A novel band selection and spatial noise reduction method for hyperspectral image classification.

FU, H., ZHANG, A., SUN, G., REN, J., JIA, X., PAN, Z. and MA, H.

2022

A novel band selection and spatial noise reduction method for hyperspectral image classification

Hang Fu, Aizhu Zhang, *Member, IEEE* , Genyun Sun, *Member, IEEE*, Jinchang Ren, *Senior Member, IEEE*, Xiuping Jia, *Fellow, IEEE*, Zhaojie Pan, and Hongzhang Ma

Abstract—As an essential reprocessing method, dimensionality reduction (DR) can reduce the data redundancy and improve the performance of hyperspectral image (HSI) classification. A novel unsupervised DR framework with feature interpretability, which integrates both band selection (BS) and spatial noise reduction method, is proposed to extract low-dimensional spectral-spatial features of HSI. We proposed a new Neighboring band Grouping and Normalized Matching Filter (NGNMF) for BS, which can reduce the data dimension whilst preserve the corresponding spectral information. An enhanced 2-D singular spectrum analysis (E2DSSA) method is also proposed to extract the spatial context and structural information from each selected band, aiming to decrease the intra-class variability and reduce the effect of noise in the spatial domain. The support vector machine (SVM) classifier is used to evaluate the effectiveness of the extracted spectral-spatial low-dimensional features. Experimental results on three publicly available HSI datasets have fully demonstrated the efficacy of the proposed NGNMF-E2DSSA method, which has surpassed a number of state-of-the-art DR methods.

Index Terms—Hyperspectral image (HSI), dimensionality reduction, band selection, enhanced 2DSSA, image classification.

I. INTRODUCTION

HYPERSPECTRAL image (HSI), with a three-dimensional data structure as a hypercube, has abundant spectral and spatial information [1]. Covering the wavelengths from the visible light to infrared, the spectral profile can be used to identify the changing properties of the objects in terms of moisture, temperature, and chemical components [2]. For each spectral band, the corresponding grayscale image formed by the spectral reflectance response contains certain spatial structure information of the objects. These characteristics have enabled HSI to be widely applied in urban mapping [3], land cover classification [4], mineral detection [5], and precision agriculture [6]. However, two we face extra challenges when dealing with HSI for remote sensing applications [7].

The classification accuracy of HSI can be easily affected by the Hugh's phenomenon, i.e. less samples than the number of spectral bands [8, 9]. To tackle this issue, a number of dimensionality reduction (DR) methods have been proposed [10-12], which can in general divided into two categories: feature extraction (FE) and band selection (BS) based ones.

In FE methods, the original high-dimensional data are often transferred into a lower-dimensional representation to reveal the associated distinctive properties. Typical examples include the principal component analysis (PCA) and its variations [13], and linear discriminant analysis (LDA) [14]. In order to further extract the nonlinear structure of the HSI, manifold learning methods have been introduced for improving the classification accuracy, such as the local preserving projection (LPP) [15], local Fisher discriminant analysis (LFDA) [16], and spatial-spectral manifold reconstruction preserving embedding (SSMRPE) [17]. However, due to the lack of spectral characteristic analysis of the ground objects, the feature-space metrics obtained after DR has no real meaning of the associated physical and/or chemical properties [18, 19].

BS aims to select a subset of the original spectral bands, based on certain predefined criteria. Conventional BS methods can be divided into two main groups: i.e. ranking-based and clustering-based methods, yet they both unfortunately suffer from some drawbacks. For ranking-based methods, the correlation among the selected representative bands can be quite high. While clustering-based methods can reduce data correlation and redundancy, most of them are sensitive to noisy bands. Therefore, the combination of these two is found to be particular useful. In [20], an enhanced fast-peak-based clustering (E-FDPC) method was proposed, in which local density and intra-cluster distance were combined for band ranking. Sun *et al.* [21] proposed an adaptive distance-based band hierarchy (ADBH) clustering with E-FDPC for ranking. Wang *et al.* [22] proposed a fast neighborhood grouping BS method (FNGBS), in which a coarse-to-fine grouping strategy was used for band clustering with reduced redundancy, whilst the product of local density and information entropy for band ranking. Compared with feature extraction, band selection can better preserve the spectral information of the original data, i.e. clear interpretability and representability of the original data, along with a limited classification improvement [3, 23].

Another important factor that affects the accuracy of HSI classification is the spectral variability, which is often degraded by the improvement of the spatial resolution [8]. Specifically, rich information provided by high-resolution images may increase the intra-class variability and decrease the interclass variability [24], leading to poor classification performance.

