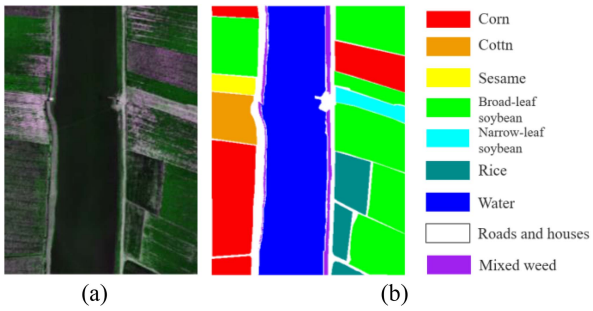

Fig. 7. WHU-HI-LongKou dataset. (a) False color image. (b) Ground truth.

B. Experimental Setup

In our experiments, randomly divide the HSI data into a training set and a testing set. For the Dafeng Natural Reserve and WHU-Hi-LongKou datasets, 30 samples per class are randomly selected as training samples, and for the Houston dataset, 20 samples per class are selected as training samples. The training set is used to learn the DR model and the testing set is chosen to validate of the model. The low-dimensional testing samples are then classified using a support vector machine, and finally the effectiveness of the algorithm is assessed by five metrics: CA per category, overall classification accuracy (OA), average classification accuracy (AA), Kappa coefficient (KC), and computation time (in seconds).

In the experiments, the proposed MFS-PE is compared with some of the most advanced DR algorithms, including NPE LRCRP, JSaCCPGT, S^3-PCA, PCA [69], collaboration preserving GE (CPGE) [39], block collaborative graph-based discriminant analysis (BCGDA) [70], and spatial–spectral manifold reconstruction preserving embedding (SSMRPE) [48]. For PCA, NPE, CPGE, BCGDA, and LRCRP, these five algorithms are all spectral-based DR methods, where BCGDA is enabled to use the labeled information of the samples to enhance manifold discrimination, and LRCRP combines minimizing local scattering with maximizing total divisibility to obtain the optimal projection matrix. For JSaCCPGT, S^3-PCA, and SSMRPE algorithms, spatial features are fully taken into account. Of these, JSaCCPGT applies the data removed from noise by mean filters to a CR model with multiple feature constraint terms to obtain an optimal low-dimensional manifold space, S^3-PCA performs spatial–spectral reconstruction of samples from each homogeneous region after superpixel segmentation and performs both global and local PCA to extract multiple feature information, and SSMRPE performs weighted mean filters on HSI, which designs a combined distance to capture spatial and spectral information and adjust the reconstruction weights. To make the results more reliable, the number of KNN for NPE, S^3-PCA, and MFS-PE is set to 7, and the window size is set to 7 for JSaCCPGT, SSMRPE, and MFS-PE.

C. Classification Results

To compare the classification results of various DR algorithms on three datasets, the training set is randomly selected from 30 samples per class on the Dafeng Natural Reserve dataset and WHU-Hi-LongKou dataset, and 20 samples per class on the Houston dataset, and the testing set consists of the remaining samples. Tables I–III represent the classification results for each of the three datasets, and Figs. 8–10 represent the relevant classification maps for the entire scenes of the three datasets.

In Table I, the combined spatial–spectral methods—JSaCCPGT, S^3-PCA, MMSRPE, and MFS-PE—have better classification results on the Houston dataset than algorithms that only consider spectral information, such as NPE, CPGE, and BCGDA. Algorithms that consider both local and global features, i.e., S^3-PCA and MFS-PE, are able to distinguish each feature class better than other DR methods. The proposed MFS-PE method has better classification results for most classes and achieves the best AA, OA, and KC. This indicates that MFS-PE constructs a robust graph that can efficiently extract spatial–spectral features and local–global features to reveal the internal manifold structure of the HSI. The results of the accuracy evaluation in Table I are accordingly reflected in Fig. 8.

Table II shows the classification results for the Dafeng Natural Reserve dataset. In most classes, MFS-PE provides the best CA and achieves the best AA, OA, and KC. As shown in Fig. 9, compared to other methods, MFS-PE is able to smoothly denoise the HSI data without loss of feature detail, which indicates that combining the two embedding models is not only effective in constructing adjacency graph, but also produces better maps of classification results.

Table III reports the classification results of the nine DR methods on the WHU-Hi-LongKou dataset, with S^3-PCA and MFS-PE being well ahead of the other algorithms in terms of AA, OA, and KC, whereas the proposed algorithms can spend less time cost to obtain higher accuracy. The results illustrate that the combination of multiple feature information applied to DR is extremely effective, which can result in a more robust projection space. In the classification map shown in Fig. 10, MFS-PE is able to effectively remove noise while retaining as much detail as possible in the features.

In conclusion, the proposed MFS-PE algorithm not only fully considers local–global features and spatial–spectral features, but also combines two DR models to solve the upper limit of single model classification performance, which can effectively construct robust graph to fully exploit the intrinsic structure of HSI data and obtain better classification maps.

V. Discussion

In this section, selection of some key parameters is adjusted. First, in the spatial–spectral CRP model, the spectral and spatial regularization parameters λ and β need to be adjusted, then the number of superpixel blocks $\mathcal{K}$ used for local enhancement and sample augmentation is analyzed, as well as the balance parameter α when integrating the two models with different mechanism. Next, the reduced subspace d is analyzed to obtain the optimal dimensionality. Finally, the effectiveness of the proposed local enhancement and KNN selection method are demonstrated. The training set is randomly selected from 30 samples per class on the Dafeng Natural Reserve dataset and WHU-Hi-LongKou dataset, and 20 samples per class on the

