Spectral-spatial and Superpixelwise Unsupervised Linear Discriminant Analysis for Feature Extraction and Classification of Hyperspectral Images

Pengyu Lu, Xinwei Jiang, Yongshan Zhang, Xiaobo Liu, Zhihua Cai, Junjun Jiang and Antonio Plaza, *Fellow, IEEE*

Abstract—Dimensionality reduction (DR) is important for feature extraction and classification of hyperspectral images (HSIs). Recently proposed superpixel-based DR models have shown promising performance, where superpixel segmentation techniques were applied to segment an HSI and then DR models like principal component analysis (PCA) or linear discriminant analysis (LDA) were employed to extract the local and/or global features. However, superpixelwise PCA based local features are unsatisfactory because PCA aims to extract features with high variance, which could be inefficient in superpixels with mixed objects or strong noise/outliers. In addition, superpixelwise unsupervised LDA based global features may neglect local (spatial-contextual) information. To address these issues, we propose a new spectral-spatial and superpixelwise unsupervised LDA (S^3-ULDA) model for unsupervised feature extraction from HSIs. Specifically, the HSI is first segmented into various superpixels with pseudo labels. Then, superpixel based local reconstruction for HSI denoising is conducted. Next, superpixelwise unsupervised LDA (SuperULDA) is performed on both the original HSI and locally reconstructed data to extract global features. Then, superpixelwise unsupervised local Fisher discriminant analysis (SuperULFDA) is developed for local feature extraction, where each superpixel and its adjacent superpixels (along with their pseudo-labels) are fed into local Fisher discriminant analysis (LFDA) to extract local features. The superpixel-level local manifold structures can be effectively modeled by the proposed SuperULFDA. Finally, by fusing the extracted global and local features, novel global-local and spectral-spatial features can be obtained. Our experimental results on several benchmark HSIs demonstrate the superiority of the proposed method over state-of-the-art methods. The code of the proposed model is available at **https://github.com/XinweiJiang/S3-ULDA**.

Index Terms—Hyperspectral images (HSIs), superpixel segmentation, principal component analysis (PCA), linear discriminant analysis (LDA), dimensionality reduction (DR), spectral-spatial features.

P. Lu, X. Jiang, Y. Zhang and Z. Cai are with the School of Computer Science, and also with Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China (e-mail: hello-lpy@qq.com, ysjxw@hotmail.com, yszhang.cug@gmail.com, zh-cai@cug.edu.cn).

X. Liu is with the School of Automation, and also with the Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, China University of Geosciences, Wuhan 430074, China (e-mail:xbliu@cug.edu.cn).

J. Jiang is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail:jiangjunjun@hit.edu.cn).

A. Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politecnica, University of Extremadura, 10003 Caceres, Spain (e-mail: aplaza@unex.es).

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) contain abundant spectral bands that greatly assist classification tasks [1]. However, accurate HSI classification is challenging due to the relatively small number of labeled samples available and the high dimensionality of spectral features, which leads to the problem called 'curse of the dimensionality' or 'Hughes phenomenon' [2]. Many dimensionality reduction (DR) methods have been designed to mitigate this problem by removing noise and extracting lower-dimensional features from HSIs.

Generally speaking, DR models can be divided into two categories: feature selection and feature extraction. The former try to pick a representative subset of features from the original high-dimensional observations [3], while the latter transform the observed data to a low-dimensional subspace [2]. As feature selection is often sub-optimal compared to feature extraction [2], we focus on feature extraction models in this paper. According to the availability of labeled samples, feature extraction algorithms can be roughly divided into unsupervised and supervised models. Principal component analysis (PCA) is a classic unsupervised DR model which aims at finding the projection directions along which the data have the maximum variance. Variants of PCA include orthogonal total variation component analysis (OTVCA) [4] and Tensor PCA [5]. However, facing the complex nonlinear relationships between high-dimensional spectral features and the projected low-dimensional features is challenging for the aforementioned methods. As a result, many nonlinear DR models have been introduced for effective feature extraction from HSIs, including manifold learning techniques such as neighborhood preserving embedding (NPE), locality preserving projection (LPP) [2], local pixel NPE (LPNPE) [6], local neighborhood structure preserving embedding (LNSPE) [7] and laplacian regularized collaborative representation projection (LRCRP) [8]. In contrast, supervised DR models as linear discriminant analysis (LDA) [2] project the high-dimensional spectral features into a low-dimensional subspace where the data have maximum inter-class distance and minimum intra-class separation. To address the limitations of LDA, which does not consider the local manifold structures embedded into the observations, models like nonparametric weighted feature

extraction (NWFE) [2], local Fisher discriminant analysis (LFDA) [9], multistructure unified discriminative embedding (MUDE) [10], superpixel-based spatial-spectral dimension reduction (SSDR) [11] and various sparse or collaborative graph-based discriminant analysis methods have been also introduced [12]. Also, there are some semisupervised DR models which try to simultaneously use the few labeled data and numerous unlabeled data [13].

Although the aforementioned methods have been successful in the task of extracting features from HSIs, the performance of these models could be compromised by their sensitivity to noise and also by the fact that they disregard spatial information [14]–[16]. To address this issue, typical image filtering models such as mean, Gabor, and propagation filters [15], [17] have been adopted to denoise HSIs and extract spatial features with physical meaning.

Also, many deep learning models have been developed to extract hierarchical spectral-spatial features for HSI classification. For example, convolutional neural networks (CNNs) and their 3D version (3DCNNs) have been proposed to simultaneously extract the spatial and spectral information from HSIs with 3D convolutions [18]. Inspired by graph convolutional networks (GCNs), which can naturally model long-range spatial relations in HSIs, the minibatch GCN (miniGCN) was fused with CNNs to extract effective spectral-spatial features and conduct end-to-end classification, leading to the so-called fusion network (FuNet) [19]. To further improve the performance of GCN, the superpixel-based graph-in-graph (GiG) model was introduced into GCN, leading to GiG convolutional network (GiGCN), which could extract effective local and global features for HSI classification [20]. Hypergraph convolution and high-order structure preservation were introduced in structure-preserved hyper GCN (SPHGCN) [21], which could achieve more discriminative features from multiple pixel nodes. To further boost the performance of deep learning models, various deep learning models were fused. For example, attention multihop graph and multi-scale convolutional fusion network (AMGCFN) were proposed to fuse multiscale CNN and multihop GCN to extract the multilevel information of HSIs [22]. A global spatial feature representation subnetwork and a dual-view spectral aggregation subnetwork were fused in dual-view spectral and global spatial feature fusion network (DSGSF) [23] to learn the discriminative spatial-spectral features. Deep high-order tensor convolutional sparse coding (DHTCSCNet) [24] tried to use tensor based deep feature extraction network to extract combined features which incorporated shallow, deep, spectral, and spatial features. Also, self-supervised learning techniques such as masked autoencoder were introduced in masked auto-encoding spectral–spatial transformer (MAEST) [25] and spectral–spatial masked transformer (SS-MTr) as well as its variant contrastive SS-MTr (C-SS-MTr) [26]. However, the high dimensionality of spectral features and the complex spatial structures in HSI data may compromise the performance of the filtering models, notwithstanding the fact that a relatively large amount of labeled data (along with a laborious and time-intensive hyperparameter tuning phase) is needed for training these deep learning models.

In contrast, many superpixel-based shallow spectral-spatial feature extraction models have been developed which exhibit promising performance in terms of accuracy and complexity [27], [28]. For example, entropy rate superpixel (ERS) segmentation [29] was applied for HSI segmentation, followed by PCA-based local feature extraction in each homogeneous superpixel block, which results in the superpixelwise PCA method (SuperPCA) [30]. To further boost the performance of SuperPCA, superpixel-based local reconstruction for HSI denoising and PCA-based global-local and spectral-spatial feature extraction were introduced in the spectral-spatial and superpixelwise PCA method (S^3-PCA) [16]. Motivated by the exploitation of pseudo-labels from superpixel segmentation, the flexible Gabor-based superpixel-level unsupervised LDA (FG_SuULDA) [15] was proposed to conduct unsupervised global feature extraction. To further consider local/nonlocal spatial–spectral correlation information among/between superpixels, the superpixel-level hybrid discriminant analysis (SHDA) [31] was introduced based on pseudo-labels of superpixels and unsupervised LDA. However, the superpixel-level global discriminant analysis based on superpixels could be inaccurate when the superpixels include mixed objects or noise/outliers. Additionally, superpixel-level local graph-based LDA could also be ineffective, mainly because the global and local discriminant analysis models are conducted simultaneously, leading to the extraction of global features that could decrease the effectiveness of the local model. Recently, superpixel-based feature extraction techniques have been reviewed [28] and it is acknowledged that superpixelwise PCA based local features in SuperPCA and S^3-PCA could be not good enough for accurate HSI classification, mainly because the application of the classic PCA to each superpixel is dominated by spectral features with high variance, which could fail to characterize superpixels with mixed objects or strong noise/outliers. In comparison, unsupervised LDA based global features in FG_SuULDA and SHDA could also be unsatisfactory due to the fact that they neglect the local (spatial-contextual) information in HSIs.

To address these issues, in this paper we propose a new unsupervised DR model called spectral-spatial and superpixel-wise unsupervised LDA (S^3-ULDA) which extracts effective global-local and spectral-spatial features through superpixel-based unsupervised LDA and LFDA models. Specifically, we first conduct superpixel segmentation on the HSI, followed by superpixel-based local reconstruction for HSI denoising. Then, we simultaneously make use of the original and locally reconstructed HSI data to extract low-dimensional global features based on unsupervised LDA. Next, unsupervised LFDA with pseudo-labels of adjacent superpixels is performed to conduct unsupervised local feature extraction on a superpixel-by-superpixel basis. Compared to PCA-based local feature extraction in SuperPCA and S^3-PCA, and to unsupervised LDA-based global feature extraction in SHDA, the extracted local features based on the proposed unsupervised LFDA are more informative, mainly because superpixel-level local features can be directly extracted and our new model can effectively fit the local multimodality of data in the adjacent superpixels inherited from classic LFDA. The main contributions of the

proposed models are as follows:

1) A novel unsupervised feature reduction model (SuperULDA) is proposed to extract global features from HSIs based on both the original and locally reconstructed data, as well as the pseudo-labels of superpixels provided by ERS. By simultaneously using the two data sources, the extracted global features are more discriminative.

2) To extract effective local features from HSIs, a novel unsupervised feature reduction model (SuperULFDA) is conducted locally on each superpixel along with its adjacent superpixels as well as the corresponding pseudo-labels of superpixels. By performing the new model on a superpixel-by-superpixel basis, the extracted superpixel-level local features could effectively model the local spatial structures of the HSI and then provide discriminative local features.

3) To further improve the discrimination of the extract features, the global-local and spectral-spatial features are obtained by fusing the obtained global and local features, leading to the proposed S^3-ULDA which significantly improves the quality of the learned feature representation and thus benefits the downstream classification tasks.

The rest of the paper is organized as follows. Section II reviews some related works. Section III describes the proposed method. Section IV provides experimental results and analyses. Section V concludes the paper with some remarks and hints at plausible future research lines.

II. RELATED WORKS

In this section, we briefly some methods that are directly related with our new approach, including entropy rate superpixel (ERS) segmentation, PCA and SuperPCA (unsupervised DR models), LDA and LFDA (supervised DR models).

A. Entropy Rate Superpixel (ERS) Segmentation

Superpixel segmentation aims to segment the image into homogeneous regions (also termed superpixel blocks) that share similar properties. In the field of HSI processing, ERS is a popular superpixel segmentation algorithm. It represents the HSI by a graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, where the set of vertices $\mathbf{V}$ corresponds to the pixels in the HSI and the set of edges $\mathbf{E}$ denotes the pairwise similarities of vertices. ERS tries to pick a subset of edges $\mathbf{A} \subseteq \mathbf{E}$ such that the obtained graph $\mathcal{G}^* = (\mathbf{V}, \mathbf{A})$ consists of K connected sub-graphs. The objective function of ERS is defined by

$$\mathbf{A}^* = \arg\max_{\mathbf{A}} \mathrm{Tr}\left\{H(\mathbf{A}) + \alpha B(\mathbf{A})\right\}, \mathrm{s.t.} \mathbf{A} \subseteq \mathbf{E}, \quad (1)$$

where $H(\mathbf{A})$ is an entropy rate function aimed to obtain homogeneous and compact clusters, $B(\mathbf{A})$ is a balancing term to obtain clusters with similar size, and α is a balancing parameter. $\mathbf{A}$ greedy heuristic algorithm can be applied to solve Eq. (1). ERS has been successfully adopted to identify spatially homogeneous regions in HSIs [15], [16], [30].