To alleviate this issue, contextual and structural information from the spatial domain are needed to tackle the ambiguity in the spectral domain, leading to a series of spatial feature extraction methods. As one of the classic spatial methods, the extended morphological attribute profiles (EMAPs) [25] were used to extract multilevel spatial structural features with a series of attribute profiles, providing a rich description of the scene. In [26], the Gabor filters at different scales and orientations were used to represent the abundant spatial texture features of HSI. In addition, some edge-preserving filtering based methods, such as bilateral filtering (BF) [27] and domain transform recursive filtering (DTRF) [28], have been found effective in extracting spatial context information, which can smooth out texture and noise while preserving spatial details. Zabalza *et al*. [29] proposed an interesting 2D singular spectrum analysis (2DSSA) method, decomposing each band image into varying spatial feature components. By selecting certain components as spatial features, 2DSSA can achieve state-of-the-art classification performance. However, it still has some limitations, such as feature ambiguity and loss of structural information [30, 31].

In order to address both of these issues above, some spectral-spatial methods, combining FE/BS with spatial feature extraction methods, were further proposed for effective feature extraction. Kang *et.al*. [32-34] proposed a few spectral and spatial feature integration frameworks, in which the DR approach (e.g., PCA, band averaging) was utilized as pre-processing, followed by the subsequent spatial technique (e.g., intrinsic image decomposition, Gaussian pyramid and edge-preserving filtering) for extracting spectral-spatial features. In [35], a novel ICA and EMAP based spectral-spatial DR method was proposed, in which spectral features extracted by ICA was suitable for representing each class, whilst the improved and reduced AP (rAP) could further address the spatial redundancy of EMAP. In [36], the guided filter (GF) was used to extract spatial context information, with the LFDA for low-dimensional embedding. Zheng *et.al* [18] performed graph-based feature extraction on the selected band subset, which achieved superior classification results yet with a high computational cost. Although being successful in reported work, there are still some shortcomings in these methods. First of all, the interpretability of the extracted spectral-spatial features decreases after applying various transformation techniques. In addition, some spatial methods need quite a few parameters for spatial processing, which again increases the feature dimensionality and calculation cost to some extent.

In recent years, deep learning techniques [37, 38] have been widely used, including hyperspectral image classification, due to their strong ability to mine high-level spatially invariant and discriminant features [39, 40]. In some typical models [41-43], feature transformation or feature selection for hyperspectral pre-processing is used, followed by a deep network, e.g. the convolutional neural network, (CNN), to further extract robust features before data classification [8, 44]. He *et al*. [45] used the maximum noise fraction (MNF) for DR and covariance features construction, where a classical 2-D CNN (2DCNN) model was adopted for HSI classification. In [46], the proposed adaptive DR (ADR) was used to reduce the high dimensionality along with a semi-supervised 3-D deep neural network to further extract the spatial-spectral features. Unfortunately, these deep networks still face serious problems, such as time-consuming and error-prone process, huge hyper-parameters, high computational cost, and more importantly, low interpretability of the extracted deep features compared with the handcrafted features [47].

The key to improve the classification accuracy of HSI is the extraction of significant features [30], which usually show several characteristics, including high interpretability, low redundancy, robustness to noise, and strong land cover separability. It is thus essential to investigate how to extract such features from HSIs within a simple yet effective framework.

To this end, we propose in this paper a novel unsupervised DR framework that combine BS and 2DSSA based spatial methods to extract significant features of HSI. Firstly, we designed an effective BS method, namely neighborhood grouping normalized matched filter (NGNMF), by combining clustering and ranking approaches to fully reduce the spectral redundancy. Then, an improved enhanced 2DSSA, E2DSSA, was proposed and applied to each selected band to extract significant spatial context features while eliminating the effect of noise within the scene. Finally, the support vector machine (SVM) classifier was applied for classification to evaluate the efficacy of the extracted features.

The major contributions of our paper are highlighted below:

1) A DR framework, NGNMF-E2DSSA, is proposed for significant low-dimensional spectral-spatial feature extraction in HSIs. By combining NGNMF based band selection and E2DSSA based spatial processing, DR and feature extraction are carried out simultaneously, benefiting from high efficiency, low spectral redundancy, noise robustness, and feature efficacy.

2) With a new band grouping strategy to reduce the spectral redundancy of the raw data, and an improved MF-based ranking strategy to further eliminate the effect of noisy bands, NGNMF is able to obtain the most representative spectral bands with assured quality of band selection.

3) By introducing local spatial similarity-based adaptive embedding to construct a low-rank trajectory matrix, the proposed E2DSSA can extract more effective spatial context and structural features in terms of improved intra-class similarity and inter-class variability of land covers, which appears to be superior to most existing techniques.

4) Experiments on three datasets demonstrate that NGNMF-E2DSSA achieves superior classification performance to several state-of-the-art DR methods and deep learning methods, even with a small number of training samples.

The remainder of this paper is organized as follows. Section II introduces the proposed NGNMF-E2DSSA method. In Section III, we discussed and analyzed the experimental results. Finally, some concluding remarks are provided in Section IV.
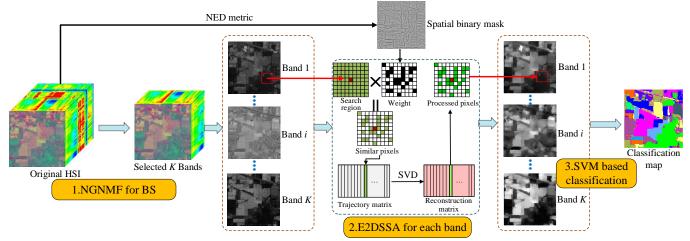
Fig. 1. Flowchart of the proposed NGNMF-E2DSSA method for HSI classification.