TABLE I
CA FOR THE HOUSTON DATASET

Class	PCA	NPE	CPGE	BCGDA	LRCRP	JSaCCGPT	S^3-PCA	SSMRPE	MSF-PE
1	88.06	83.76	**92.43**	88.86	89.81	81.34	87.12	88.37	88.38
2	93.76	**98.99**	91.44	96.7	89.63	90.83	84.04	95.22	88.75
3	98.44	99.71	**100**	94.96	99.57	99.57	99.29	89.13	**100**
4	90.06	**98.54**	96.58	98.14	96.86	93.18	86.64	91.25	85.88
5	92.9	95.89	90.94	86.25	91.26	**97.26**	94.68	80.84	88.46
6	85.16	99.65	86.61	**100**	87.83	97.01	92.71	97.96	87.39
7	86.74	75.34	73.58	80.88	76.21	79.64	85.3	82.3	**90.2**
8	87.94	48.66	76.04	72.9	81.39	**93.51**	81.7	69.17	84.74
9	71.78	53.09	62.47	63	74.94	71.17	84.24	**85.1**	74.86
10	77.44	49.55	54.74	64.37	68.56	83.85	90.11	77.52	**90.32**
11	79.76	83.33	68.95	85.44	81.01	62.33	80.05	79.09	**86.81**
12	73.71	56.09	69.51	52.7	69.75	**81.92**	81.56	74.66	74.87
13	61.67	41.87	45.05	34.92	42.31	71.51	**71.9**	60.46	67.29
14	83.07	89.38	91.76	70.95	82.97	89.35	**100**	87.8	93.6
15	97.44	97.9	96.31	95.03	97.89	99.84	**100**	98.57	97.5
AA	84.52	76.12	79.63	78.6	81.34	82.33	87.18	83.67	**87.23**
OA	84.35	75.15	77.3	77.32	80.89	80.19	85.54	82.88	**85.94**
KC	0.8309	0.7314	0.7546	0.7548	0.7935	0.786	0.8438	0.815	**0.848**
Time(s)	1.5997	3.2915	3.9939	3.3015	2.7209	5.5751	630.5241	5.5057	35.3429

The bold entity represents CA, OA, AA and KC which have the highest accuracy among all the algorithms.

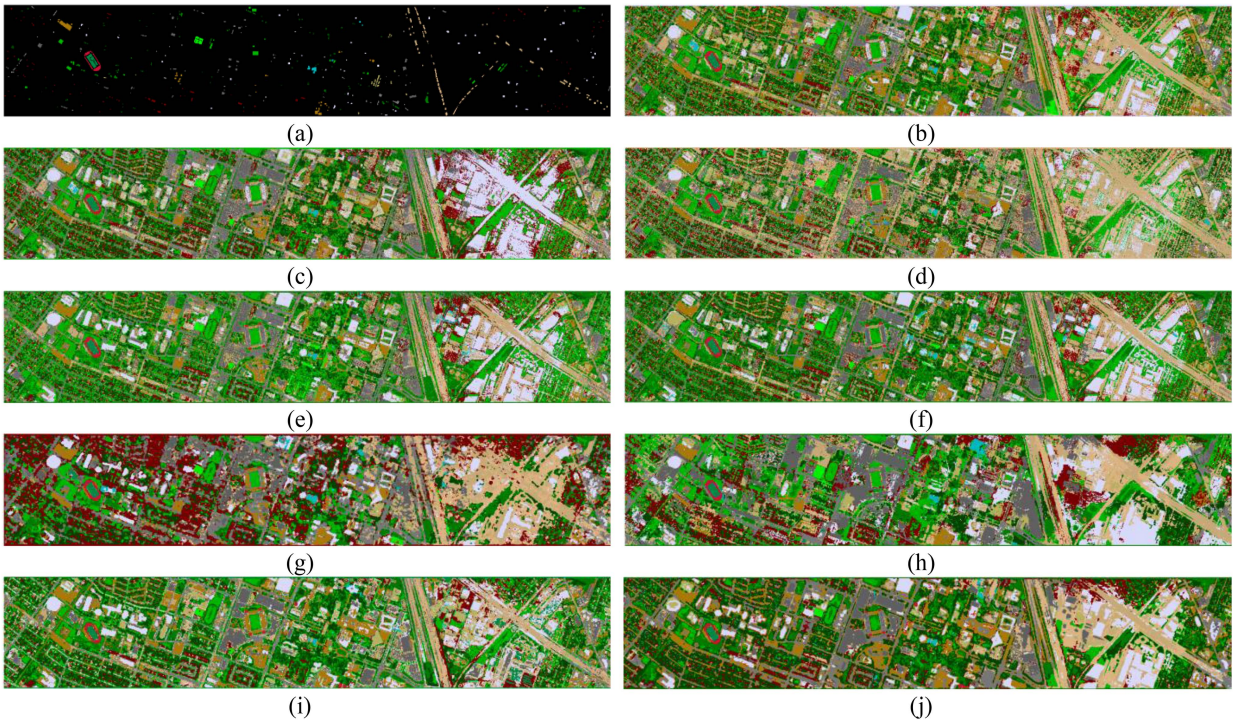

Fig. 8. Classification results of all samples from the Houston data. (a) False color image. (b) PCA. (c) NPE. (d) CPGE. (e) BCGDA. (f) LRCRP. (g) JSaCCPGT. (h) S^3-PCA. (i) SSMRPE. (j) MFS-PE.

Houston dataset, and the testing set consists of the remaining samples.

A. Analysis of Regularization Parameters λ and β

The classification performance of the proposed MFS-PE is seriously affected by the different regularization parameters λ and β, to evaluate algorithm performance, the parameters λ and β are tuned with a set of $\{0.0001, 0.001, \ldots, 1\}$. Fig. 11 shows the CA with different regularization parameters λ and β.

From Fig. 11, we can see that the parameters λ and β have a significant impact on the classification performance of MFS-PE, indicating that both Tikhonov regularization with local enhancement and regular terms with spatial characteristics help the spatial–spectral CR model to obtain more robust collaboration coefficients to construct the graph. Based on the experimental classification results, we set λ and β to 0.001 and 0.01 for the Houston dataset, 0.01 and 0. 01 for the Dafeng Natural Reserve dataset, and 0.001 and 0.01 for the WHU-Hi-LongKou dataset to obtain the best accuracy.