B. PCA and SuperPCA

PCA is a classic DR algorithm which tries to obtain low-dimensional features with maximum variance. Due to its simplicity and effectiveness, it has been widely used for HSI data pre-processing and feature extraction. Given a set of high-dimensional observations $\mathbf{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\} \in \mathbb{R}^{B \times N}$ with dimensionality B and a number of samples N, PCA aims to find low-dimensional features $\boldsymbol{H} = \boldsymbol{P}^T \boldsymbol{X}$ with projection matrix $\boldsymbol{P} \in \mathbb{R}^{B \times d}$ by maximizing the total variance of the projection

$$\boldsymbol{P^*} = \arg\max_{\boldsymbol{P}} \mathrm{Tr}(\boldsymbol{P}^\mathrm{T} \mathrm{Cov}(\mathbf{X}) \boldsymbol{P}), s.t. \boldsymbol{P}^T \boldsymbol{P} = \mathbf{I}, \quad (2)$$

where $Cov(\boldsymbol{X}) = \frac{1}{N} \boldsymbol{X} \boldsymbol{X}^T$ is the covariance matrix of the observation. The optimization of PCA can be performed by selecting the eigenvectors corresponding to the top$-d$ eigenvalues to form the projection matrix $\boldsymbol{P}$.

SuperPCA first performs ERS to obtain homogeneous superpixels made up of spatially similar pixels, and then performs PCA on each superpixel to extract local spectral-spatial features. However, it only extracts local features without considering global information. To handle this issue, S^3-PCA was proposed to perform superpixel-based local reconstruction for HSI denoising plus extraction of PCA-based global features along with superpixelwise, PCA-based local features for global-local and spectral-spatial feature extraction, achieving better performance than SuperPCA. However, the superpixelwise PCA-based local features are not enough for accurate HSI classification because the application of the classic PCA to each superpixel is dominated by spectral features with high variance. In the case of superpixels with mixed objects or strong noise/outliers, both PCA and SuperPCA may provide unsatisfactory features.

C. LDA and LFDA

As a classic supervised DR model, LDA makes use of labels to extract more discriminative features than PCA. LDA finds the projection matrix P by maximizing the between-class variance and minimizing the within-class variance of the projection. Given a set of observations $\boldsymbol{X} = \{\boldsymbol{x_n}\} \in \mathbb{R}^{B \times N}$, along with the corresponding labels $\boldsymbol{L} = \{c_n\}(n = 1, ..., N)$, $\boldsymbol{c_n} = \{1, 2, ..., C\}$, LDA finds a projection matrix $\boldsymbol{P}$ to obtain low-dimensional features $\boldsymbol{H} = \boldsymbol{P}^T \boldsymbol{X}$ with the objective function

$$\boldsymbol{P} = \arg\max_{\boldsymbol{P}} \left(\frac{\boldsymbol{P}^T \boldsymbol{S_b} \boldsymbol{P}}{\boldsymbol{P}^T \boldsymbol{S_w} \boldsymbol{P}}\right) \quad \text{with} \quad (3)$$

$$\boldsymbol{S_w} = \sum_{c=1}^{C} \sum_{j=1}^{n_c} (\boldsymbol{x_j^c} - \boldsymbol{u^c})(\boldsymbol{x_j^c} - \boldsymbol{u^c})^T, \quad (4)$$

$$\boldsymbol{S_b} = \sum_{c=1}^{C} (\boldsymbol{u^c} - \boldsymbol{u})(\boldsymbol{u^c} - \boldsymbol{u})^T,$$

where $\boldsymbol{S_w}$ is the within-class scatter matrix, $\boldsymbol{S_b}$ is the between-class scatter matrix, n_c and $\boldsymbol{x_j^c}$ are the number of samples belonging to class c and the j-th sample in the c-th class, respectively, $\boldsymbol{u^c}$ denotes the mean vector of the c-th class and u denotes the mean vector of the whole data set $\boldsymbol{X}$.

The optimization of LDA can also be transformed to solve an eigendecomposition problem.

Although LDA can find discriminative features, these features are low-dimensional (with dimensionality up to $C-1$). In addition, LDA only focuses on the global information, without considering the local (spatial-contextual) information. If there are multiple streams and multiple categories in the observation, LDA may be unable to find effective features for classification. To address these issues, LFDA was introduced by fusing LDA and LPP. The objective function of LFDA is similar to that of LDA, with two differences: the redefinition of the local within-class scatter matrix $\widetilde{S}^{w}$, and the redefinition of the between-class scatter matrix $\widetilde{S}^{b}$

$$\widetilde{S}^{w} = \frac{1}{2} \sum_{i,j=1}^{N} \widetilde{W}_{i,j}^{w}(x_i - x_j)(x_i - x_j)^T, \quad (5)$$

$$\widetilde{S}^{b} = \frac{1}{2} \sum_{i,j=1}^{N} \widetilde{W}_{i,j}^{b}(x_i - x_j)(x_i - x_j)^T,$$

with

$$A_{i,j} = exp\left(-\frac{||x_i - x_j||_2^2}{\delta^2}\right) \quad (6)$$

$$\widetilde{W}_{i,j}^{w} \equiv \begin{cases} A_{i,j}/n_c & if \ label(x_i) = label(x_j) = c \\ 0 & if \ label(x_i) \neq label(x_j) \end{cases}$$

$$\widetilde{W}_{i,j}^{b} \equiv \begin{cases} A_{i,j}(1/N - 1/n_c) & if \ label(x_i) = label(x_j) = c \\ 1/N & if \ label(x_i) \neq label(x_j) \end{cases}$$

where the parameter δ controls the 'decay' of the affinity.

LFDA can effectively retain the local multimodality of data from the same class and address the issue that LDA only focuses on the global structure of the observations.

Interestingly, LDA has been successfully extended for unsupervised feature extraction from HSIs by adopting LDA and pseudo-labels of superpixels in [15]. The resulting FG_SuULDA method has shown promising results. Although a 3D flexible Gabor transformation was introduced to enhance spatial features, its performance may be unsatisfactory due to the lack of local feature representation. To address this issue, superpixel-level hybrid discriminant analysis was introduced in [31], but the resulting SHDA model may be also ineffective because the global and local discriminant analysis models are conducted simultaneously, resulting in the global features that may decrease the effectiveness of local feature representation.

III. PROPOSED METHOD

In this section, we introduce the proposed spectral-spatial and superpixelwise unsupervised LDA (S^3-ULDA) method for feature extraction and classification of HSIs. As it can be seen from the flowchart in Fig. 1, PCA is first utilized to reduce the dimensionality of the HSI, and then the first principal component is fed to the ERS image segmentation model to obtain multiple homogeneous superpixel blocks, where pixels in each superpixel belong to the same class (with pseudo-labels provided by ERS). Secondly, in order to denoise the HSI, we locally reconstruct each pixel in every superpixel by using their S nearest neighbors within the same superpixel.

Thirdly, we use unsupervised LDA and pseudo-labels from ERS to global feature extraction from the HSI, where the original and locally reconstructed pixels are simultaneously adopted to construct the within-superpixel scatter matrix and the between-superpixel scatter matrix in the LDA framework, leading to superpixelwise unsupervised LDA (SuperULDA). Next, to extract effective local features from the HSI, we propose a superpixelwise unsupervised LFDA (SuperULFDA) where each superpixel (along with its adjacent superpixels and pseudo-labels) are modeled by means of an unsupervised LFDA to obtain local features from the HSI (on a superpixel-by-superpixel basis). The application of unsupervised LFDA on local areas of the HSI can effectively fit the local multi-modality of samples in local superpixels and address the issue that LDA only focuses on global structures. As a result, the local features extracted by the proposed SuperULFDA are more effective for classification than those provided by S^3-PCA [16] and SHDA [31]. Finally, by fusing the global features from SuperULDA and the local features from SuperULFDA, the proposed S^3-ULDA can obtain effective global-local and spectral-spatial features.

Let us denote an HSI data set by $X^3 \in \mathbb{R}^{M \times N \times B}$, where M, N and B are the width, height, and number of spectral bands, respectively. Typically, we reshape a HSI data cube into a matrix $X = \{x_1, x_2, ..., x_{MN}\} \in \mathbb{R}^{B \times MN}$, in which each x_i corresponds to a pixel. We first perform ERS superpixel segmentation to segment the whole HSI into K superpixels, denoted by $X^k = \{x_1^k, x_2^k, ..., x_{n_k}^k\}, (k = 1, ..., K)$, where x_i^k and n_k are the i-th pixel and the number of the pixels in the k-th superpixel, respectively. In addition, with superpixel-based local reconstruction, we can obtain the denoised HSI data $X^* = \{x_1^*, x_2^*, ..., x_{MN}^*\}$, in which each x_i^* corresponds to the locally reconstructed pixel of x_i in the original HSI data. Similarly, we use X^{*k} to denote all the pixels of the k-th locally reconstructed superpixel, corresponding to superpixel X^k in the original HSI data. The goal of feature extraction is to obtain effective, low-dimensional, spectral-spatial features $H = P^T X$ with the optimal projection matrix $P \in \mathbb{R}^{B \times d}$, where d is the dimensionality of the reduced space.

As superpixel-based local reconstruction in S^3-PCA can be used to denoise HSI data with good results, we adopt such method in this paper. However, different from S^3-PCA (where only the locally reconstructed data are employed in the subsequent steps), we exploit both the locally reconstructed data and the original data. Specifically, after ERS-based superpixel segmentation, each pixel of the HSI is locally reconstructed by its surrounding S neighboring pixels in the same superpixel block. For the i-th pixel x_i^k in the k-th superpixel, its S nearest spatial neighbors belonging to the same superpixel are $\{x_i^{n_1}, ..., x_i^{n_S}\}$ and we can obtain the reconstructed data $X^* = \{x_i^{*k}\}_{i=1,...,MN}^{k=1,...,K}$ by

$$x_i^{*k} = \sum_{j=1}^{S} w_j \times x_i^{n_j}, w_j = \frac{exp(-||x_i^k - x_i^{n_j}||_2^2/2t^2)}{\delta^2}, \quad (7)$$

where $t = (1/S) \sum_{j=1}^{S} ||x_i^k - x_i^{n_j}||_2^2$, $\delta = \sum_{j=1}^{S} exp(-||x_i^k - x_i^{n_j}||_2^2/2t^2)$, and w_j is defined as the similarity weight between x_i^k and $x_i^{n_j}$.
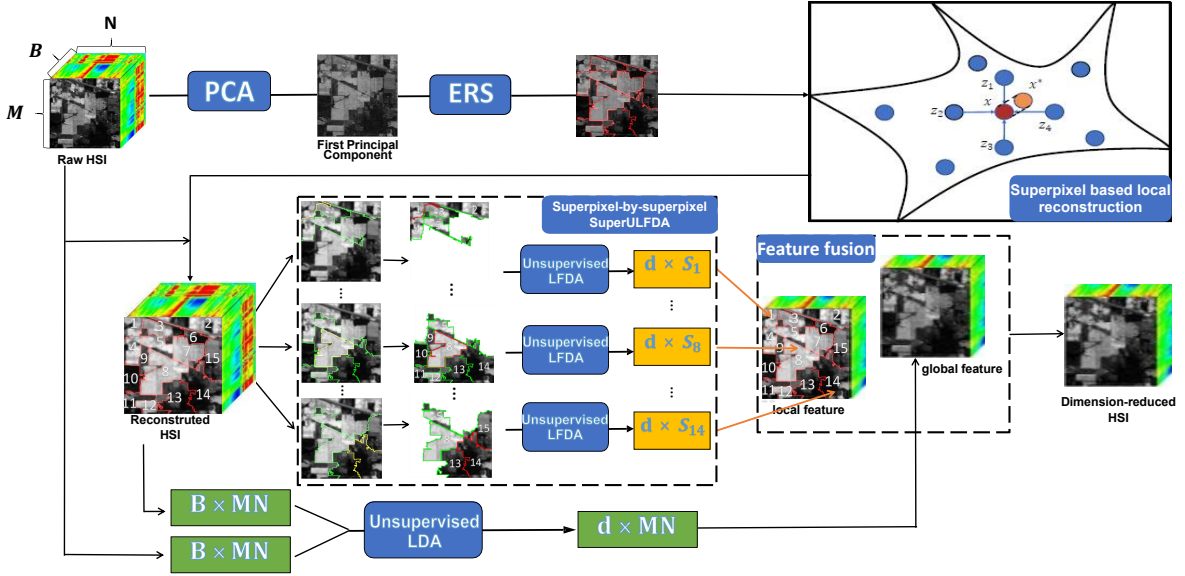
Fig. 1. Outline of the proposed S^3-ULDA method.

A. SuperULDA-based Global Feature Extraction

After the superpixel-based local reconstruction, we extract the global spectral–spatial features from the HSI by the proposed SuperULDA. Compared to typical PCA-based feature extraction, LDA provides more discriminative features by using label information. However, for unsupervised HSI feature reduction tasks without labels, LDA cannot be adopted. Inspired by the recently proposed FG_SuULDA, where classic LDA and the pseudo-labels provided by superpixel segmentation were adopted for unsupervised feature extraction, we simultaneously make use of the locally reconstructed data and the original data to conduct unsupervised feature extraction based on LDA. Our new model can extract effective global features from high-dimensional and noisy HSI data.