II. Proposed Methodology

The workflow of the proposed NGNMF-E2DSSA method is shown in Fig. 1, which contains three main steps as detailed below. First, NGNMF is used for DR and spectral feature extraction on the original HSI data. Second, E2DSSA is applied to extract and enhance the spatial contextual features on each selected band in order to generate low-dimensional spectral-spatial features. Third, the resulting spectral-spatial features are classified using the SVM classifier. In terms of notations, we use italic letters to denote scalars (e.g., a and A), boldface lowercase letters for vectors (e.g., $\mathbf{a}$), boldface capital letters for matrices (e.g., $\mathbf{A}$), and calligraphic letters for tensors (e.g., $\mathcal{A}$). The input HSI data is denoted as $\mathcal{X} \in \mathbb{R}^{W \times H \times B}$, in which W, H, and B denote the row, column and number of bands, respectively.

A. NGNMF for band selection

The BS method can effectively reduce the dimension of a HSI whilst preserving the relevant information of the hypercube. To tackle the noise sensitivity of clustering based methods and the band redundancy of ranking based methods, we propose NGNMF, which combines clustering and ranking together to select bands of high significance from the HSI. NGNMF consists of two main steps: i.e. i) grouping of neighboring bands, and ii) band ranking with the normalization matched filter (MF). Detailed implementation of these two steps are described below.

1) Grouping of neighboring bands: The B bands of $\mathcal{X}$ are first divided equally into K ($K<B$) groups, and each band group $\mathbf{X}_k$ ($k=1,\ldots,K$) is given by:

$$\mathbf{X}_k = \begin{cases} \{\mathbf{X}^{(k-1)[B/K]+1},...,\mathbf{X}^{k[B/K]}\} & k[B/K] \leqslant B \\ \{\mathbf{X}^{(k-1)[B/K]+1}, \ \ldots \ ,\mathbf{X}^{B}\} & k[B/K] > B \end{cases} \quad (1)$$

where $\mathbf{X}^{(i)}$ represents the ith band image of $\mathcal{X}$, and $[B/K]$ is the smallest integer no less than B/K. Then, the fine partition algorithm [22] is performed on the initial band group $\mathbf{X}_k$ ($k=1,\ldots,K$) to obtain a new band group $\mathbf{X}_k'$ ($k=1,\ldots,K$) below, where the number of bands in each group is no longer the same.

$$\mathcal{X} = \{\mathbf{X}_1',...,\mathbf{X}_k',...,\mathbf{X}_K'\}, \quad \mathbf{X}_k' \in \mathbb{R}^{W \times H \times B_k} \quad (2)$$

where B_k represents the number of bands in the new group $\mathbf{X}_k'$. After the fine band partition, highly correlated spectral bands

will be grouped and the correlation between band groups is low. As a result, it is also more reasonable to select representative bands from each band group.

2) Band ranking with the normalization matched filter (MF): During the selection of representative bands, water absorption and/or low signal-to-noise bands may be incorrectly selected. Inspired by the MF ranking based bad band removal [48], a modified band normalization MF ranking strategy is developed for more effective section of the representative bands as follows.

Firstly, a band normalization is applied to each band group $\mathbf{X}_k'$, and the normalization range is [0, 1]. Compared with the normalization using a normalization matrix in [48], which significantly reduces the band difference, this band normalization is more conducive to band distinction. Each subset $\mathbf{X}_k'$ is then converted into a two-dimensional matrix $\{x_1^k,...,x_i^k,...,x_{WH}^k\} \in \mathbb{R}^{WH \times B_k}$, in which x_i^k represents the ith pixel vector with a dimension of B_k. Each pixel x_i^k is selected as the target pixel, and its weight w_i^k can be obtained by the MF detector as:

$$w_i^k = \kappa \mathbf{C}^{-1}(x_i^k - m^k) \in \mathbb{R}^{B_k \times 1}, (i=1,...,WH; \ k=1,...,K) \quad (3)$$

$$\kappa = \frac{1}{(x_i^k - m)^T \mathbf{C}^{-1}(x_i^k - m)} \quad (4)$$

where m^k is the mean of all x_i^k. $\mathbf{C} \in \mathbb{R}^{B_k \times B_k}$ is the covariance matrix, and κ is a normalization constant.

Finally, we calculate the average of the absolute weight vectors to obtain the weights of each band below:

$$|w^k|_{mean} = \frac{1}{WH} \sum_{i=1}^{WH} |w_i^k| \in \mathbb{R}^{B_k \times 1} \quad (5)$$

where $|w^k|_{mean}$ represents the weight vector corresponding to the band subset $\mathbf{X}_k'$. The bands with larger weights are assumed to have higher signal-to-noise ratios and image quality, and smaller weights indicate a potential bad bands. Therefore, the band with the maximum weight is selected as the most representative band.