TABLE II
CA FOR THE DAFENG NATURAL RESERVE DATASET

Class	PCA	NPE	CPGE	BCGDA	LRCRP	JSaCCGPT	S^3-PCA	SSMRPE	MSF-PE
1	**100**	99.9	**100**	**100**	**100**	**100**	99.9	**100**	**100**
2	91.68	80.52	92.79	92.05	95.22	97.04	97.78	98.6	**99.25**
3	52.61	36.19	58.26	67.14	46.82	50.18	**98.03**	66.2	71.43
4	**100**	**100**	95.56	**100**	**100**	94.57	**100**	**100**	**100**
5	92.93	91.38	98.15	98.55	97.77	97.32	99.3	**100**	99.88
6	86.51	56.55	72.84	82.11	79.09	84.98	94.87	**96.75**	95.09
7	89.34	**95.38**	87.4	93.84	93.39	88.73	94.64	91.92	94.06
8	97.82	96.59	92.59	97.78	99.54	98.74	95.31	99.32	**100**
9	83.52	59.46	92.86	76.29	76	77.45	95.18	95.18	**100**
AA	89.31	78.19	90.25	89.28	87.94	83.81	93.81	94.53	**95.43**
OA	91.87	84.31	93.67	93.48	92.72	89.14	95.75	97.02	**97.05**
KC	0.9035	0.8143	0.9247	0.9223	0.9136	0.8715	0.9439	0.9645	**0.9648**
Time(s)	2.249	12.9938	12.2355	15.8129	16.3695	21.1139	363.8874	41.8773	105.2567

The bold entity represents CA, OA, AA and KC which have the highest accuracy among all the algorithms.

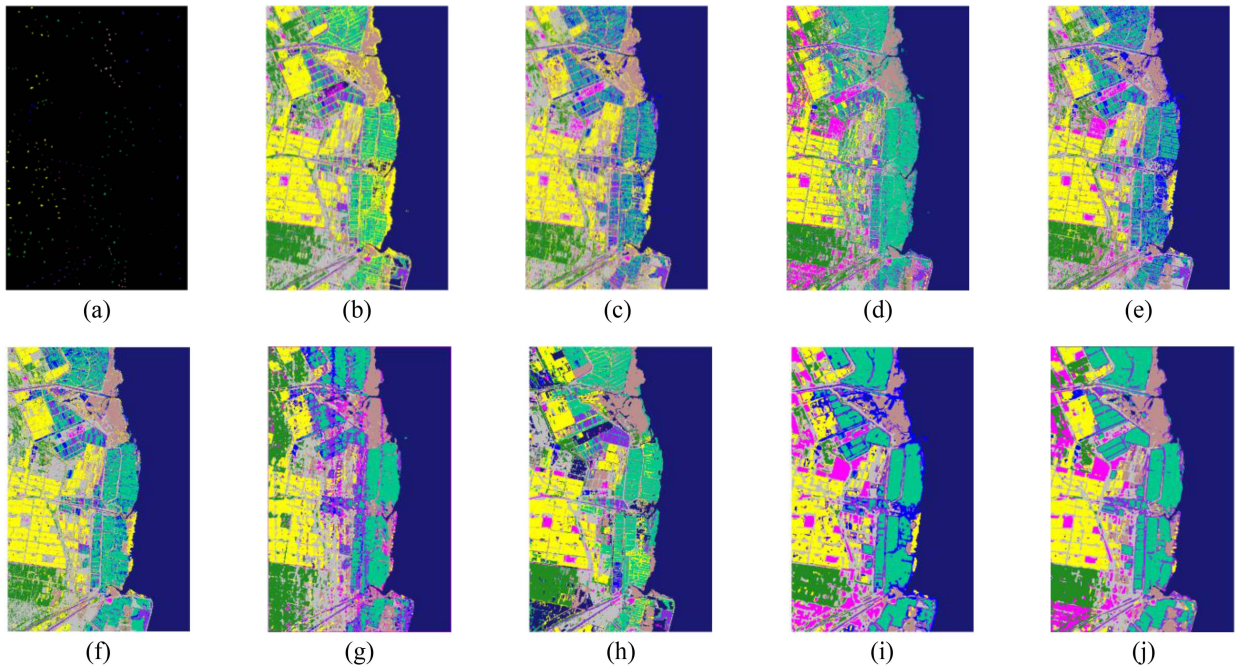

Fig. 9. Classification results of all samples from the Dafeng Natural Reserve data. (a) False color image. (b) PCA. (c) NPE. (d) CPGE. (e) BCGDA. (f) LRCRP. (g) JSaCCPGT. (h) S^3-PCA. (i) SSMRPE. (j) MFS-PE.

TABLE III
CA FOR THE WHU-HI-LONGKOU DATASET

Class	PCA	NPE	CPGE	BCGDA	LRCRP	JSaCCGPT	S^3-PCA	SSMRPE	MSF-PE
1	91.2	84.71	85.51	89.57	95.53	**98.74**	95.96	88.84	96.42
2	37.83	37.74	64.2	41.44	58.76	60.84	85.22	85.65	**87.19**
3	56.12	15.93	31.54	14.97	32.94	52.59	**99.24**	88.42	98.75
4	98.05	93.67	98.29	95.41	97.3	99.22	98.98	98.34	**99.48**
5	14.79	24.12	9.51	36.78	50.31	36.3	**95.59**	54.97	66.72
6	74.02	68.6	88.68	95.1	98.85	94.73	95.4	77.52	**99.1**
7	99.99	98.68	99.97	99.98	99.94	99.77	99.56	**100**	99.87
8	56.97	63.7	70.48	75.68	75.42	84.84	89.91	91.2	**96.92**
9	46.71	15.16	49.96	42.65	51.86	31.6	83.45	56.41	**95.21**
AA	63.39	54.55	64.02	64.99	73.99	72.49	**94.31**	81.21	92.21
OA	75.48	71.39	71.35	82.01	89.03	87.73	**97.26**	92.19	96.89
KC	0.6944	0.6495	0.6438	0.7732	0.8594	0.8428	**0.9642**	0.8984	0.9606
Time(s)	1.0245	1.8429	2.2802	1.8912	2.57	4.2744	135.5693	3.3959	9.0516

The bold entity represents CA, OA, AA and KC which have the highest accuracy among all the algorithms.