To efficiently make use of the superpixel-based local reconstruction and pseudo-labels, we extend LDA to utilize both the original data X and the locally reconstructed data X^* to globally calculate the within-superpixel scatter matrix and the between-superpixel scatter matrix. Let us denote the mean vector u^k of the k-th superpixel and the mean vector of the whole HSI u by

$$u^k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i^k, \qquad u = \frac{1}{M \times N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} x_i^k. \quad (8)$$

In addition, we can similarly define the mean vector of the k-th reconstructed superpixel u^{*k} and the mean vector of the whole reconstructed HSI u^* in the same way. Based on the above notations, the objective function of SuperULDA is

$$P_g = \arg\max_{P_g} \left(\frac{P_g^T S_b^g P_g}{P_g^T S_w^g P_g} \right), \quad (9)$$

where P_g denotes the projection matrix, S_w' is the global within-superpixel scatter matrix and S_b' is the global between-superpixel scatter matrix.

Different from FG_SuULDA, in the proposed SuperULDA the original and locally reconstructed data are simultaneously

utilized to globally derive the within-superpixel scatter matrix S_w^g and the between-superpixel scatter matrix S_b^g in the LDA framework. Mathematically, S_w^g and S_b^g can be redefined by

$$S_w^g = \sum_{k=1}^{K} \left(\sum_{i=1}^{n_k} (x_i^k - u^k)(x_i^k - u^k)^T + \right. \quad (10)$$

$$\left. \sum_{i=1}^{n_k} (x_i^{*k} - u^{*k})(x_i^{*k} - u^{*k})^T \right),$$

$$S_b^g = \sum_{k=1}^{K} n_k (u^k - u)(u^k - u)^T + \quad (11)$$

$$\sum_{k=1}^{K} n_k (u^{*k} - u^*)(u^{*k} - u^*)^T,$$

where S_w^g depicts the distances between pixels in the same superpixel and the mean vector, and S_b^g measures the distances between mean vectors of different superpixels. It is expected that the elements of S_w^g are as small as possible, so that the samples from the same superpixel can be closer to each other, while the elements of S_b^g are as large as possible, so that the pixels from different superpixels are far away from each other.

The solution of the proposed SuperULDA can be found by solving the following eigendecomposition problem

$$S_w^{g^{-1}} S_b^g P_g = \lambda P_g \quad (12)$$

The eigenvectors corresponding to the largest d eigenvalues are used to construct the projection matrix $P_g \in \mathbb{R}^{B \times d}$, which will be used to extract the global spectral–spatial feature H_g as follows

$$H_g = P_g^T X^*, \quad (13)$$

where X^* is the HSI data denoised by superpixel-based local reconstruction.
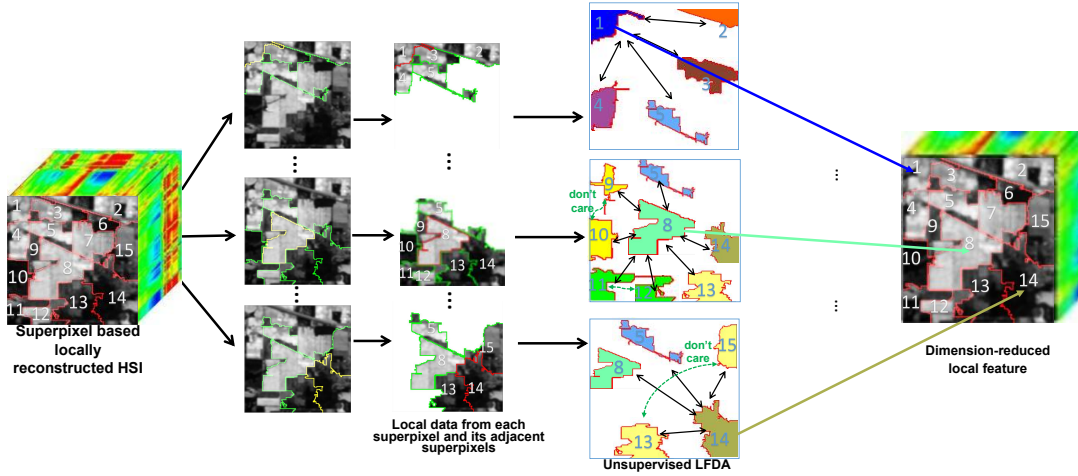
Fig. 2. Outline of the proposed SuperULFDA for local feature extraction on a superpixel-by-superpixel basis. For each superpixel, the local data (consisting of the superpixel and its adjacent superpixels, along with the pseudo-labels of the superpixels) are fed into unsupervised LFDA to extract local features.

B. SuperULFDA-based Local Feature Extraction

SuperULDA is designed to extract effective global features, but it ignores the local (spatial-contextual) structures in HSIs and can only reduce the dimensionality of low-dimensional features up to $K-1$, where K is the number of superpixels. To handle these issues, the SuperULFDA framework (illustrated in in Fig. 2) is proposed to extract local features on a superpixel-by-superpixel basis, where the local data from each superpixel and its adjacent superpixels (along with the pseudo-labels of the superpixels) are fed into unsupervised LFDA for local feature extraction. LFDA can manage the multimodality of superpixels in the HSI data, where samples in different superpixels are separated and far-apart samples belonging to the same class are made close. Just as an example, for the 14-th superpixel in Fig. 2 in case its adjacent (13-th and 15-th) superpixels belong to the same class, LFDA could effectively group them together, because LFDA does not care about the similarity between the 13-th and 15-th superpixels. Moreover, LFDA addresses the issue that LDA only focuses on the global structures with a limitation on the number of reduced features (up to $K-1$, where K is the number of adjacent superpixels). Thanks to the above advantages of LFDA, the proposed SuperULFDA can effectively extract local features by considering superpixel-level local structures.

Let us denote by $k = 1, 2, ..., K$ the pseudo-labels of superpixels provided by ERS. For each locally reconstructed superpixel $\boldsymbol{X^{*k}}$ [with pseudo-label $PL(X^{*k}) = k$], where n_0^k is the number of pixels, its adjacent superpixels are given by $\boldsymbol{X_1^{*k}}, \boldsymbol{X_2^{*k}}, ..., \boldsymbol{X_S^{*k}}$, with $n_1^k, n_2^k, ..., n_S^k$ pixels. The number of pixels in the k-th superpixel (along with its adjacent superpixels) is $N_k = n_0^k + n_1^k + n_2^k + ... + n_S^k$. Let $\boldsymbol{A}$ be the similarly matrix with the (i, j)-th element representing the similarity between $\boldsymbol{x_i^*}$ and $\boldsymbol{x_j^*}$. The objective function of the proposed SuperULFDA for each superpixel and its adjacent superpixels is

$$\boldsymbol{P_l} = \arg\max_{\boldsymbol{P_l}} \left[tr\left((\boldsymbol{P_l}^T \boldsymbol{S_w^l} \boldsymbol{P_l})^{-1} \boldsymbol{P_l}^T \boldsymbol{S_b^l} \boldsymbol{P_l} \right) \right], \quad (14)$$

where $\boldsymbol{P_l} \in \mathbb{R}^{B \times d}$ is the projection matrix, and the local within-superpixel scatter matrix $\boldsymbol{S_w^l}$ and the local between-superpixel scatter matrix $\boldsymbol{S_b^l}$ are defined in Eq. (15) based on the locally reconstructed data $\boldsymbol{X^*}$ as

$$\boldsymbol{S_w^l} = \frac{1}{2} \sum_{i,j=1}^{N_k} \widetilde{W}_{i,j}^w (\boldsymbol{x_i^*} - \boldsymbol{x_j^*})(\boldsymbol{x_i^*} - \boldsymbol{x_j^*})^T, \quad (15)$$

$$\boldsymbol{S_b^l} = \frac{1}{2} \sum_{i,j=1}^{N_k} \widetilde{W}_{i,j}^b (\boldsymbol{x_i^*} - \boldsymbol{x_j^*})(\boldsymbol{x_i^*} - \boldsymbol{x_j^*})^T, \quad (16)$$

with

$$A_{i,j} = exp\left(-\frac{||\boldsymbol{x_i^*} - \boldsymbol{x_j^*}||_2^2}{\delta^2} \right)$$

$$\widetilde{W}_{i,j}^w \equiv \begin{cases} A_{i,j}/n_s^k & if\ PL(\boldsymbol{x_i^*}) = PL(\boldsymbol{x_j^*}) \\ 0 & if\ PL(\boldsymbol{x_i^*}) \neq PL(\boldsymbol{x_j^*}) \end{cases}$$

$$\widetilde{W}_{i,j}^b \equiv \begin{cases} A_{i,j}(1/N_k - 1/n_s^k) & if\ PL(\boldsymbol{x_i^*}) = PL(\boldsymbol{x_j^*}) \\ 1/N_k & if\ PL(\boldsymbol{x_i^*}) \neq PL(\boldsymbol{x_j^*}) \end{cases}$$

and δ being a parameter that controls the 'decay' of the affinity.

The solution of the proposed SuperULFDA is similar to that of classical LFDA, i.e., by solving the generalized eigen-decomposition problem in Eq. (12). Once the projection matrix $\boldsymbol{P_l}$ is optimized, the local spectral–spatial features with respect to the samples in the k-th superpixel can be learnt by

$$\boldsymbol{H_l^k} = \boldsymbol{P_l}^T \boldsymbol{X^{*k}}, \quad (17)$$

and then the local spectral–spatial features of the HSI ($\boldsymbol{H_l}$) can obtained by aggregating all the superpixels as follows: $\boldsymbol{H_l} = \{\boldsymbol{H_l^1}, \boldsymbol{H_l^2}, ..., \boldsymbol{H_l^K}\}$.

Compared to SHDA, where the local structures represented by superpixel-level local graphs are considered in the unsupervised LDA framework (along with superpixel-level global graphs), the proposed SuperULFDA directly extracts local features on a superpixel-by-superpixel basis.

C. S^3-ULDA-based Global-Local and Spectral–spatial Feature Extraction

The global features(H_g) from SuperULDA and the local features (H_l) are fused by concatenating them along the feature dimension, leading to a set of global–local and spectral–spatial features (H_s) which is simply defined as follows

$$H_s = [H_g, H_l] \in \mathbb{R}^{MN \times 2d} \tag{18}$$

Specifically, the proposed S^3-ULDA adopts SuperULDA to extract global features and SuperULFDA to extract local features. We provide the pseudo-code of S^3-ULDA in Algorithm 1. As for the complexity of the proposed model, the superpixel-based local reconstruction costs $O(MN)$, and to solve Eq. (12) and Eq. (14) requires $O(B^3)$, which leads to $O(MN + KB^3)$ for the proposed model with M, N, B, K being the width, height, number of spectral bands and number of superpixels, respectively. With parallelization strategies such as multithreading techniques, the complexity of the proposed model can be further reduced to $O(MN/K + B^3)$ because solving Eq. (7) and Eq. (14) in superpixel-by-superpixel fashion is independent from each other.

Algorithm 1 The proposed S^3-ULDA for unsupervised feature extraction from HSIs

---

Input: HSI data $X \in \mathcal{R}^{M \times N \times B}$, number of superpixels K for ERS-based segmentation, number of nearest neighbors S for superpixel-based local reconstruction, dimensionality of reduced features d.

1: Apply PCA to the HSI data X to obtain the first principal component;
2: Conduct ERS-based image segmentation on the feature-reduced data;
3: **for** superpixel $k = 1, ..., K$ **do**
4:    **for** each pixel in the k-th superpixel $x_i^k (i = 1, ..., n_k)$ **do**
5:       Conduct superpixel-based local reconstruction using Eq. (7);
6:    **end for**
7: **end for**
8: Solve Eq. (12) to obtain the global features $H_g \in \mathbb{R}^d$ in the original data X and in the locally reconstructed data X^*, plus the corresponding pseudo-labels;
9: **for** each superpixel $k = 1, ..., K$ **do**
10:    Obtain local data from the k-th superpixel X_k^* and its adjacent superpixels, along with the corresponding pseudo-labels;
11:    Solve Eq. (14) to obtain the local features $H_l^k \in \mathbb{R}^d$ with respect to the k-th superpixel using the local data;
12: **end for**
13: Concatenate the global features H_g and the local features H_l to obtain a set of global–local and spectral–spatial features $H_s \in \mathbb{R}^{2d}$.
Output: Extracted features H_s.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

To demonstrate the effectiveness of the proposed S^3-ULDA, we first conduct a parameter sensitivity analysis to select the optimal parameters of the proposed model. Then, ablation experiments are conducted to show the performance of SuperULDA, SuperULFDA and S^3-ULDA. Finally, we compare the proposed S^3-ULDA to classical and state-of-the-art feature extraction techniques, including classic unsupervised models (PCA, NPE), classic supervised models (LFDA, NWFE), recent unsupervised models (OTVCA [4], LNSPE [7]) superpixel-based unsupervised models (SuperPCA [30], S^3-PCA [16], FG_SuULDA [15], SHDA [31]) and deep learning models (3DCNN [18], FuNet-C [19], GiGCN [20], DSGSF [23] and C-SS-MTr [26]). For all the considered DR models, when dimensionally-reduced features are extracted, an SVM classifier with radial basis function (RBF) kernel and grid searched parameters is trained in the learnt low-dimensional feature space randomly selected training samples from the labeled data, and the overall accuracy (OA), average accuracy (AA) and Kappa coefficients corresponding to the prediction of test data are utilized for objectively comparing the different models. For all the considered deep learning models, the classification accuracies are based on the embedded end-to-end classifiers with the optimal hyperparameters reported in the original papers. All the comparative deep learning models are re-trained with the reported optimal hyperparameters on the training data, which are the same for all comparative models in experiments to ensure the fairness of comparison. Moreover, to subjectively compare the proposed models to state-of-the-art superpixel-based unsupervised DR models, visualization experiments are performed in the learnt 2D/3D subspace.