By selecting a representative band in each group, we can obtain a spectral band set $\mathcal{Y}$ with a dimension K.

$$\mathcal{Y} = \{\mathbf{Y}^1,...,\mathbf{Y}^i,...,\mathbf{Y}^K\} \in \mathbb{R}^{W \times H \times K} \quad (6)$$

where $\mathbf{Y}^i$ represents the selected bands from the ith band group. The low dimensional spectral feature $\mathcal{Y}$ is then taken as the input to E2DSSA for further refinement.
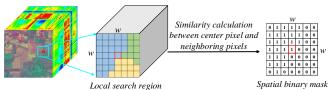
Fig. 2. Spatial binary mask generation using NED metrics. Each pixel corresponds to a local search area, and these search regions overlap.
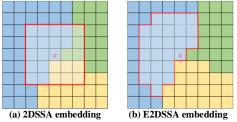

(a) 2DSSA embedding   **(b) E2DSSA embedding**

Fig. 3. The regular embedding window in 2DSSA (a), and adaptive embedding in E2DSSA (b) with the same number of embedding pixels.

B. E2DSSA for spatial feature extraction

After BS, the remaining K bands contain major spectral features yet with a lower dimension. To effectively extract spectral-spatial joint features, E2DSSA is applied to $\mathcal{Y}$ to extract the spatial context features in two steps, i.e. i) spatial binary mask generation based on spectral similarity; and ii) band-based spatial processing guided by a spatial binary mask.

1) Spatial binary mask generation: From the original hypercube, a spatial binary mask is generated according to the similarity between the center pixel and the neighboring pixels. Taking each spectral pixel $\mathbf{x}_i \in \mathbb{R}^{B \times 1}$ ($i = 1, \ldots, WH$) as the central pixel, a local search window with a size of $w \times w$ can be determined. The similarity of pixels in the search window to the central one can be calculated using the normalized Euclidean distance (NED) [49]. The similarity of any two spectral pixel vectors is given by:

$$s_{ij} = \sqrt{\sum \left(\frac{\mathbf{x}_j}{\|\mathbf{x}_j\|_2} - \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|_2} \right)^2} \qquad (7)$$

where s_{ij} represents the similarity of pixels $\mathbf{x}_i$ and $\mathbf{x}_j$, and $\|\cdot\|_2$ represents the 2-norm of the spectral vector.

By initializing a mask of size $w \times w$ and defining the number of embedded pixels to be L ($L < w \times w$), we set the position of the first L highly similar pixels, according to s_{ij}, including the center pixel, to 1, and the rest to 0. In this way we can obtain a spatial binary mask corresponding to each pixel, as shown in Fig. 2. Note that the search window size $w \times w$ is set to 17×17 to ensure balance between the efficiency and efficacy in selection of qualified similar pixels [50, 51]. As the search windows can be overlapped to each other, the pixel mirror filling of the original image is needed to meet the construction of the search window of the edge pixels.

2) band-based spatial processing: In this step, the spatial binary mask obtained above is used to guide the E2DSSA to extract features in each band in four steps, i.e. adaptive embedding, singular value decomposition (SVD), grouping, and reprojection as detailed below.

a. Adaptive embedding: For each band image $\mathbf{Y}^i$, each pixel and its (L-1) similar neighbors guided by the binary mask are selected to build the trajectory matrix. Compared with 2DSSA that uses a fixed window (Fig. 3a), our method can more
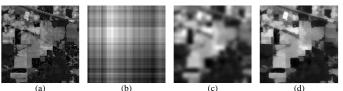

Fig. 4. (a) Test image. Feature image corresponding to maximum singular value of (b) SVD, (c) 2DSSA, and (d) E2DSSA.

effectively utilize local spatial information (Fig. 3b). The L pixels corresponding to each central pixel are expanded into a column vector, and all the column vectors are arranged to a matrix, i.e., a trajectory matrix $\mathbf{M}^i \in \mathbb{R}^{L \times WH}$. Each band image can obtain a corresponding trajectory matrix, thus the resulting trajectory matrix contains both global and local spatial features of the image for enhanced image characterization [52]. The matrix $\mathbf{M}^i$ has several characteristics: First of all, it is not strictly a Hankel-block-Hankel (HbH) matrix compared with the trajectory matrix of 2DSSA, that is, the elements on the anti-diagonal line of the matrix are not exactly the same. In addition, $\mathbf{M}^i$ has pixel similarity in column direction and high correlation in row direction, i.e. a low-rank characteristic.