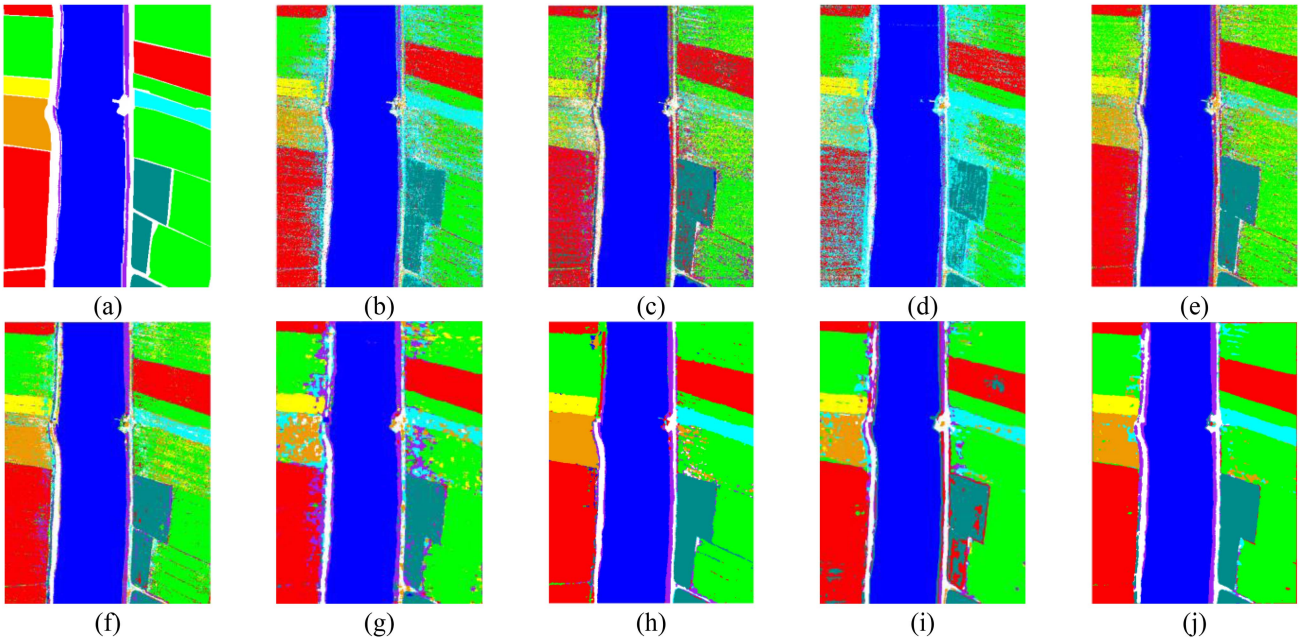

Fig. 10. Classification results of all samples from the WHU-Hi-LongKou data. (a) False color image. (b) PCA. (c) NPE. (d) CPGE. (e) BCGDA. (f) LRCRP. (g) JSaCCPGT. (h) S^3-PCA. (i) SSMRPE. (j) MFS-PE.

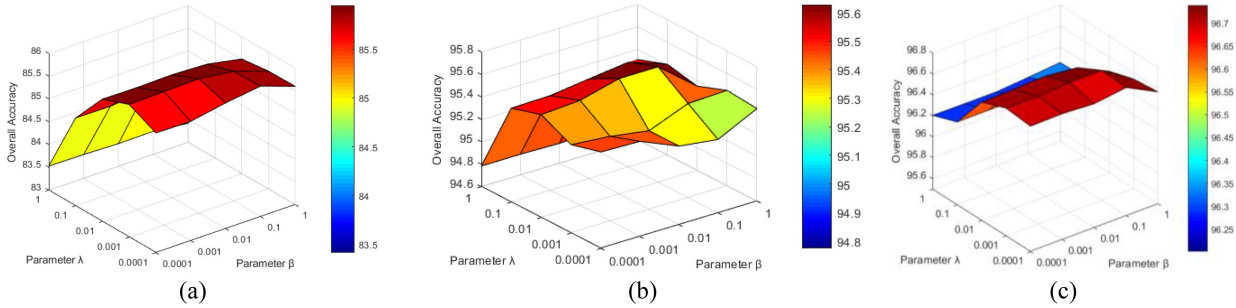

Fig. 11. Classification results with different regularization parameters λ and β. (a) Houston. (b) Dafeng Natural Reserve. (c) WHU-Hi-LongKou.

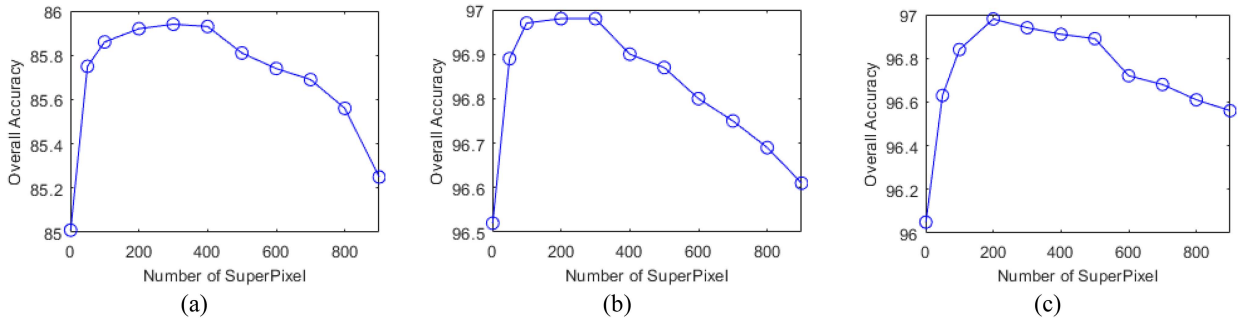

Fig. 12. Classification results of MFS-PE with different number of superpixel blocks. (a) Houston. (b) Dafeng Natural Reserve. (c) WHU-Hi-LongKou.

B. Analysis of Number of Superpixel Segmentation Blocks $\mathcal{K}$

To analyze the effect of different numbers of superpixels on classification performance, the number of segmentation blocks $\mathcal{K}$ needs to be analyzed. Evaluate three datasets by setting the number of superpixel blocks $\{1, 50, 100, 200, 300, \ldots, 900\}$. Experiments were conducted with trade-off parameter α set to 0.5 to ensure that the graphs constructed by both models had the same weights. Fig. 12 shows the CA of different superpixel blocks.