TABLE I
THE NUMBER OF SAMPLES FOR EACH CLASS IN INDIAN PINES, UNIVERSITY OF PAVIA AND HOUSTON 2013 DATA

Indian Pines		University of Pavia		Houston 2013	
Class Names	Numbers	Class Names	Numbers	Class Names	Numbers
Alfalfa	46	Asphalt	6631	Healthy grass	1251
Corn-notill	1428	Bare soil	18649	Stressed grass	1254
Corn-mintill	830	Bitumen	2099	Synthetic grass	697
Corn	237	Bricks	3064	Trees	1244
Grass-pasture	483	Gravel	1345	Soil	1242
Grass-trees	730	Meadows	5029	Water	325
Grass-pasture-mowed	28	Metal sheets	1330	Residential	1268
Hay-windrowed	478	Shadows	3682	Commercial	1244
Oats	20	Trees	947	Road	1252
Soybean-notill	972			Highway	1227
Soybean-mintill	2455			Railway	1235
Soybean-clean	593			Railway	1233
Wheat	205			Parking Lot 2	469
Woods	1265			Tennis Court	428
Buildings-Grass-Trees-Drives	386			Running Track	660
Stone-Steel-Towers	93				
Total Number	10249	Total Number	42776	Total Number	15029

Three data sets are used in the experiments:

- The Indian Pines image was gathered by the airborne visible infra-red imaging spectrometer (AVIRIS). By removing the water absorption bands, the number of effective spectral bands is 200 and the image comprises 145×145 pixels and 16 classes.
- The University of Pavia image was collected by the reflective optics spectrographic imaging system (RO-SIS). The number of effective spectral bands is 103 after removing 12 noisy bands and the image comprises 610×340 pixels and 9 classes.
- The Houston 2013 image was acquired by the compact airborne spectrographic imager (CASI) over the campus of the University of Houston and the neighboring urban

(a) Indian Pines

(b) University of Pavia

(c) Houston 2013
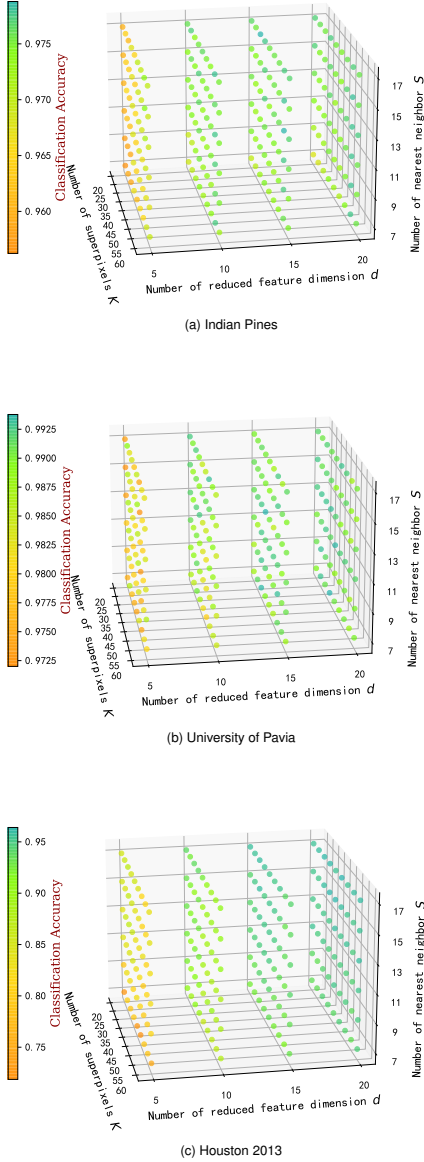
Fig. 3. Parameter sensitivity analysis for the proposed S^3-ULDA on three data sets.

area. The number of effective spectral bands is 144 and the image comprises 349×1905 pixels and 15 classes.

The sample sizes regarding each groundtruth class in the three data sets are shown in Table I, and we use the same random seed values while creating training and testing data, so that we can generate the same training and testing data for all the models in the experiments to fairly assess and compare the performance of different models. Each experiment is repeated ten times and the average results are provided. Methods were implemented on MATLAB R2016a, PyCharm and executed in Windows 10 64-bit platform equipped with AMD R7-5800H CPU, NVIDIA GeForce RTX 3060 GPU and 24-GB RAM.

A. Parameter Sensitivity Analysis

There are a few parameters in the proposed S^3-ULDA, including the number of nearest neighbors S in superpixel-

based local reconstruction, the number of superpixels K in ERS-based segmentation, and the dimensionality of reduced features d (in fact, the number of reduced features in the proposed model is $2d$ as SuperULDA and SuprULFDA provide d features each. In order to pick the optimal parameters, five-fold cross-validation is used to conduct the parameter sensitivity analysis for $S \in \{7, 9, ..., 17\}, K \in \{20, 25, 30, ..., 60\}, d \in \{5, 10, 15, 20\}$. Specifically, we choose 50 labeled samples from each class, where 80% of these samples are for training and the remaining samples for validation.

We select the parameters corresponding to the highest average OAs from the five-fold cross-validation as the optimal parameters. The experimental results in Fig. 3(a-c) show that the optimal parameters are $d = 10, 15, 15$, $K = 35, 35, 30$ and $S = 17, 13, 15$ for the three considered data (Indian Pines, University of Pavia and Houston 2013), respectively. Also, we can see that the parameters of the proposed model are relatively stable for the three data (with different spatial and spectral resolution), so we anticipate that the same parameters can be used for other HSIs. In fact, we simply suggest using as default parameters for the proposed model: $d = 15$, $K = 35$ and $S = 15$. For the SVM classifier, the RBF kernel is selected with the optimal parameters set to $C = 100000$ and $\gamma = 20, 30, 50$ for the three data sets, respectively.

B. Ablation Experiments

We conduct ablation experiments to show the effectiveness of the extracted global, local, and fused global-local features from SuperULDA, SuprULFDA and S^3-ULDA, respectively. Specifically, to show the effectiveness of the locally reconstructed data, ablation experiments are conducted for the proposed models by only using the original (ori), locally reconstructed (rec) and both the original and locally reconstructed data (all), which results in the following terminology:

1) SuperULDA(ori), SuperULFDA(ori) and S^3-ULDA(ori) denote models applied to the original HSIs.
2) SuperULDA(rec), SuperULFDA(rec) and S^3-ULDA(rec) refer to models applied to locally reconstructed HSIs.
3) SuperULDA(all), SuperULFDA(all) and S^3-ULDA(all) refer to models applied to both the original and locally reconstructed HSIs.

Table II provides an overview of the ablation experiments. As the proposed SuperULFDA can be utilized to extract the global features with all superpixels or the local features with adjacent superpixels on a superpixel-by-superpixel basis, we further refer to them as global SuperULFDA and local SuperULFDA, respectively. In contrast, there is no local SuperULDA because SuperULDA cannot be directly used to extract local features with adjacent superpixels. This is due to the fact that the number of reduced features is limited to K, i.e., the number of pseudo-labels of local adjacent superpixels that is typically less than 5. For the proposed S^3-ULDA, consisting of global SuperULFDA(all) and local SuperULFDA(rec), we extract the global features by using SuperULFDA on all of the original and locally reconstructed data, and concatenate to

TABLE II
OVERVIEW OF ABLATION EXPERIMENTS

Models	Unsupervised LDA	Using original data	Using reconstructed data	Unsupervised LFDA	Using original data	Using reconstructed data	With all superpixels	With superpixels by superpixels	Fusing the global and local features
global SuperULDA(ori)	Y	Y					Y		
global SuperULDA(rec)	Y		Y				Y		
global SuperULDA(all)	Y	Y	Y				Y		
global SuperULFDA(ori)				Y	Y		Y		
global SuperULFDA(rec)				Y		Y	Y		
global SuperULFDA(all)				Y	Y	Y	Y		
local SuperULFDA(ori)				Y	Y			Y	
local SuperULFDA(rec)				Y		Y		Y	
local SuperULFDA(all)				Y	Y	Y		Y	
S^3-ULDA(ori)	Y	Y		Y	Y		Y	Y	Y
S^3-ULDA(rec)	Y		Y	Y		Y	Y	Y	Y
S^3-ULDA(all)	Y	Y	Y	Y	Y	Y	Y	Y	Y
S^3-ULDA	Y	Y	Y	Y		Y	Y	Y	Y

TABLE III
OAS OBTAINED BY DIFFERENT METHODS IN ABLATION EXPERIMENTS (IN ALL CASES, SVM IS USED FOR THE FINAL CLASSIFICATION. BEST RESULTS IN BOLD.)

Datasets	T.N.s/C	Global						Local			Gloabl-Local			
		global SuperULDA(ori)	global SuperULDA(rec)	global SuperULDA(all)	global SuperULFDA(ori)	global SuperULFDA(rec)	global SuperULFDA(all)	local SuperULFDA(ori)	local SuperULFDA(rec)	local SuperULFDA(all)	S^3-ULDA (ori)	S^3-ULDA (rec)	S^3-ULDA (all)	S^3-ULDA
Indian Pines	10	65.89%	83.26%	**84.10%**	55.76%	75.87%	76.56%	83.12%	85.65%	**85.83%**	83.83%	85.12%	**88.13%**	87.67%
	20	72.23%	90.98%	**91.49%**	66.03%	84.19%	86.17%	91.30%	93.40%	**93.84%**	92.34%	93.01%	**94.96%**	94.95%
	30	74.87%	93.45%	**94.41%**	69.40%	88.27%	89.94%	92.68%	94.77%	**95.73%**	93.74%	94.96%	96.56%	**96.68%**
University of Pavia	10	73.15%	81.63%	**86.66%**	74.41%	86.51%	84.49%	77.96%	**88.63%**	79.51%	87.51%	91.63%	93.22%	**94.81%**
	20	76.74%	87.22%	91.16%	79.21%	**91.66%**	90.25%	84.33%	**93.85%**	87.67%	92.17%	96.34%	97.27%	**98.14%**
	30	79.59%	90.07%	93.03%	81.10%	**93.34%**	92.61%	85.73%	**95.63%**	90.81%	93.91%	97.32%	98.16%	**98.69%**
Houston 2013	10	64.76%	71.03%	84.63%	74.64%	**85.97%**	83.35%	69.41%	**82.13%**	79.96%	83.43%	83.49%	84.85%	**86.17%**
	20	71.74%	78.61%	90.69%	78.99%	90.87%	**90.97%**	77.82%	**89.15%**	87.23%	90.09%	89.79%	90.49%	**91.33%**
	30	74.85%	83.07%	**93.73%**	81.07%	92.95%	93.02%	83.15%	**92.72%**	91.82%	93.63%	93.09%	94.10%	**94.42%**

them the local features obtained by using SuperULFDA on the locally reconstructed data.

Table III shows the obtained classification results when the number of training samples from each class is set to $T = 10, 20, 30$, with the number of superpixels set to $K = 35, 35, 30$, and the number of spatial neighbors set to $S = 17, 13, 15$ on Indian Pines, University of Pavia and Houston 2013. The dimensionality of the reduced features for all the models is set to 30. For the fused S^3-ULDA(ori), S^3-ULDA(rec), S^3-ULDA(all) and S^3-ULDA, $d = 15$ is set to be the number of the extracted global and local features respectively, which leads fused global-local features with dimensionality of 30. As it can be seen from Table III, superpixel-based local reconstruction can effectively denoise the data and then improve the performance of SuperULDA and SuperULFDA. This can be observed by comparing the global SuperULDA(ori) and SuperULDA(rec), the global SuperULFDA(ori) and SuperULFDA(rec), as well the local SuperULFDA(ori) and SuperULFDA(rec) in terms of OAs.