b. Singular value decomposition (SVD): SVD is applied to the trajectory matrix $\mathbf{M}^i$, a process equivalent to calculating the eigenvalues ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$) and their corresponding eigenvectors ($u_1, u_2, \ldots u_L$) of $\mathbf{M}^i \mathbf{M}^{iT}$. Therefore, the trajectory matrix $\mathbf{M}^i$ is decomposed by:

$$\mathbf{M}^i = \mathbf{M}_1^i + \cdots + \mathbf{M}_j^i + \cdots + \mathbf{M}_L^i, \; j = 1, \ldots, L \qquad (8)$$

each submatrix $\mathbf{M}_j^i$ can be defined as:

$$\mathbf{M}_j^i = \sqrt{\lambda_j} \, \mathbf{u}_j \mathbf{v}_j, \; \mathbf{v}_j = (\mathbf{M}^i)^T \mathbf{u}_j / \sqrt{\lambda_j}, \; j = 1, \ldots, L \qquad (9)$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are denoted as the left and right singular vectors, respectively, and $\sqrt{\lambda_i}$ is the singular value of the trajectory matrix. Different submatrices $\mathbf{M}_j^i$ usually contain different spatial structure contents, including the trend, texture, edge and noise, etc.

c. Grouping: In order to retain the main information and remove the noise or insignificant components, in this step, we select the submatrix corresponding to the larger singular value for reconstruction [29-31]. As the trajectory matrix $\mathbf{M}^i$ has a low rank (or rank-1) characteristic, we choose $\mathbf{M}_1^i$ to approximate the matrix $\mathbf{M}^i$ for efficiency.

d. Reprojection: The selected submatrix $\mathbf{M}_1^i$ is reprojected to the image size as the refined feature image. For each column of the matrix $\mathbf{M}_1^i$, these processed pixels are returned to their original position. This requires to record the position of similar pixels according to each central pixel in the previous steps. For multiple pixel values at the same position, the average value of these pixels is used at the position. Homogeneous regions in the image include more similar pixels, and the regional consistency is enhanced after the averaging. While in the heterogeneous areas (e.g., edges), there are fewer similar pixels to be averaged, pixel differences are preserved.

Compared with 2DSSA, the main advantage of E2DSSA is the construction of a low rank trajectory matrix based on adaptive embedding. In the low rank trajectory matrix, the pixels in the column direction have high similarity and the

Fig. 5. Indian Pines. (a) False-color image; (b) Ground truth image.

Fig. 6. PaviaU. (a) False-color image; (b) Ground truth image.

Fig. 7. Salinas. (a) False-color image; (b) Ground truth image.

columns have a high correlation. This low-rank feature makes most of the spatial information after SVD decomposition concentrate on the first component, which is crucial for the distinction of ground objects.

The spatial feature maps of E2DSSA, as well as SVD and 2DSSA are compared in Fig. 4. Compared with SVD, the feature maps structured by 2DSSA and E2DSSA improve the continuity of land covers, mainly because the trajectory matrix retains the local spatial information. Moreover, the feature map of E2DSSA contains more edges and textures features, which further improves the distinguishability of different land covers.

C. SVM based classification

The obtained low-dimensional spectral-spatial features $\mathcal{Z}$ after E2DSSA can be represented as:

$$\mathcal{Z} = \{\mathbf{Z}^1,...,\mathbf{Z}^i,...,\mathbf{Z}^K\} \in \mathbb{R}^{W \times H \times K} \qquad (10)$$

where $\mathbf{Z}^i$ represents the feature band image of $\mathbf{Y}^i$.

The spectral direction of feature $\mathcal{Z}$ contains the main spectral identification information for distinguishing the ground objects. While in the spatial direction, it retains the spatial context and structure information of the original band, where the effects of noise are removed whilst improving the intra-class

consistency. As a result, the low-dimensional feature $\mathcal{Z}$ has an improved interpretability.

In addition, the choice of a classifier is crucial for the performance evaluation of the obtained features, especially with a limited number of labelled training samples in HSI. Among many classifiers, the support vector machine (SVM) has achieved excellent performance in HSI classification due to its applicability to small samples and robustness to the data dimensionality [53,54]. Therefore, we chose the SVM to perform the classification using the obtained low-dimensional features as input, and the results are given in the next section.

III. EXPERIMENTAL SETUP AND RESULTS

In this section, we described the details of three benchmark HSI datasets and experimental settings. The effectiveness of NGNMF and E2DSSA were validated respectively. Finally, we compared the proposed framework with several state-of-the-art methods using small random training samples, as well as deep learning methods with spatial disjoint samples.

A. Datasets description

The three hyperspectral datasets utilized in our experiments include Indian Pines, Pavia University (PaviaU), and Salinas. The Indian Pines dataset was gathered by the AVIRIS sensor at the Indian Pines test site in Northwest Indiana, USA. The spatial size of this data is 145×l45 pixels with a spatial resolution of 20m, and these pixels are divided into 16 classes. It has 220 spectral bands covering the range from 0.4 to 2.5μm. In this experiment, a total of 200 bands were used after removing 20 water absorption bands (104-108, 150-163 and 220). More detailed information is given in Fig. 5, which shows the false-color image with corresponding ground truth, as well as the name and the number of samples for each class.