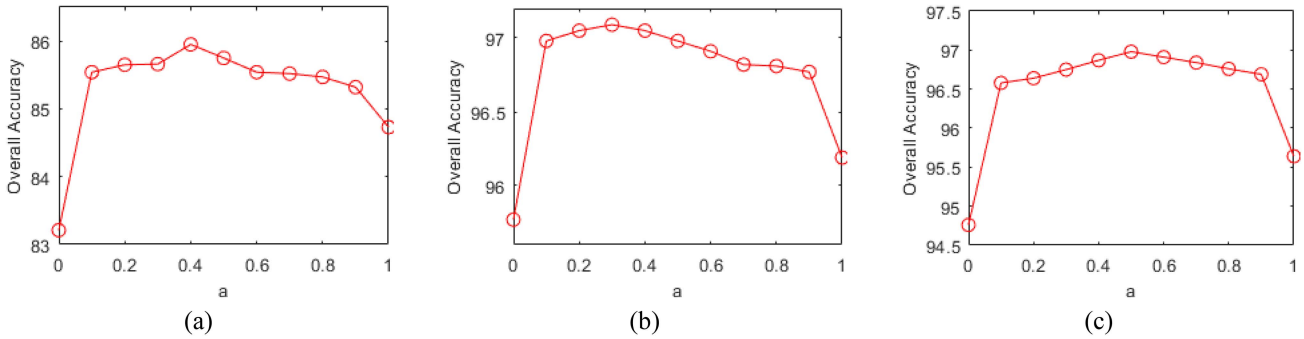

Fig. 13. Classification results of MFS-PE with different tradeoff parameter α. (a) Houston. (b) Dafeng Natural Reserve. (c) WHU-Hi-LongKou.

As shown in Fig. 12, CA is influenced by the number of superpixel segmentation blocks. When the number of superpixels is 1, the HSI is considered to be not segmented, so both sample enhancement and local enhancement do not operate. As the number of superpixel blocks increases, OA tends to increase and then decrease. The reason for this is that as the number of blocks increases from 1, sample enhancement and local enhancement work together to increase the accuracy until it reaches its maximum. However, as the number of blocks continues to increase, the effect of local enhancement begins to diminish, when fewer training samples have the same superpixel labels as the reconstructed samples, resulting in the spatial–spectral distance of most training samples from the reconstructed samples being set to a maximum, which directly affects the construction of spatial–spectral CR graph. Since the amplified samples are not labeled, the improvement in accuracy from sample enhancement does not improve significantly even if the number of superpixel blocks continues to increase. Thus, the accuracy tends to decrease mainly due to local enhancements. In subsequent experiments, to balance local enhancement with sample unlabeled sample augmentation for optimal classification, the number of superpixel blocks is set to 300 for the Houston dataset, 300 for the Dafeng Natural Reserve dataset, and 200 for the WHU-Hi-LongKou dataset.

C. Analysis of Tradeoff Parameter α

In the experiment, the tradeoff parameter α is able to balance the contribution between the spatial–spectral CR graph and the sample augmented neighborhood-constrained graph. Thus, α largely influences the classification performance of MFS-PE. To obtain the best classification results, tradeoff parameter α is chosen within a set of $\{0, 0.1, 0.2, \ldots, 0.9, 1\}$. When α is 0 or 1, it indicates the use of single neighborhood constrained model or spatial–spectral CR model, respectively. Fig. 13 shows the effect of different parameters α on the CA of MFS-PE.

According to Fig. 13, the accuracy tends to rise and then fall as α increases, indicating that the combination of the two graphs contribute to the CA. To achieve the best classification performance, α is set to 0.4 for the Houston dataset, 0.3 for the Dafeng Natural Reserve dataset, and 0.5 for the WHU-Hi-LongKou dataset in the following experiments.

D. Investigation of Embedding Dimension d

For the aforementioned DR algorithm, the embedding dimension d represents the number of effective features to be retained, and therefore the size of d determines how well the classification performs. In Fig. 14, the effect of different embedding dimensions d on the CA is analyzed.

Some conclusions can be drawn from the three datasets that follow Fig. 14. As d increases from 1 to 40, the overall classification accuracy of all DR methods tends to increase first and then smooth out. This indicates that more embedding features can retain richer discriminative information, resulting in more robustly constructed graphs and better classification results. However, due to the limited number of training samples, the manifold information between samples is limited, so an increase in d causes OA to peak and then remain stable. Considering all DR methods for the three datasets, the OA reaches a maximum for embedding dimensions d greater than 20. All methods are compared down to 30 dimensions by CA.

E. Effectiveness Analysis of New KNN Method

In order to demonstrate the effectiveness of the proposed KNN selection method, the superpixel-segmentation-based KKN method is compared with the traditional spectral-domain-based KNN method, and Fig. 15 shows the OAs of the two methods.

From the figure, it can be seen that the proposed KNN method has different degrees of improvement compared to the traditional method on all three datasets with equal K values. This suggests that there is a relationship between the spectral similarity of different blocks of superpixels and the spatial location of the samples, and that a clever combination of them can lead to the selection of more useful nearest neighbors.

F. Effectiveness Analysis of Local Enhancement

To demonstrate the effectiveness of the local enhancement method, the experiments were validated under conditions with and without local enhancement, and Fig. 16 shows the OAs in both cases.

As shown in Fig. 16, with the use of local enhancement, the CA of the three datasets is improved by about 1% over that without local enhancement. This indicates that not all the

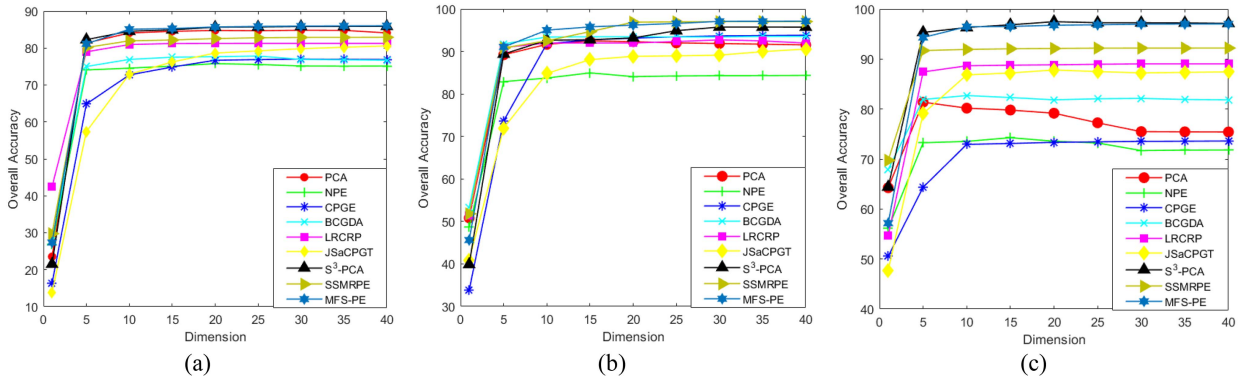

Fig. 14. Classification results with different embedding dimensions *d*. (a) Houston. (b) Dafeng Natural Reserve. (c) WHU-Hi-LongKou.