Another important observation is that global SuperULDA(all) significantly outperforms the corresponding global SuperULDA(ori) and global SuperULDA(rec). Similar results can be obtained for global SuperULFDA, where global SuperULFDA(all) outperforms global SuperULFDA(ori) and global SuperULFDA(rec) in most cases. However, for the proposed local SuperULFDA, using only locally reconstructed data appears to be the best solution, because the results in Table III show that local SuperULFDA(rec) outperforms local SuperULFDA(ori) and local SuperULFDA(all) in the last two datasets. This is probably due to the fact that the superpixel-by-superpixel local feature extraction model (local SuperULFDA) is sensitive to noise/outliers in the original data, which could largely compromise the performance of the local model, especially in the University of Pavia and Houston 2013 data sets, that comprise high spatial resolution and lots of mixed objects. Quite opposite, in the Indian Pines data (with lower spatial resolution, few mixed objects but many mixed pixels), local SuperULFDA(all) still outperforms local SuperULFDA(ori) and local SuperULFDA(rec), possibly because there are many mixed pixels that contain more than one material type in the low spatial resolution HSI. Overall, we can conclude that global SuperULDA using both the original
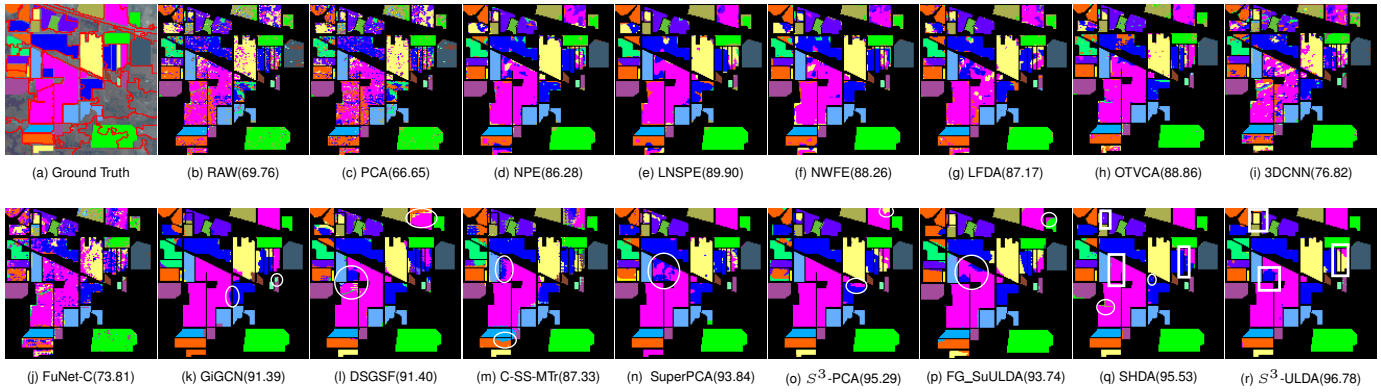
Fig. 4. Classification maps obtained by different models for the Indian Pines image. (a) Ground Truth. (b) Raw. (c) PCA. (d) NPE. (e) LNSPE. (f) NWFE. (g) LFDA. (h) OTVCA. (i) 3DCNN. (j) FuNet-C. (k) GiGCN. (l) DSGSF. (m) C-SS-MTr. (n) SuperPCA. (o) S3 -PCA. (p) FG_SuULDA. (q) SHDA. (r) S^3 -ULDA.
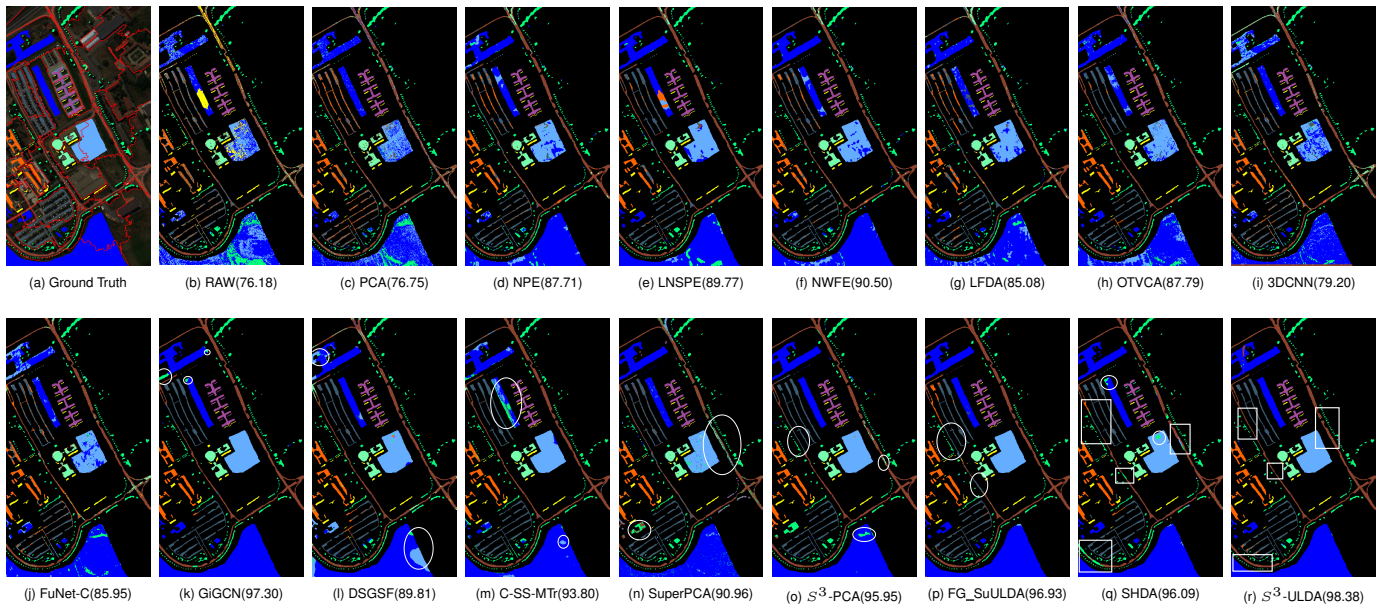
Fig. 5. Classification maps obtained by different models for the University of Pavia image. (a) Ground Truth. (b) Raw. (c) PCA. (d) NPE. (e) LNSPE. (f) NWFE. (g) LFDA. (h) OTVCA. (i) 3DCNN. (j) FuNet-C. (k) GiGCN. (l) DSGSF. (m) C-SS-MTr. (n) SuperPCA. (o) S3 -PCA. (p) FG_SuULDA. (q) SHDA. (r) S^3 -ULDA.

and locally reconstructed data is optimal for global feature extraction model, while only using the locally reconstructed data appears to be the best choice for the local SuperULFDA. Also, although the global SuperULFDA appears to provide sub-optimal performance, the local SuperULFDA provides significantly higher OAs compared to the global SuperULDA and the global SuperULFDA in all cases. This suggests that effective local features can be extracted by the local SuperULFDA on a superpixel-by-superpixel basis, because the unsupervised LFDA can effectively deal with the multimodality of superpixels and address the issue that unsupervised LDA only focuses on the global structures of the data. Thus, we fuse the global SuperULDA(all) and the local SuperULFDA(rec), leading to the proposed S^3-ULDA which outperforms S^3-ULDA(ori), S^3-ULDA(rec) and S^3-ULDA(all). Although S^3-ULDA(all) could further improve the proposed S^3-ULDA in the Indian Pines data set, it provides worse results than the

proposed S^3-ULDA in the other two images due to their higher spatial resolution (and to the presence of mixed objects). Thus, the proposed S^3-ULDA will be based on the global features from SuperULDA(all) and the local features from SuperULFDA(rec) in the following experiments.

C. Model Comparison

In this subsection, we conduct extensive experiments to compare the proposed S^3-ULDA with classical and state-of-the-art DR models as well as deep learning models as follows,

1) PCA, NPE: They are the classic unsupervised DR models to conduct spectral feature reduction.
2) LFDA, NWFE: They are the classic supervised DR models to conduct spectral feature reduction.
3) OTVCA [4], LNSPE [7]: They are the variants of PCA and NPE for unsupervised spectral feature reduction.

4) SuperPCA [30]: It uses PCA in each homogeneous superpixel to obtain effective local spatial-spectral features.

5) S^3-PCA [16]: It uses superpixels based local reconstruction to denoise HSIs, and then global features are extracted by PCA, while local features are obtained by performing PCA in each superpixel, followed by the fusion of global and local spatial-spectral features.

6) FG_SuULDA [15]: It uses unsupervised LDA to extract the global features, where the pseudo labels of superpixels are considered as the labels of classic LDA.

7) SHDA [31]: It uses unsupervised global and local discriminant analysis models based on the pseudo labels of superpixels to extract global features by considering local/nonlocal spatial–spectral correlation information among/between superpixels.

8) 3DCNN [18]: It uses 3D convolution kernels to extract the spatial-spectral information and conduct end-to-end classification.

9) FuNet-C [19]: It fuses minibatch GCN (miniGCN) and CNNs to extract effective spectral-spatial features and conduct end-to-end classification.

10) GiGCN [20]: It uses superpixel-based graph-in-graph model to extract effective features of superpixels and then superpixel-based GCN is adopted for classification.

11) DSGSF [23]: It fuses a global spatial feature representation subnetwork and a dual-view spectral aggregation subnetwork to learn the discriminative spatial-spectral features and conduct classification.

12) C-SS-MTr [26]: It uses masked autoencoder to pre-train a vanilla Transformer to learn effective features and then fine-tune the model with additional contrastive loss, which brings instance discriminability.

Generally, the comparative methods cover traditional unsupervised as well as supervised DR models, superpixels based unsupervised DR models, CNN based models, GNN based models, Transformer based models, which are complete and sufficient for comparison. For classic models PCA, NPE, LFDA and NWFE, the default parameters are used. For OTVCA, LNSPE, SuperPCA, S^3-PCA, FG_SuULDA and SHDA, the experimental settings are the same as their original papers. For deep learning models 3DCNN, Funet-C, GiGCN, DSGSF and C-SS-MTr, the suggested architectures and the corresponding hyperparameters settings in the original papers and source codes are used to conduct comparative experiments.

We first show the classification results when the number of labeled samples from each class is set to $T = 30$, with the numbers of superpixels set to $K = 35, 35, 30$, the number of spatial neighbors set to $S = 17, 13, 15$, and the dimensionality of reduced features set to $d = 10, 15, 15$ ($2d = 20, 30, 30$ for the fused global-local features) in the Indian Pines, University of Pavia and Houston 2013 data, respectively. For the Indian Pines data, if the number of samples of a class is less than the number of labeled samples T, up to half of all labeled samples from this class are picked as the training data.

The classification maps obtained for Indian Pines, University of Pavia and Houston 2013 data are provided in Figs. 4, 5 and Fig. S1 in the supplementary material of the paper due to the oversize of the figures. As can see that the traditional unsu-

pervised DR methods (PCA, NPE) and the classic supervised DR models (NWFE, LFDA) perform poorly because they ignore spatial information. In contrast, some state-of-the-art unsupervised DR algorithms could achieve better performance, such as LNSPE (which utilizes local manifold structures). Recently, superpixelwise DR models such as SuperPCA and S^3-PCA have shown outstanding performance for superpixel segmentation-based spatial information extraction, followed by PCA-based local spatial-spectral feature extraction for each superpixel. By further using the pseudo-labels of superpixels, FG_SuULDA and SHDA exhibit promising performance by adopting unsupervised LDA and superpixel-based local manifold representation for unsupervised global feature extraction. Although SuperPCA, S^3-PCA, FG_SuULDA and SHDA provide high OAs, the proposed S^3-ULDA outperforms them. As shown in the classification maps, the classification errors denoted by the circles in the classification maps of SuperPCA, S^3-PCA, FG_SuULDA and SHDA can be rectified by the proposed S^3-ULDA. Moreover, compared to SHDA (which is also based on pseudo-labels of superpixels and unsupervised LDA for global feature extraction, with the consideration of local manifold structures), the proposed S^3-ULDA could simultaneously extract more effective global and local features. Especially, the local feature extraction by the proposed S^3-ULDA is conducted on a superpixel-by-superpixel basis, which is more effective than SHDA in the areas marked by rectangular boxes in the classification maps reported in Figs. 4(q, r) and 5(q, r). In addition, although the deep learning models GiGCN based on superpixels and GCN, plus C-SS-MTr based on masked autoencoder and Transformer show outstanding accuracy compared to other deep learning methods, the proposed S^3-ULDA could significantly outperform them, which demonstrates the effectiveness of the model. The full classification results in Tables IV, V and VI support the aforementioned conclusions. Obviously, the proposed S^3-ULDA achieved the best performance in terms of OA, AA, and Kappa in the three data.

As for the time complexity, the comparison of the training times for all the models is given in the tables. Since FG_SuULDA needs a series of 3D flexible Gabor filters to exploit the spatial–spectral features, its complexity is higher than that of other models. In terms of the proposed S^3-ULDA, the SuperULFDA based local feature extraction on a superpixel-by-superpixel basis significantly increases the complexity of the model compared to that of SuperPCA, S^3-PCA and SHDA. However, it is worth noting that the proposed S^3-ULDA can be easily accelerated through parallelization techniques, because the (time-consuming) local feature extraction model (SuperULFDA) in the proposed S^3-ULDA is conducted in superpixel-by-superpixel fashion, where each superpixel calculation is independent from others. As shown in Tables IV, V and VI, with multithreading techniques the training times of the proposed S^3-ULDA can be significantly reduced (from 12.24s, 483.51s and 2853.65s to 2.48s, 37.39s, 356.74s in the three HSIs, respectively).

To further verify the proposed S^3-ULDA, we randomly choose $T = \{5, 10, 20, 30, 40, 50, 60\}$ labeled samples from each class to show the changes in OA as the number of training

TABLE IV

OA, AA AND KAPPA OBTAINED BY SEVERAL FEATURE EXTRACTION ALGORITHMS (COMBINED WITH THE SVM CLASSIFIER) ON THE INDIAN PINES DATA SET. BEST RESULTS IN BOLD.