The PaviaU dataset was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor, which has a spatial size of 610×340 pixels and 115 bands with a wavelength range of 0.43–0.86μm. In this experiment, the number of bands was reduced to 103 by removing 12 noisy bands. Nine land cover categories are used in this dataset. The false-color image with corresponding ground truth is shown in Fig. 6, and the sample size of each class is indicated by values in brackets.

The Salinas dataset was also collected by the AVIRIS sensor over an agricultural area of Salinas Valley, California. With a spatial size of 512×217 pixels, this dataset contains 224 bands, presenting a spatial resolution of 3.7m. Again, 20 water absorption bands were removed and the remaining 204 bands were used for classification. In Fig. 7, the false color image and corresponding ground reference map are shown, as well as the detailed information on 16 land cover categories.

B. Experimental setup

In our experiment, two types of sampling strategies are used to generate the training set, i.e. random sampling and spatially disjoint sampling, which are mainly used in subsections C-E and F, respectively. In order to avoid systematic errors and reduce random discrepancies, all experiments were carried out ten times independently, aiming to obtain a better presentation of results. As for the classifier, the kernel SVM to classify the data was implemented using the LIBSVM library [55]. An RBF
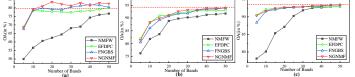
Fig. 8. OA comparisons of different BS methods on (a) Indian Pines, (b) PaviaU, and (c) Salinas.
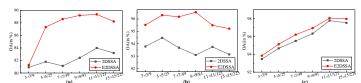

Fig. 9. Parameters comparisons between E2DSSA and 2DSSA on (a) Indian Pines, (b) PaviaU, and (c) Salinas.

kernel was adopted with a ten-fold cross-validation for tuning the parameter. In addition, five metrics, including class-by-class accuracy, overall accuracy (OA), average accuracy (AA), kappa coefficient (kappa), and computation time are used for the assessment of different methods' performance. All the experiments were performed on a computer with a 3.5-GHz CPU, 8-GB memory and 64-bit Windows 10 using MATLAB 2017a.

C. Analysis of NGNMF

In this subsection, classification accuracy is used to evaluate the performance of the selected band subsets on the three datasets. The proposed NGNMF was compared with the classical EFDPC [20], state-of-the-art NMFW [48], and FNGBS [22] methods. NMFW acts directly on the original HSI data instead of combining it with other BS methods as in [46] in our experiment. In addition, the number of selected bands varies from 5 to 50 with an interval of 5, and 10% of labeled samples are used in SVM for classification on all band subsets. The classification accuracy in terms of OA is shown in Fig. 8.

As seen in Fig. 8, NGNMF has outperformed all other BS methods on all the three datasets, which have shown the superiority of the proposed algorithm. For the ranking-based NMFW method, it only achieves limited classification accuracies, mainly because the correlation between selected bands is relatively high, leading to loss of some spectral discrimination information. By contrast, the EFDPC and FNGBS methods combine the clustering and ranking strategies and achieve better classification accuracies on the three datasets. For NGNMF, it firstly retains the advantages of combining clustering and ranking strategies. Moreover, compared with FNGBS, it further reduces the possibility of selecting noisy bands as representative bands and thus effectively improves the classification accuracy.

D. Analysis of E2DSSA

In this subsection, the performance of E2DSSA is analyzed and compared with 2DSSA, including embedding window parameters, number of singular values, and corresponding sub-images. Moreover, E2DSSA is compared with several classical spatial feature extraction methods in terms of feature images and classification accuracy for the first time.

1) Parameters Sensitivity Analysis

Fig. 10. Feature maps obtained by 2DSSA (a-d), EMAP (e-h), Gabor (i-l), DTRF (m-p) and E2DSSA (q-t) with different parameters on band-20 of Indian Pines.

TABLE I
CLASSIFICATION ACCURACY (%) OF DIFFERENT SPATIAL METHODS USING THE SVM CLASSIFIER.

Methods	OA	AA	kappa	time(s)
2DSSA	81.74 ±0.74	76.15 ±1.06	0.79 ±0.01	10.68
EMAP	82.64 ±1.37	81.32 ±1.43	0.80 ±0.02	2.79
Gabor	82.02 ±0.62	81.40 ±2.36	0.80 ±0.01	**1.09**
DTRF	83.55 ±1.65	83.26 ±2.82	0.81 ±0.02	4.16
E2DSSA	**88.01 ±0.83**	**87.50 ±0.45**	**0.86 ±0.01**	36.09

Similar to 2DSSA, the embedding window size of E2DSSA affects the classification performance, though the former uses a fixed rectangle window while the latter adopts an irregularly shaped mask, containing a certain number of pixels similar to the central one. In order to objectively and consistently compare the performance of these two methods, the number of pixels in the respective embedding window was set to be the same in our experiment, for example, the window of 2DSSA is 5×5 and the similar pixels of E2DSSA is 25 (including the center pixel). In the experiment, six different window sizes from 3×3 (9) to 15×15 (225) were used for comparison, and from the three datasets 2%, 1%, and 1% of labeled samples per class were randomly selected for training, with the detailed results given in Fig. 9.