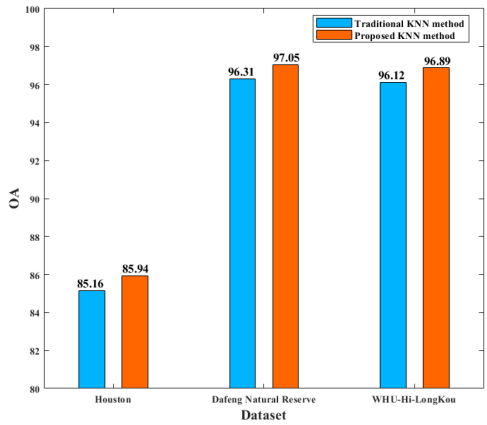

Fig. 15. Effectiveness analysis of the proposed KNN method.

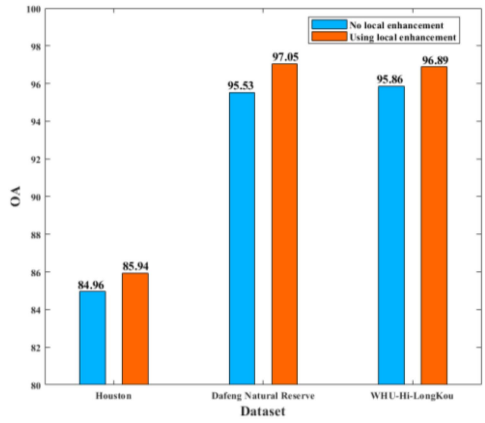

Fig. 16. Effectiveness analysis of local enhancement.

remaining training samples are useful when constructing the spatial–spectral CR graph for each training sample, and the weights of certain training samples, though small, still have bad effects. Therefore, more robust graph can be constructed efficiently by forcing the weights of these samples that play side effects to a minimum.

VI. Conclusion

This article proposes an unsupervised DR method that can fully take into account multiple feature information. Specifically, two GE models are combined to explore the collaborative relationships and local features between the HSI data to solve the performance upper limit of the single model. Then, sample augmentation and local enhancement are achieved through superpixel segmentation to improve the stability of the new model. Finally, a new KNN selection method is designed to select more suitable neighbors. However, the proposed MFS-PE method is performed without using the label information of samples. In practical applications, the class label information of the samples usually plays a large role. Therefore, the algorithm will be extended to semisupervised or supervised versions with available label information in the future research.

References

[1] A. Plaza et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. Suppl. 1, pp. S110–S122, 2009, doi: 10.1016/j.rse.2007.07.028.

[2] B. Rasti et al., "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020, doi: 10.1109/MGRS.2020.2979764.

[3] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013, doi: 10.1109/JPROC.2012.2197589.

[4] P. Zheng, H. Su, H. Lu, and Q. Du, "Adaptive hypergraph regularized multilayer sparse tensor factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5502918, doi: 10.1109/TGRS.2023.3241115.

[5] R. N. Patro, S. Subudhi, P. K. Biswal, and F. Dell'Acqua, "A review of unsupervised band selection techniques: Land cover classification for hyperspectral earth observation data," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 3, pp. 72–111, Sep. 2021, doi: 10.1109/MGRS.2021.3051979.

[6] D. Hong et al., "Interpretable hyperspectral artificial intelligence," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 52–87, Jun. 2021.

[7] D. Manolakis et al., "Longwave infrared hyperspectral imaging: Principles, progress, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 72–100, Jun. 2019, doi: 10.1109/MGRS.2018.2889610.

[8] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, Mar. 2022, doi: 10.1109/MGRS.2021.3105440.

[9] Y. Gao, H. Su, H. Lu, and Q. Du, "Self-balancing dictionary learning for relaxed collaborative representation of hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5539918, doi: 10.1109/TGRS.2022.3211209.

[10] J. Hu, X. Shen, H. Yu, X. Shang, Q. Guo, and B. Zhang, "Extended subspace projection upon sample augmentation based on global spatial and local spectral similarity for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8653–8664, 2021, doi: 10.1109/JSTARS.2021.3107105.

[11] H. Su, H. Zhang, Z. Wu, and Q. Du, "Relaxed collaborative representation with low-rank and sparse matrix decomposition for hyperspectral anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6826–6842, 2022, doi: 10.1109/JSTARS.2022.3193315.

[12] H. Su, Y. Hu, H. Lu, W. Sun, and Q. Du, "Diversity-driven multikernel collaborative representation ensemble for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2861–2876, 2022, doi: 10.1109/JSTARS.2022.3162121.

[13] 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, doi: 10.1109/TGRS.2019.2957135.

[14] Y.-J. Deng, H.-C. Li, X. Song, Y.-J. Sun, X.-R. Zhang, and Q. Du, "Patch tensor-based multigraph embedding framework for dimensionality reduction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1630–1643, Mar. 2020, doi: 10.1109/TGRS.2019.2947200.

[15] H. Fu et al., "A novel band selection and spatial noise reduction method for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5535713, doi: 10.1109/TGRS.2022.3189015.

[16] Y. Zhou, J. Peng, and C. L. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015, doi: 10.1109/TGRS.2014.2333539.

[17] 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, no. 85, pp. 2399–2434, 2006.

[18] X. Xu, Z. Huang, L. Zuo, and H. He, "Manifold-based reinforcement learning via," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 934–947, Apr. 2017.

[19] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014, doi: 10.1109/MSP.2013.2279894.