Class	Training Samples	PCA	NPE	LNSPE [7]	NWFE [2]	LFDA [9]	OTVCA [4]	3DCNN [18]	FuNet-C [19]	GiGCN [20]	DSGSF [23]	C-SS-MTr [26]	SuperPCA [30]	S^3-PCA [16]	FG_SuULDA [15]	SHDA [31]	S^3-ULDA
Alfalfa	23	95.65%	100.00%	100.00%	100.00%	100.00%	95.65%	95.65%	95.65%	100.00%	61.90%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Corn-notill	30	50.07%	80.47%	82.12%	79.26%	82.98%	73.39%	75.54%	61.52%	83.05%	87.58%	83.54%	92.85%	90.70%	80.62%	92.06%	97.35%
Corn-mintill	30	69.13%	88.63%	95.75%	90.38%	93.13%	92.13%	68.13%	46.38%	91.20%	80.04%	88.07%	92.50%	99.00%	94.00%	97.88%	89.50%
Corn	30	79.23%	93.72%	91.79%	93.24%	94.69%	85.99%	92.75%	85.99%	100.00%	81.82%	99.58%	95.65%	97.58%	100.00%	95.17%	98.55%
Grass-pasture	30	91.17%	94.70%	92.72%	95.58%	94.26%	91.17%	86.98%	91.17%	92.96%	99.28%	84.06%	99.34%	99.56%	95.58%	89.40%	98.23%
Grass-trees	30	86.14%	98.86%	97.86%	99.00%	99.00%	97.86%	93.43%	95.57%	96.57%	97.09%	96.85%	97.14%	97.14%	97.86%	100.00%	100.00%
Grass-pasture-mowed	14	85.71%	100.00%	100.00%	92.86%	100.00%	92.86%	100.00%	100.00%	100.00%	55.00%	100.00%	92.86%	85.71%	100.00%	92.86%	100.00%
Hay-windrowed	30	86.61%	98.44%	99.55%	98.88%	99.78%	97.54%	96.88%	97.32%	99.79%	99.78%	100.00%	99.55%	100.00%	99.78%	99.78%	99.78%
Oats	10	60.00%	90.00%	90.00%	100.00%	100.00%	100.00%	90.00%	100.00%	55.00%	87.50%	100.00%	100.00%	100.00%	100.00%	85.77%	98.30%
Soybean-notill	30	71.02%	77.39%	83.86%	82.80%	81.53%	83.65%	76.75%	60.62%	71.60%	85.12%	86.32%	83.55%	86.20%	91.93%	96.12%	94.89%
Soybean-mintill	30	46.60%	78.43%	84.95%	81.03%	75.67%	90.10%	61.57%	74.19%	97.92%	96.73%	78.66%	91.59%	94.89%	94.80%	98.93%	92.72%
Soybean-clean	30	68.21%	88.28%	94.49%	92.90%	90.05%	91.47%	72.65%	69.80%	75.72%	80.71%	71.33%	92.72%	95.38%	96.27%	99.43%	99.92%
Wheat	30	97.71%	100.00%	99.43%	100.00%	99.43%	98.86%	99.43%	99.43%	91.71%	93.96%	99.02%	99.43%	99.43%	100.00%	100.00%	100.00%
Woods	30	85.99%	90.85%	92.96%	95.14%	93.20%	93.20%	88.58%	83.32%	99.84%	97.44%	97.87%	99.84%	99.84%	97.81%	100.00%	100.00%
Buildings-Grass-Trees-Drives	30	56.46%	96.07%	99.16%	96.07%	95.79%	91.29%	73.88%	67.70%	98.45%	99.43%	98.19%	98.60%	96.91%	97.19%	99.44%	
Stone-Steel-Towers	30	98.41%	100.00%	100.00%	98.41%	98.41%	96.82%	100.00%	100.00%	97.85%	72.29%	98.92%	98.41%	98.41%	100.00%	100.00%	100.00%
OA		66.65%	86.28%	89.90%	88.26%	87.17%	88.86%	76.82%	73.82%	91.39%	91.40%	87.33%	93.84%	95.29%	93.74%	95.53%	96.78%
AA		76.76%	92.24%	94.04%	93.44%	93.62%	92.00%	85.76%	83.04%	90.73%	85.98%	92.65%	95.88%	96.30%	96.63%	96.61%	98.04%
Kappa		62.68%	84.43%	88.50%	86.64%	85.42%	87.28%	73.83%	70.04%	90.20%	90.18%	85.70%	92.94%	94.61%	92.85%	94.87%	96.32%
Time (s)		0.07	9.33	26.24	2.48	1.67	36.08	35.11	74.06	68.29	253.12	279.07	0.31	2.23	414.56	8.13	12.24(2.48)

TABLE V

OA, AA AND KAPPA OBTAINED BY SEVERAL FEATURE EXTRACTION ALGORITHMS (COMBINED WITH THE SVM CLASSIFIER) ON THE UNIVERSITY OF PAVIA DATA SET. BEST RESULTS IN BOLD.

Class	Training Samples	PCA	NPE	LNSPE [7]	NWFE [2]	LFDA [9]	OTVCA [4]	3DCNN [18]	FuNet-C [19]	GiGCN [20]	DSGSF [23]	C-SS-MTr [26]	SuperPCA [30]	S^3-PCA [16]	FG_SuULDA [15]	SHDA [31]	S^3-ULDA
Asphalt	30	78.91%	82.84%	84.87%	87.11%	75.44%	89.38%	77.87%	84.37%	95.42%	94.91%	96.29%	79.64%	93.21%	91.79%	86.82%	99.33%
Meadows	30	77.24%	90.50%	91.04%	92.75%	87.93%	83.80%	78.61%	87.58%	98.56%	98.97%	89.27%	93.08%	94.65%	99.08%	98.88%	98.07%
Bitume	30	80.04%	88.84%	84.20%	88.06%	72.16%	94.39%	78.78%	76.12%	99.81%	87.63%	95.38%	97.54%	99.81%	98.84%	97.58%	98.26%
Bricks	30	95.02%	91.83%	93.24%	93.05%	95.22%	93.74%	94.86%	98.62%	90.24%	90.45%	97.62%	85.33%	93.47%	92.78%	91.23%	96.04%
Gravel	30	99.92%	100.00%	100.00%	100.00%	100.00%	100.00%	97.41%	99.77%	99.09%	99.59%	96.88%	100.00%	100.00%	100.00%	99.47%	99.85%
Meadow	30	64.69%	83.40%	89.98%	87.00%	85.10%	83.90%	62.69%	81.82%	100.00%	65.25%	98.05%	94.80%	99.68%	100.00%	97.94%	99.84%
Metal sheet	30	83.31%	90.77%	91.92%	92.85%	92.23%	92.31%	89.00%	84.15%	99.42%	70.85%	100.00%	92.38%	98.31%	100.00%	100.00%	99.38%
Shadows	30	53.48%	76.75%	86.34%	82.97%	77.33%	92.94%	79.93%	73.36%	97.23%	96.49%	96.85%	92.55%	99.48%	91.16%	98.08%	97.04%
Trees	30	99.45%	95.09%	95.20%	96.38%	91.49%	99.78%	96.73%	100.00%	81.31%	97.63%	99.79%	95.53%	98.15%	96.95%	90.51%	99.78%
OA		76.75%	87.71%	89.77%	90.50%	85.08%	87.79%	79.20%	85.95%	97.30%	89.81%	93.80%	90.96%	95.95%	96.93%	96.09%	98.38%
AA		81.34%	88.89%	90.75%	91.34%	86.32%	92.25%	83.99%	87.31%	95.83%	89.03%	96.87%	91.97%	97.42%	96.76%	95.65%	98.62%
Kappa		70.11%	83.89%	86.62%	87.48%	80.55%	84.17%	73.09%	81.66%	96.43%	86.84%	91.95%	88.16%	94.68%	95.93%	94.82%	97.86%
Time (s)		0.36	564.24	1679.26	2.26	1.41	313.95	74.05	247.25	366.24	302.82	143.47	0.50	154.02	915.10	74.38	483.51(37.39)

TABLE VI

OA, AA AND KAPPA OBTAINED BY SEVERAL FEATURE EXTRACTION ALGORITHMS (COMBINED WITH THE SVM CLASSIFIER) ON THE HOUSTON 2013 DATA SET. BEST RESULTS IN BOLD.

Class	Training Samples	PCA	NPE	LNSPE [7]	NWFE [2]	LFDA [9]	OTVCA [4]	3DCNN [18]	FuNet-C [19]	GiGCN [20]	DSGSF [23]	C-SS-MTr [26]	SuperPCA [30]	S^3-PCA [16]	FG_SuULDA [15]	SHDA [31]	S^3-ULDA
Healthy grass	30	94.43%	97.54%	97.71%	96.72%	98.36%	95.99%	97.30%	97.54%	88.01%	98.45%	79.86%	90.09%	95.09%	99.75%	85.59%	96.81%
Stressed grass	30	98.77%	98.86%	98.86%	99.10%	97.88%	99.67%	94.11%	99.59%	77.99%	85.22%	88.76%	93.22%	96.81%	87.09%	95.10%	94.69%
Synthetic grass	30	99.70%	96.85%	99.40%	97.15%	98.80%	99.40%	99.25%	99.85%	100.00%	92.97%	98.13%	99.70%	98.35%	98.85%	100.00%	99.85%
Trees	30	95.72%	98.60%	97.54%	98.93%	96.61%	94.56%	97.94%	94.23%	91.80%	98.93%	95.66%	85.83%	97.03%	94.32%	83.53%	90.28%
Soil	30	97.78%	99.34%	98.93%	99.59%	100.00%	99.09%	99.59%	98.93%	96.16%	98.95%	95.21%	97.29%	100.00%	96.53%	98.27%	
Water	30	98.64%	97.29%	100.00%	99.32%	99.32%	98.31%	90.17%	88.81%	98.15%	98.64%	89.23%	96.61%	97.63%	100.00%	97.63%	96.27%
Residential	30	84.65%	80.94%	97.97%	91.11%	89.66%	92.33%	78.59%	76.66%	93.93%	79.66%	89.20%	86.83%	88.37%	90.15%	89.58%	97.09%
Commercial	30	79.58%	69.77%	90.95%	73.64%	69.11%	73.81%	80.56%	77.10%	87.70%	82.77%	85.61%	71.58%	72.98%	79.16%	92.09%	92.67%
Road	30	76.76%	79.30%	71.83%	76.76%	86.58%	81.83%	79.21%	72.83%	91.61%	84.03%	97.44%	72.67%	77.82%	94.60%	89.85%	78.31%
Highway	30	79.11%	93.48%	83.47%	95.66%	74.60%	96.91%	92.73%	86.63%	99.43%	72.40%	97.64%	97.99%	98.33%	100.00%	98.08%	99.93%
Railway	30	78.42%	93.86%	90.48%	97.01%	81.24%	91.95%	74.11%	76.27%	88.74%	83.38%	97.33%	97.84%	94.85%	92.78%	96.59%	95.68%
Parking Lot 1	30	74.81%	93.52%	93.78%	91.77%	79.38%	92.19%	83.62%	87.86%	87.92%	80.95%	92.86%	75.31%	88.69%	82.71%	97.67%	97.51%
Parking Lot 2	30	54.44%	90.43%	96.01%	87.93%	85.65%	75.63%	93.39%	79.73%	90.19%	47.09%	85.07%	69.48%	87.24%	99.54%	94.99%	88.15%
Tennis Court	30	99.75%	99.50%	92.48%	99.25%	99.50%	100.00%	98.74%	94.97%	100.00%	95.79%	100.00%	97.99%	100.00%	100.00%	99.50%	100.00%
Running Track	30	98.57%	97.14%	98.49%	96.98%	100.00%	98.73%	100.00%	100.00%	100.00%	95.09%	97.58%	98.41%	99.52%	100.00%	100.00%	100.00%
OA		86.88%	91.45%	93.12%	92.69%	89.06%	92.34%	89.34%	87.98%	91.95%	86.03%	92.73%	87.74%	91.71%	93.35%	93.48%	94.64%
AA		87.41%	92.3%	93.96%	93.40%	90.51%	92.69%	90.63%	88.73%	93.03%	86.10%	92.89%	88.59%	92.65%	94.66%	94.45%	95.03%
Kappa		85.81%	90.76%	92.56%	92.09%	88.17%	91.72%	88.48%	91.30%	87.00%	84.89%	92.15%	86.74%	91.03%	92.82%	92.95%	94.20%
Time (s)		0.86	3081.97	9218.79	8.14	2.19	1026.13	56.99	310.68	1412.44	408.34	190.68	9.95	645.26	5407.09	290.83	2853.65(356.74)

TABLE VII

OA OBTAINED BY SEVERAL FEATURE EXTRACTION ALGORITHMS (COMBINED WITH THE SVM CLASSIFIER) WITH DIFFERENT TRAINING SAMPLES ON THE THREE CONSIDERED DATA SETS. BEST RESULTS IN BOLD.