As seen in Fig. 9, the performance of E2DSSA is superior to 2DSSA for all window parameters, with an improved OA averaging about 6%, 2%, and 0.5%, respectively. According to the classification accuracies of 2DSSA, it can be inferred that

Fig. 11. Classification accuracy (OA) with different numbers of training samples on the (a) Indian Pines, (b) PaviaU, and (c) Salinas datasets.

TABLE II
CLASSIFICATION ACCURACY (%) OF DIFFERENT COMPARED METHODS ON THE INDIAN PINES DATASET (5% TRAINING).

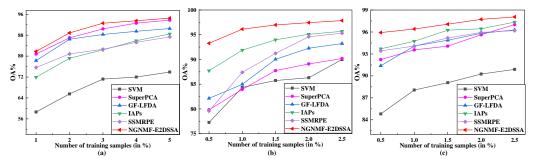| Class | Samples | | SVM | SuperPCA | GF-LFDA | IAPs | SSMRPE | NGNMF-E2DSSA |
	Train	Test	Dimension=200			Dimension=35		
1	3	43	10.85±8.17	**100.00±0.0**	94.57±1.34	81.40±22.2	61.11±15.1	84.50±12.8
2	72	1356	66.37±7.78	87.30±5.45	90.81±1.11	87.83±6.26	88.50±2.37	**93.17±0.37**
3	42	788	55.29±5.29	89.35±1.01	89.21±0.71	85.91±6.38	77.68±0.83	**93.02±0.89**
4	12	225	28.59±2.72	62.08±0.31	67.41±7.47	62.81±2.60	45.57±4.24	**88.15±10.8**
5	25	458	86.54±2.53	**96.66±0.68**	96.43±1.98	96.00±2.53	92.94±0.26	96.00±1.20
6	37	693	92.74±3.60	96.20±3.14	99.37±0.58	99.13±0.76	**99.45±0.09**	98.03±0.96
7	2	26	60.26±13.5	96.30±0.20	**100.00±0.0**	98.72±2.22	96.00±0.02	91.03±5.88
8	24	454	94.42±2.42	**99.57±0.00**	97.72±1.50	99.05±0.89	98.63±0.79	97.21±2.33
9	1	19	12.28±10.9	**100.00±0.0**	**100.00±0.0**	75.44±4.53	70.37±6.42	85.96±4.31
10	49	923	68.33±7.43	91.61±3.17	86.75±2.39	85.41±3.24	86.09±0.84	**91.84±1.02**
11	123	2332	81.63±1.29	92.57±4.34	89.11±0.72	90.38±2.80	87.13±0.36	**95.21±0.76**
12	30	563	46.95±2.00	88.56±7.07	84.78±2.21	81.88±7.23	85.02±0.42	**90.47±6.98**
13	11	194	95.19±2.54	**99.50±0.56**	98.97±0.52	98.80±0.60	98.76±0.31	96.05±0.30
14	64	1201	91.73±2.62	**99.84±0.21**	94.98±2.06	97.06±1.84	96.68±0.84	98.95±0.38
15	20	366	40.98±6.70	**98.67±0.29**	88.34±3.48	83.24±5.37	81.03±0.65	95.81±2.48
16	5	88	67.05±15.9	80.00±21.1	88.26±5.83	90.53±9.11	71.37±4.75	**99.24±0.66**
	OA		73.84±0.50	92.53±0.60	90.81±0.23	90.03±1.58	87.97±0.24	**94.78±0.29**
	AA		62.45±0.89	92.39±1.26	91.67±0.49	88.35±5.29	83.52±1.33	**93.41±1.08**
	kappa		0.699±0.006	0.915±0.007	0.895±0.003	0.886±0.019	0.863±0.003	**0.94±0.003**
	Computation time (s)		3.9	6.5	14.5	16.3	12.2	8.7

the performance of 2DSSA is severely affected by the shape of land covers. The bigger embedding window can obtain better classification results on Indian Pines and Salinas datasets because these datasets have a large area of ground features, while the smaller window is better for the PaviaU due to its finely broken and striped ground features. In contrast, the strategy of selecting similar pixels in E2DSSA can adaptively extract features of ground objects, thus it has a more robust performance for complex land covers, with the best parameters of the three datasets concentrated in the range of 49-121 in Fig. 9.