[20] Y. Jin, Y. Dong, Y. Zhang, and X. Hu, "SSMD: Dimensionality reduction and classification of hyperspectral images based on spatial-spectral manifold distance metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5538916, doi: 10.1109/TGRS.2022.3205178.

[21] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003, doi: 10.1162/089976603321780317.

[22] W. Li, L. Zhang, L. Zhang, and B. Du, "GPU parallel implementation of isometric mapping for hyperspectral classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1532–1536, Sep. 2017, doi: 10.1109/LGRS.2017.2720778.

[23] Y. Hou, P. Zhang, X. Xu, X. Zhang, and W. Li, "Nonlinear dimensionality reduction by locally linear inlaying," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 300–315, Feb. 2009, doi: 10.1109/TNN.2008.2005582.

[24] S. T. Tu, J. Y. Chen, W. Yang, and H. Sun, "Laplacian eigenmaps-based polarimetric dimensionality reduction for SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 170–179, Jan. 2012, doi: 10.1109/TGRS.2011.2168532.

[25] Y. Zhang, Z. Zhang, J. Qin, L. Zhang, B. Li, and F. Li, "Semi-supervised local multi-manifold Isomap by linear embedding for feature extraction," *Pattern Recognit.*, vol. 76, pp. 662–678, 2018, doi: 10.1016/j.patcog.2017.09.043.

[26] Q. Li and H. Ji, "Multimodality image registration using local linear embedding and hybrid entropy," *Neurocomputing*, vol. 111, pp. 34–42, 2013, doi: 10.1016/j.neucom.2012.11.032.

[27] H. Yuan and Y. Y. Tang, "Learning with hypergraph for hyperspectral image feature extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 8, pp. 1695–1699, Aug. 2015, doi: 10.1109/LGRS.2015.2419713.

[28] L. Fang, S. Li, X. 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, doi: 10.1109/TGRS.2015.2392755.

[29] Z. Wang and B. He, "Locality perserving projections algorithm for hyperspectral image dimensionality reduction," in *Proc. 19th Int. Conf. Geoinform.*, 2011, pp. 1–4, doi: 10.1109/GeoInformatics.2011.5980790.

[30] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, 2005, vol. 2, pp. 1208–1213, 10.1109/ICCV.2005.167.

[31] S. Yan, D. Xu, B. Zhang, H.-J. 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, doi: 10.1109/TPAMI.2007.250598.

[32] Q. Liu, Z. Wu, Y. Xu, and Z. Wei, "A unified attention paradigm for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5506316, doi: 10.1109/TGRS.2023.3257321.

[33] Z. Wu et al., "Hyperspectral anomaly detection with relaxed collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5533417, doi: 10.1109/TGRS.2022.3190327.

[34] 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, doi: 10.1109/TGRS.2011.2129595.

[35] H. Su, F. Shao, Y. Gao, H. Zhang, W. Sun, and Q. Du, "Probabilistic collaborative representation based ensemble learning for classification of wetland hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509517, doi: 10.1109/TGRS.2023.3267638.

[36] N. H. Ly, Q. Du, and J. E. Fowler, "Sparse graph-based discriminant analysis for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 3872–3884, Jul. 2014, doi: 10.1109/TGRS.2013.2277251.

[37] F. Cao, Z. Yang, X. Hong, Y. Cheng, Y. Huang, and J. Lv, "Supervised dimensionality reduction of hyperspectral imagery via local and global sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3860–3874, 2021, doi: 10.1109/JSTARS.2021.3069030.

[38] F. Luo, H. Huang, Z. Ma, and J. Liu, "Semisupervised sparse manifold discriminative analysis for feature extraction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6197–6211, Oct. 2016, doi: 10.1109/TGRS.2016.2583219.

[39] F. Luo, Z. Zou, J. Liu, and Z. Lin, "Dimensionality reduction and classification of hyperspectral image via multistructure unified discriminative embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517916, doi: 10.1109/TGRS.2021.3128764.

[40] Y. Zhong, X. Wang, L. Zhao, R. Feng, L. Zhang, and Y. Xu, "Blind spectral unmixing based on sparse component analysis for hyperspectral remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 119, pp. 49–63, 2016, doi: 10.1016/j.isprsjprs.2016.04.008.

[41] F. Luo, H. Huang, J. Liu, and Z. Ma, "Fusion of graph embedding and sparse representation for feature extraction and classification of hyperspectral imagery," *Photogrammetry Eng. Remote Sens.*, vol. 83, no. 1, pp. 37–46, 2017, doi: 10.14358/PERS.83.1.37.

[42] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 471–478, doi: 10.1109/ICCV.2011.6126277.

[43] N. H. Ly, Q. Du, and J. E. Fowler, "Collaborative graph-based discriminant analysis for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2688–2696, Jun. 2014, doi: 10.1109/JSTARS.2014.2315786.

[44] N. Liu, W. Li, and Q. Du, "Unsupervised feature extraction for hyperspectral imagery using collaboration-competition graph," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1491–1503, Dec. 2018, doi: 10.1109/JSTSP.2018.2877474.

[45] W. Yang, Z. Wang, and C. Sun, "A collaborative representation based projections method for feature extraction," *Pattern Recognit.*, vol. 48, no. 1, pp. 20–27, 2015, doi: 10.1016/j.patcog.2014.07.009.

[46] X. Jiang, L. Xiong, Q. Yan, Y. Zhang, X. Liu, and Z. Cai, "Unsupervised dimensionality reduction for hyperspectral imagery via Laplacian regularized collaborative representation projection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6007805.

[47] M. Han, R. Cong, X. Li, H. Fu, and J. Lei, "Joint spatial-spectral hyperspectral image classification based on convolutional neural network," *Pattern Recognit. Lett.*, vol. 130, pp. 38–45, 2020, doi: 10.1016/j.patrec.2018.10.003.

[48] Y. Jin, Y. Dong, Y. Zhang, and X. Hu, "SSMD: Dimensionality reduction and classification of hyperspectral images based on spatial–spectral manifold distance metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5538916, doi: 10.1109/TGRS.2022.3205178.

[49] R. Mohanty, S. L. Happy, and A. Routray, "A semisupervised spatial spectral regularized manifold local scaling cut with HGF for dimensionality reduction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3423–3435, Jun. 2019, doi: 10.1109/TGRS.2018.2884771.