Datasets	T.N.s/C	RAW	PCA	NPE	LNSPE [7]	NWFE [2]	LFDA [9]	OTVCA [4]	3DCNN [18]	FuNet-C [19]	GiGCN [20]	DSGSF [23]	C-SS-MTr [26]	SuperPCA [30]	S^3-PCA [16]	FG_SuULDA [15]	SHDA [31]	S^3-ULDA
Indian Pines	5	44.87%±6.50%	45.47%±5.53%	53.26%±6.42%	55.01%±4.48%	60.08%±2.68%	59.12%±3.75%	54.71%±5.58%	46.73%±3.38%	49.45%±2.63%	77.25%±3.93%	52.03%±5.87%	59.74%±3.94%	63.49%±5.77%	72.05%±5.39%	74.77%±4.31%	80.26%±3.594%	78.97%±3.86%
	10	55.77%±3.27%	55.07%±2.87%	71.44%±3.39%	73.73%±2.94%	70.03%±2.52%	64.06%±2.95%	72.95%±2.07%	56.23%±3.89%	58.80%±1.97%	82.69%±3.86%	77.65%±1.56%	74.21%±3.11%	85.75%±3.06%	87.37%±1.26%	83.20%±4.49%	87.64%±2.51%	87.67%±4.27%
	20	63.81%±3.37%	62.16%±2.64%	82.84%±1.37%	85.28%±1.42%	85.08%±1.67%	82.48%±2.02%	85.24%±2.52%	69.50%±3.22%	68.59%±2.49%	88.55%±2.05%	87.81%±1.33%	82.80%±2.45%	92.47%±1.35%	94.01%±1.64%	90.85%±1.55%	93.83%±1.46%	94.95%±1.68%
	30	68.77%±1.27%	66.24%±0.58%	86.45%±1.91%	88.78%±1.91%	88.86%±1.61%	87.58%±1.14%	88.67%±1.74%	75.33%±1.90%	73.24%±1.15%	90.38%±1.07%	90.73%±0.79%	87.27%±2.00%	94.61%±0.81%	95.79%±0.77%	93.82%±1.13%	95.41%±0.75%	96.68%±1.05%
	40	71.64%±1.05%	68.90%±1.06%	88.54%±1.32%	90.60%±1.19%	90.37%±1.48%	89.68%±0.60%	90.27%±1.12%	79.78%±2.97%	77.03%±1.04%	91.52%±1.22%	93.02%±0.69%	91.92%±0.81%	95.33%±0.97%	96.21%±0.90%	95.60%±1.33%	96.06%±5.86%	97.29%±0.72%
	50	74.18%±1.20%	70.70%±1.01%	89.95%±1.26%	91.96%±1.15%	92.71%±1.32%	90.97%±0.75%	91.80%±0.95%	81.84%±2.04%	79.08%±1.29%	91.32%±1.49%	94.50%±0.91%	92.14%±1.88%	95.42%±1.02%	96.55%±0.78%	96.64%±0.80%	96.22%±0.54%	97.75%±0.56%
	60	75.24%±1.12%	71.79%±1.33%	90.97%±0.72%	93.01%±0.85%	92.52%±0.81%	91.71%±0.91%	93.02%±0.62%	83.37%±1.99%	80.40%±1.66%	91.70%±0.94%	95.77%±0.88%	95.02%±076%	95.72%±0.56%	96.95%±0.61%	97.37%±0.66%	96.50%±0.68%	98.01%±0.51%
University of Pavia	5	64.59%±5.07%	65.26%±5.14%	68.66%±6.05%	68.26%±8.56%	72.18%±5.46%	75.27%±2.83%	68.01%±6.30%	63.91%±4.89%	69.87%±5.34%	86.72%±4.57%	59.60%±0.56%	71.67%±2.92%	74.36%±3.65%	83.70%±3.41%	76.86%±4.23%	83.05%±6.11%	87.08%±4.46%
	10	70.22%±3.05%	67.00%±3.06%	77.76%±4.24%	78.71%±6.26%	81.31%±2.27%	78.90%±2.84%	77.84%±4.09%	72.26%±3.41%	77.33%±3.45%	94.03%±2.57%	64.20%±0.43%	80.86%±3.19%	83.39%±3.33%	90.88%±2.58%	88.79%±3.23%	91.15%±2.77%	94.81%±2.82%
	20	75.85%±2.37%	75.80%±2.31%	84.70%±1.91%	87.71%±2.02%	89.14%±1.94%	81.65%±2.14%	86.85%±1.99%	75.26%±2.20%	83.61%±2.53%	96.43%±1.50%	82.68%±2.70%	90.28%±2.87%	89.34%±1.54%	95.63%±1.10%	94.56%±2.21%	94.84%±1.56%	98.14%±0.80%
	30	76.45%±1.50%	76.13%±1.85%	87.33%±1.69%	90.61%±1.41%	90.64%±1.60%	85.56%±1.62%	87.63%±1.79%	78.49%±1.32%	86.60%±1.57%	97.26%±1.18%	89.45%±0.98%	93.41%±1.26%	91.26%±0.99%	96.24%±1.33%	97.11%±1.04%	95.96%±0.94%	98.69%±0.41%
	40	77.76%±1.26%	77.41%±1.49%	89.13%±1.17%	92.50%±1.08%	91.77%±1.19%	88.27%±1.33%	90.04%±1.28%	81.17%±1.78%	88.56%±1.29%	97.53%±0.86%	94.79%±0.34%	95.45%±0.89%	92.19%±0.81%	96.72%±1.14%	98.13%±0.36%	96.48%±1.04%	98.94%±0.64%
	50	78.77%±1.31%	78.77%±1.31%	91.01%±1.46%	93.42%±0.79%	92.66%±0.65%	89.75%±1.58%	91.71%±1.19%	82.96%±1.71%	90.04%±1.10%	97.74%±0.57%	95.71%±0.23%	96.60%±0.99%	93.59%±0.90%	97.34%±0.77%	98.60%±0.37%	96.98%±0.99%	99.01%±0.61%
	60	80.52%±1.32%	79.87%±1.16%	92.05%±1.07%	94.59%±0.62%	93.62%±0.92%	91.40%±1.13%	92.99%±0.85%	84.97%±2.19%	91.40%±0.91%	97.96%±0.62%	96.19%±0.46%	96.95%±0.72%	93.99%±1.08%	97.84%±0.49%	98.95%±0.29%	97.27%±0.32%	
Houston 2013	5	65.82%±4.04%	65.92%±3.98%	65.13%±2.10%	71.63%±2.61%	73.86%±2.24%	64.69%±2.09%	67.73%±2.65%	70.41%±2.39%	70.79%±2.40%	73.72%±0.41%	56.42%±2.11%	70.61%±3.31%	58.29%±4.21%	65.44%±2.78%	73.84%±2.94%	74.29%±2.49%	74.83%±3.14%
	10	77.09%±1.90%	76.93%±1.80%	79.21%±2.06%	83.62%±1.49%	84.10%±1.72%	68.67%±2.29%	81.22%±1.85%	79.29%±2.78%	79.71%±0.80%	88.97%±1.38%	70.69%±2.19%	79.18%±2.23%	81.15%±3.07%	73.87%±1.82%	78.20%±1.80%	83.75%±1.28%	86.17%±1.06%
	20	84.32%±0.85%	84.12%±1.06%	86.98%±1.55%	90.48%±0.89%	89.73%±1.94%	85.63%±1.00%	88.64%±1.18%	85.51%±1.57%	85.07%±1.35%	87.86%±1.84%	85.05%±1.20%	89.16%±1.34%	82.12%±1.22%	87.04%±0.75%	90.48%±1.36%	89.85%±1.46%	91.33%±1.10%
	30	87.10%±0.66%	86.81%±0.87%	90.42%±1.80%	93.47%±0.59%	92.99%±0.76%	89.40%±0.86%	91.97%±0.80%	87.92%±1.64%	87.73%±0.81%	91.02%±1.11%	85.48%±1.43%	92.47%±1.35%	87.30%±0.87%	91.63%±5.74%	93.53%±0.77%	93.31%±0.61%	94.42%±0.51%
	40	89.17%±0.70%	88.86%±0.87%	93.08%±0.68%	94.71%±0.93%	93.78%±0.70%	91.69%±0.47%	93.83%±0.69%	90.12%±1.74%	89.34%±1.08%	92.51%±0.83%	87.68%±1.31%	94.35%±0.69%	89.78%±0.82%	93.49%±0.66%	95.35%±0.93%	95.01%±0.48%	95.95%±0.59%
	50	90.59%±0.75%	90.22%±0.64%	94.64%±0.71%	95.88%±0.52%	95.12%±0.55%	93.34%±0.67%	94.97%±0.67%	91.70%±1.18%	91.04%±0.70%	93.31%±0.49%	92.83%±1.15%	96.04%±0.49%	91.83%±0.52%	95.44%±0.81%	96.61%±0.68%	96.30%±0.79%	96.91%±0.60%
	60	91.55%±0.63%	91.26%±0.66%	95.23%±0.61%	96.31%±0.54%	95.45%±0.52%	94.19%±0.45%	95.72%±0.49%	92.22%±1.72%	92.11%±0.50%	93.79%±1.28%	95.77%±0.95%	97.29%±0.34%	93.22%±0.59%	96.22%±0.65%	97.28%±0.64%	97.02%±0.70%	97.60%±0.57%

samples varies. Table VII shows the classification results from different models and the proposed S^3-ULDA, where the parameters of the comparative models (such as SuperPCA, S^3-PCA, FG_SuULDA, SHDA and the deep learning models) are optimally set as indicated in the original papers. It can be seen from the results that the superpixelwise DR models (such as S^3-PCA, FG_SuULDA, SHDA and the proposed S^3-ULDA) outperform other models such as typical DR models and deep learning, especially when the number of training samples is limited. In addition, for all the comparative deep learning models we find that GiGCN based on superpixels and GCN achieves the highest OAs when training data is few (T.N.s/C =5, 10) which could demonstrate the effect of superpixels. In contrast, C-SS-MTr based on masked autoencoder and Transformer show better performance than other deep learning models when the number of training data is relatively large (T.N.s/C =50, 60). Not surprisingly, the proposed S^3-ULDA could outperform all the comparative deep learning models in all cases. More interestingly, when the training data is few (T.N.s/C =5, 10) it can be observed from Table VII that the superpixels based models such as GiGCN, S^3-PCA, FG_SuULDA, SHDA and the proposed S^3-ULDA show better performance than other models, because in the cases of few training data the feature representation based on superpixels could be more effective. Overall, the proposed S^3-ULDA outperforms superpixel-based feature extraction models and state-of-the-art deep learning models in almost every scenario, which demonstrates the effectiveness of the proposed model.

In addition to above (objective) comparison, we also subjectively compare state-of-the-art superpixel-based feature reduction models such as SuperPCA, S^3-PCA, SuperULDA (which can be seen as FG_SuULDA without 3D the flexible Gabor filter), SHDA and the proposed S^3-ULDA by means of 3D data visualization. For the two fused superpixelwise feature reduction models S^3-PCA and the proposed S^3-ULDA, we first reduce the dimensionality of the three considered HSIs to 30 and then classic PCA is performed to further reduce the 30-dimensional features to a 3D feature space for visualization purposes. All labeled data in the considered HSIs are used for creating the 3D data visualization. As it can be seen from Fig. 6, the proposed S^3-ULDA is superior to the comparative models because, in the learnt 3D feature space, the samples belonging to the same class stay closer, while data from different classes appear away from other classes (especially in the circled areas, which means the extracted low-dimensional features from the proposed S^3-ULDA are more discriminative than those from comparative superpixel-based DR models).

With the above results in mind, an additional discussion is made on the superpixel-based feature reduction models (SuperPCA, S^3-PCA, FG_SuULDA, SHDA and the proposed S^3-ULDA). From the perspective of local features, PCA and unsupervised LDA/LFDA are used in SuperPCA, S^3-PCA and the proposed S^3-ULDA for local feature extraction, while FG_SuULDA and SHDA could only extract global features via unsupervised LDA. Only S^3-PCA and the proposed S^3-ULDA could extract global-local features. However, the global features based on unsupervised LDA and the superpixel-level local features based on unsupervised LFDA from the pro-posed S^3-ULDA are more discriminative than the PCA-based global-local features from S^3-PCA. Moreover, compared to SHDA (where global and local structures are modeled by the superpixel-based global and local graphs followed by global feature extraction, in a graph embedding framework), the local feature extraction model (SuperULFDA) adopted by the proposed S^3-ULDA could extract more effective local features on a superpixel-by-superpixel basis, especially in areas with mixed objects and noise/outliers.