[50] J. Jiang, C. Chen, Y. Yu, X. Jiang, and J. Ma, "Spatial-aware collaborative representation for hyperspectral remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 404–408, Mar. 2017, doi: 10.1109/LGRS.2016.2645708.

[51] C. Shah and Q. Du, "Spatial-aware collaboration-competition preserving graph embedding for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5506005, doi: 10.1109/LGRS.2021.3074328.
[52] X. Zhang, X. Jiang, J. Jiang, Y. Zhang, X. Liu, and Z. Cai, "Spectral-spatial and superpixelwise PCA for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5502210, doi: 10.1109/TGRS.2021.3057701.
[53] H. Huang, G. Shi, H. He, Y. Duan, and F. Luo, "Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2604–2616, Jun. 2020, doi: 10.1109/TCYB.2019.2905793.
[54] W. Shang et al., "Hyperspectral anomaly detection via sparsity of core tensor under gradient domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5517816, doi: 10.1109/TGRS.2023.3297627.
[55] X. Zhang, Y. He, N. Zhou, and Y. Zheng, "Semisupervised dimensionality reduction of hyperspectral images via local scaling cut criterion," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1547–1551, Nov. 2013, doi: 10.1109/LGRS.2013.2261797.
[56] W. K. Wong and H. T. Zhao, "Supervised optimal locality preserving projection," *Pattern Recognit.*, vol. 45, no. 1, pp. 186–197, 2012, doi: 10.1016/j.patcog.2011.05.014.
[57] P. Zhu, W. Zuo, L. Zhang, S. C.-K. Shiu, and D. Zhang, "Image set-based collaborative representation for face recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 7, pp. 1120–1132, Jul. 2014, doi: 10.1109/TIFS.2014.2324277.
[58] F. Yang, H. Lu, and M.-H. Yang, "Robust superpixel tracking," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1639–1651, Apr. 2014, doi: 10.1109/TIP.2014.2300823.
[59] D. R. Thompson, L. Mandrake, M. S. Gilmore, and R. Castaño, "Superpixel endmember detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4023–4033, Nov. 2010, doi: 10.1109/TGRS.2010.2070802.
[60] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2097–2104, doi: 10.1109/CVPR.2011.5995323.
[61] 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, doi: 10.1109/TPAMI.2012.120.
[62] W. Li and Q. Du, "Joint within-class collaborative representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2200–2208, Jun. 2014, doi: 10.1109/JSTARS.2014.2306956.
[63] Y. Luo, J. Han, Z. Liu, M. Wang, and G.-S. Xia, "An elliptic centerness for object instance segmentation in aerial images," *J. Remote Sens.*, vol. 2022, 2022, Art. no. 9809505, doi: 10.34133/2022/9809505.
[64] A. M. Abdi, M. Brandt, C. Abel, and R. Fensholt, "Satellite remote sensing of savannas: Current status and emerging opportunities," *J. Remote Sens.*, vol. 2022, 2022, Art. no. 9835284, doi: 10.34133/2022/9835284.
[65] J. Liu, Z. Wu, L. Xiao, and X.-J. Wu, "Model inspired autoencoder for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5522412, doi: 10.1109/TGRS.2022.3143156.
[66] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral feature extraction using total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6976–6985, Dec. 2016, doi: 10.1109/TGRS.2016.2593463.
[67] H. Su, W. Yao, Z. Wu, P. Zheng, and Q. Du, "Kernel low-rank representation with elastic net for China coastal wetland land cover classification using GF-5 hyperspectral imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 171, pp. 238–252, 2021, doi: 10.1016/j.isprsjprs.2020.11.018.
[68] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, "WHU-Hi: UAV-borne hyperspdectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," *Remote Sens. Environ.*, vol. 250, 2020, Art. no. 112012, doi: 10.1016/j.rse.2020.112012.
[69] I. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer, 2005.
[70] N. H. Ly, Q. Du, and J. E. Fowler, "Collaborative graph-based discriminant analysis for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2688–2696, Jun. 2014, doi: 10.1109/JSTARS.2014.2315786.

Kai Chen received the B.E. degree in surveying and mapping engineering from the Shandong University of Science and Technology, Taian, China, in 2021. He is currently working toward the M.E. degree in surveying and mapping with the School of Earth Sciences and Engineering, Hohai University, Nanjing, China.

His research interests include hyperspectral remote sensing image processing, collaborative representation, and hyperspectral machine learning dimensionality reduction.

Guoguo Yang received the M.S. degree in ecology from Henan Agricultural University, Zhengzhou, China, in 2001, and the Ph.D. degree in ecology from the College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing, China, in 2022.

Since 2023, she has been a Research Assistant with the School of Geographic Information and Tourism, Chuzhou University, Chuzhou, China. Her research interests include hyperspectral remote sensing and quantitative inversion of remote sensing of soil parameters.

Jing Wang received the Ph.D. degree in cartography and geography information system from the Key Laboratory of Virtual Geographic Environment (Ministry of Education), Nanjing Normal University, Nanjing, China, in 2011.

He is currently an Associate Professor with the School of Geographic Information and Tourism, Chuzhou University, Chuzhou, China. His main research interests include photogrammetry and remote sensing applications.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore, MD, USA, in 2000.

She is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a Fellow of the International Society for Optics and Photonics. She was a Co-Chair of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013, and the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She was an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, Journal of Applied Remote Sensing, IEEE Signal Processing Letters, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, PATTERN RECOGNITION, and MDPI *Remote Sensing*. From 2016 to 2020, she was the Editor-in-Chief for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She is a member of the IEEE Periodicals Review and Advisory Committee and SPIE Publications Committee.

Hongjun Su (Senior Member, IEEE) received the Ph.D. degree in cartography and geography information system from the Key Laboratory of Virtual Geographic Environment (Ministry of Education), Nanjing Normal University, Nanjing, China, in 2011.

He is currently a Full Professor with the School of Earth Sciences and Engineering, Hohai University, Nanjing, China. His main research interests include hyperspectral remote sensing dimensionality reduction, classification, and spectral unmixing.

Dr. Su is an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He was the recipient of the 2016 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society.