V. CONCLUSIONS AND FUTURE LINES

In this paper, we propose a new superpixelwise feature reduction model called S^3-ULDA, which can learn effective global-local and spectral-spatial features for accurate HSI classification. After obtaining various homogeneous superpixel blocks through ERS based superpixel segmentation technique, superpixel-based local reconstruction is conducted to denoise HSI. Then, a new SuperULDA algorithm is proposed to extract global features based on both the original HSI and locally reconstructed data, as well as the pseudo-labels of superpixels provided by ERS. By simultaneously using the two data sources, the extracted global features are more discriminative. To further extract effective local features, a new SuperULFDA algorithm is conducted locally on each superpixel (along with its adjacent superpixels) as well as the corresponding pseudo-labels, obtaining more effective local features than PCA-based local feature extraction in SuperPCA, S^3-PCA and unsupervised LDA-based local feature representation in SHDA. This is because LFDA can effectively deal with the multimodality of adjacent superpixels and address the issue that LDA only focuses on the global structures present in the HSI. Finally, by simply fusing the obtained (global and local) features, global-local and spectral-spatial features are obtained. The experimental results demonstrate that these features are more effective for classification purposes than those obtained by other state-of-the-art methods.

In future work, we will deal with how to automatically determine the optimal number of superpixels for the ERS based segmentation technique. In addition, PCA-based local features could be fused with the proposed LDA-based features to further improve the performance of feature extraction.

REFERENCES

[1] L. He, J. Li, C. Liu, and S. Li, "Recent Advances on Spectral–Spatial Hyperspectral Image Classification: An Overview and New Guidelines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 3, pp. 1579–1597, 2018.

[2] X. Jia, B.-C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 676–697, 2013.

[3] Y. Zhang, X. Wang, X. Jiang, and Y. Zhou, "Marginalized graph self-representation for unsupervised hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[4] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral feature extraction using total variation component analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 6976–6985, 2016.

[5] Y. Wang, T. Li, L. Chen, Y. Yu, Y. Zhao, and J. Zhou, "Tensor-based robust principal component analysis with locality preserving graph and frontal slice sparsity for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19, 2022.
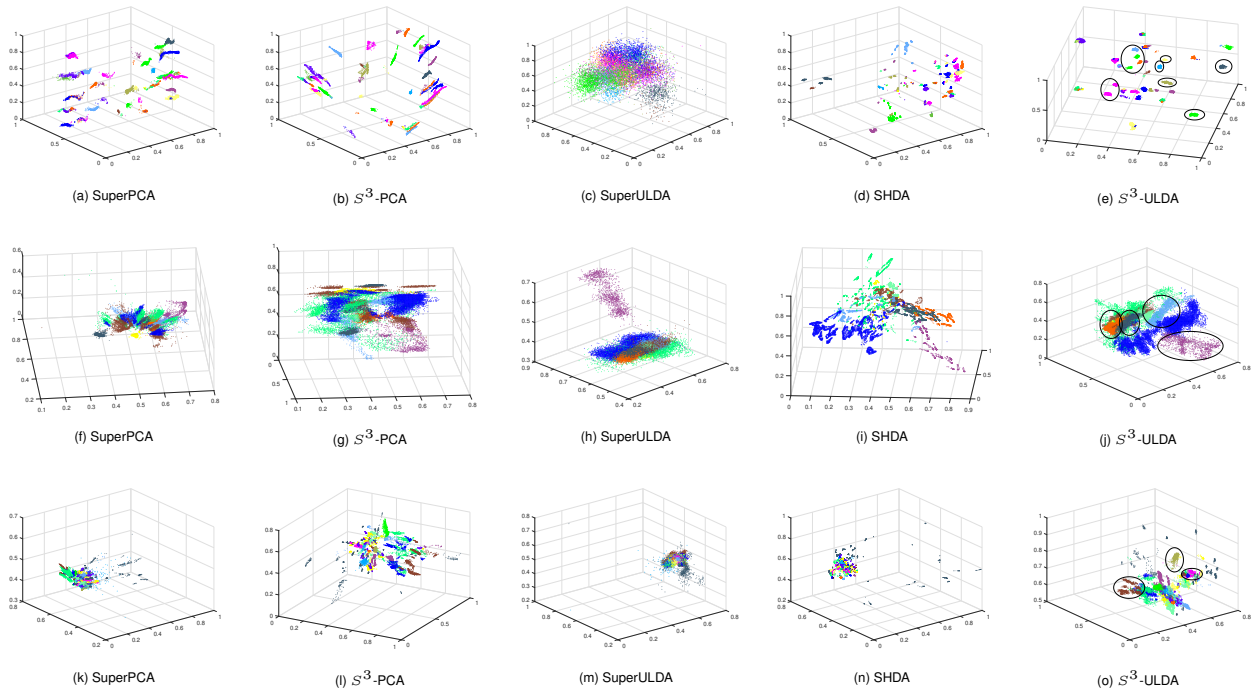
Fig. 6. 3D data visualization of Indian Pines(a-e), University of Pavia(f-j), and Houston 2013(k-o) features obtained using SuperPCA, S^3-PCA, SuperULDA, SHDA and S^3-ULDA.

[6] Y. Zhou, J. Peng, and C. L. P. Chen, "Dimension Reduction Using Spatial and Spectral Regularized Local Discriminant Embedding for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 1082–1095, 2015.

[7] G. Shi, H. Huang, and L. Wang, "Unsupervised Dimensionality Reduction for Hyperspectral Imagery via Local Geometric Structure Feature Learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1425–1429, 2020.

[8] 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 Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[9] M. Sugiyama, "Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis," *Journal of Machine Learning Research*, vol. 8, p. 1027–1061, 2007.

[10] F. Luo, Z. Zou, J. Liu, and Z. Lin, "Dimensionality reduction and classification of hyperspectral image via multistructure unified discriminative embedding," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[11] H. Xu, H. Zhang, W. He, and L. Zhang, "Superpixel-based spatial-spectral dimension reduction for hyperspectral imagery classification," *Neurocomputing*, vol. 360, pp. 138–150, 2019.

[12] Z. Ye, R. Dong, L. Bai, and Y. Nian, "Adaptive collaborative graph for discriminant analysis of hyperspectral imagery," *European Journal of Remote Sensing*, vol. 53, no. 1, pp. 91–103, 2020.

[13] F. Luo, H. Huang, Z. Ma, and J. Liu, "Semisupervised sparse manifold discriminative analysis for feature extraction of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6197–6211, 2016.

[14] J. Jiang, J. Ma, Z. Wang, C. Chen, and X. Liu, "Hyperspectral Image Classification in the Presence of Noisy Labels," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 851–865, 2019.

[15] S. Jia, Q. Zhao, J. Zhuang, D. Tang, Y. Long, M. Xu, J. Zhou, and Q. Li, "Flexible Gabor-Based Superpixel-Level Unsupervised LDA for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 12, pp. 10 394–10 409, 2021.

[16] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–10, 2022.

[17] J.-H. R. Chang and Y.-C. F. Wang, "Propagated image filtering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 10–18.

[18] A. Ben Hamida, A. Benoit, P. Lambert, and C. Ben Amar, "3-d deep learning approach for remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4420–4434, 2018.

[19] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5966–5978, 2021.

[20] S. Jia, S. Jiang, S. Zhang, M. Xu, and X. Jia, "Graph-in-graph convolutional network for hyperspectral image classification," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2022.

[21] Y. Duan, F. Luo, M. Fu, Y. Niu, and X. Gong, "Classification via structure-preserved hypergraph convolution network for hyperspectral image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023.

[22] H. Zhou, F. Luo, H. Zhuang, Z. Weng, X. Gong, and Z. Lin, "Attention multihop graph and multiscale convolutional fusion network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[23] T. Guo, R. Wang, F. Luo, X. Gong, L. Zhang, and X. Gao, "Dual-view spectral and global spatial feature fusion network for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023.

[24] C. Cheng, H. Li, J. Peng, W. Cui, and L. Zhang, "Deep high-order tensor convolutional sparse coding for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.

[25] D. Ibañez, R. Fernandez-Beltran, F. Pla, and N. Yokoya, "Masked auto-encoding spectral–spatial transformer for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[26] L. Huang, Y. Chen, and X. He, "Spectral–spatial masked transformer with supervised and contrastive learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–18, 2023.

[27] H. Zhao, F. Zhou, L. Bruzzone, R. Guan, and C. Yang, "Superpixel-Level Global and Local Similarity Graph-Based Clustering for Large Hyperspectral Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[28] S. Subudhi, R. N. Patro, P. K. Biswal, and F. Dell'Acqua, "A Survey on Superpixel Segmentation as a Preprocessing Step in Hyperspectral Image Analysis," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 5015–5035, 2021.

[29] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2097–2104.

[30] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A Superpixelwise PCA Approach for Unsupervised Feature Extraction of Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4581–4593, 2018.

[31] S. Zhang, T. Lu, W. Fu, and S. Li, "Superpixel-level hybrid discriminant analysis for hyperspectral image feature extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

Pengyu Lu is the undergraduate student in China University of Geosciences, Wuhan, China, where he is pursuing the B.Eng. degree in Computer Science. His research interests include machine learning and hyperspectral image processing.

Xinwei Jiang received the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2012. He is an Associate Professor with the China University of Geosciences, Wuhan. His research interests include nonparametric statistical models, dimensionality reduction and hyperspectral image processing.
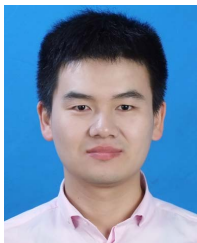
Junjun Jiang (Member, IEEE) received the B.S. degree from the Department of Mathematics, Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree from the School of Computer, Wuhan University, Wuhan, China, in 2014. From 2015 to 2018, he was an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan. From 2016 to 2018, he was a Project Researcher with the National Institute of Informatics, Tokyo, Japan. He is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. His interests mainly focus on computer vision, image processing, and hyperspectral image analysis.

Yongshan Zhang received her bachelor and Ph.D degree in Computer Science in 2014 and 2019 from China University of Geosciences, Wuhan, China, where she is an Associate Professor. She was a visiting student in the Department of Computer Science, University of Illinois at Chicago. Her research focuses on data mining and machine learning.

Xiaobo Liu (Member, IEEE) received the M.S. degree in computer science and the Ph.D. degree in geosciences information engineering from the China University of Geosciences, Wuhan, China, in 2008 and 2012, respectively. He is an Associate Professor with the School of Automation, China University of Geosciences. His research interests include machine learning, evolutionary computation, and hyperspectral remote sensing image processing.

Zhihua Cai received the B.S. degree from Wuhan University, Wuhan, China, in 1986, the M.Sc. degree from the Beijing University of Technology, Beijing, China, in 1992, and the Ph.D. degree from the China University of Geosciences, Wuhan, in 2003. He is a Faculty Member with the School of Computer Science, China University of Geosciences. He has published over 50 research articles in journals and conferences. His main research areas include data mining, machine learning, evolutionary computation, and their applications.

Antonio Plaza (Fellow, IEEE) is a Full Professor and the Head of the Hyperspectral Computing Laboratory at the Department of Technology of Computers and Communications, University of Extremadura, where he received the M.Sc. degree in 1999 and the PhD degree in 2002, both in Computer Engineering. His main research interests comprise hyperspectral data processing and parallel computing of remote sensing data. He has authored more than 800 publications in this field, including 405 JCR journal papers, 25 book chapters, and 330 peer-reviewed conference proceeding papers. He has guest edited 17 special issues on hyperspectral remote sensing for different journals. Prof. Plaza is a Fellow of IEEE "for contributions to hyperspectral data processing and parallel computing of Earth observation data," a Fellow of the Asia-Pacific Artificial Intelligence Association (AAIA) and an elected member of Academia Europaea, The Academy of Europe. He is a recipient of the recognition of Best Reviewers of the IEEE Geoscience and Remote Sensing Letters (in 2009) and a recipient of the recognition of Best Reviewers of the IEEE Transactions on Geoscience and Remote Sensing (in 2010), for which he served as Associate Editor in 2007-2012. He is also an Associate Editor for IEEE Access (receiving the recognition of Outstanding Associate Editor for the journal in 2017), and was also member of the Editorial Board of the IEEE Geoscience and Remote Sensing Newsletter (2011-2012) and the IEEE Geoscience and Remote Sensing Magazine (2013). He was also a member of the steering committee of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS). He is a recipient of the Best Column Award of the IEEE Signal Processing Magazine in 2015, the 2013 Best Paper Award of the JSTARS journal, and the most highly cited paper (2005-2010) in the Journal of Parallel and Distributed Computing. He received best paper awards at the IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, the IEEE International Conference on Space Technology, and the IEEE Symposium on Signal Processing and Information Technology. He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011-2012, and as President of the Spanish Chapter of IEEE GRSS in 2012-2016. He is currently serving as Chair of the Publications Awards Committee of IEEE GRSS and as a Vice-Chair of the Fellow Evaluations Committee of IEEE GRSS. He has reviewed more than 500 manuscripts for over 50 different journals. He served as the Editor-in-Chief of the IEEE Transactions on Geoscience and Remote Sensing journal for five years (2013-2017) and as the Editor-in-Chief of the IEEE Journal on Miniaturization for air and Space Systems (2019-2020). He has been included in the 2018, 2019, 2020, 2021 and 2022 Highly Cited Researchers List (Clarivate Analytics). Additional information: http://sites.google.com/view/antonioplaza.