2) Comparisons with classical spatial methods

In this part, the proposed E2DSSA method was further compared with some classic spatial methods including EMAP [25], Gabor [26], DTRF [28] and 2DSSA [29] on the Indian Pines dataset. For the parameters of these methods, four different embedding windows, i.e. 3×3 to 9×9 for 2DSSA and 9 to 81 for E2DSSA are used. EMAP was built with λ_a=500 for the area attribute and λ_s=15 for the standard deviation attribute in thinning and thickening operators. Regarding the Gabor filter, the four orientations $\theta = [\pi/8, 3\pi/8, 5\pi/8, 7\pi/8]$ were considered. In DTRF, the parameters δ_s and δ_r were set in the range from 50 to 200 with a step of 50 and 0.3 to 0.9 with a step of 0.2 to adjust the amount of smoothness. The corresponding spatial feature images on band-20 of these

methods are shown in Fig. 10.

According to the experimental results, 2DSSA seems to smooth out spatial detailed information such as edges, especially in the larger embedded window. For EMAP and Gabor, the image feature information under a single parameter is very limited, thus the multi-scale feature fusion in various parameters is widely used for improved classification. As for DTRF, it can smooth areas within classes and preserve the edges between classes, however, it is sensitive to noise, resulting in foggy artifacts. In contrast, E2DSSA with different parameters can retain the shape and edge features of land covers, and eliminate the influence of noise to a certain extent. Moreover, as the embedding window expands, E2DSSA can further improve the intra-class similarity and retain interclass differences of ground objects, which can be beneficial for the classification of ground objects.

Moreover, we also quantitatively evaluated the classification accuracy of the above features on the SVM classifier with 2% labelled samples. 2DSSA, DTRF and E2DSSA are applied to each band of HSI, with parameters of an embedding window of 7×7, δ_s=50 and δ_r=0.3, L=49, respectively, while EMAP and Gabor filter with multiple parameters shown in Fig. 10, both act on the first three principal components after PCA. The classification results are given in Table I. As shown in Table I, the proposed E2DSSA method outperforms all other spatial

TABLE III
CLASSIFICATION ACCURACY (%) OF DIFFERENT COMPARED METHODS ON THE PAVIAU DATASET (1% TRAINING).

Class	Samples		SVM	SuperPCA	GF-LFDA	IAPs	SSMRPE	NGNMF-E2DSSA
	Train	Test	Dimension=103			Feature Dimension=35		
1	67	6564	83.31±2.71	77.28±0.99	80.76±1.14	97.73±0.14	78.54±2.94	**98.47±0.81**
2	187	18462	92.20±1.45	95.84±1.04	90.79±0.33	**99.87±0.03**	95.95±0.68	99.32±0.54
3	21	2078	56.77±17.5	84.02±3.39	67.71±0.52	71.90±1.26	72.87±7.21	**91.72±0.43**
4	31	3033	76.64±13.9	47.62±8.57	93.66±1.40	86.05±2.61	**94.87±4.96**	90.97±6.13
5	14	1331	98.90±0.17	98.07±0.88	**100.00±0.0**	99.92±0.13	66.16±4.55	99.40±0.39
6	51	4978	71.28±3.98	91.11±1.13	79.15±2.41	77.13±3.68	92.09±5.09	**94.48±0.31**
7	14	1316	74.10±10.7	65.73±9.71	69.06±4.77	59.32±5.94	88.38±3.54	**94.38±1.28**
8	37	3645	76.35±5.73	73.42±2.41	73.02±0.52	88.65±1.79	76.23±3.60	**91.80±4.04**
9	10	937	99.79±0.02	33.69±4.38	**99.96±0.06**	93.95±1.82	98.41±2.45	86.91±7.53
OA			83.96±1.27	84.20±0.80	85.59±0.47	92.15±0.44	88.07±0.93	**96.58±0.42**
AA			81.04±1.98	74.09±2.05	83.79±0.88	86.06±0.59	84.83±0.63	**94.16±1.16**
kappa			0.786±0.018	0.788±0.012	0.807±0.006	0.894±0.006	0.841±0.012	**0.954±0.006**
Computation time (s)			2.5	8.6	32.5	26.2	33.2	47.5

TABLE IV
CLASSIFICATION ACCURACY (%) OF DIFFERENT COMPARED METHODS ON THE SALINAS DATASET (1% TRAINING)

Class	Samples		SVM	SuperPCA	GF-LFDA	IAPs	SSMRPE	NGNMF-E2DSSA
	Train	Test	Dimension=204			Feature Dimension=35		
1	21	1988	98.76±0.57	**100.00±0.0**	99.98±0.03	98.39±0.96	98.57±2.40	99.09±1.57
2	38	3688	99.01±0.52	**100.00±0.0**	99.46±0.51	99.73±0.29	94.27±8.22	99.45±0.42
3	20	1956	97.60±1.74	81.10±16.0	98.72±0.53	99.61±0.16	99.19±0.35	**99.97±0.06**
4	14	1380	98.09±0.73	70.58±14.2	94.57±6.02	96.86±1.09	97.75±0.56	**98.69±0.51**
5	27	2651	97.47±0.70	98.46±1.88	98.24±1.44	99.01±0.81	**99.99±0.02**	96.72±1.50
6	40	3919	99.57±0.10	99.92±0.02	**100.00±0.0**	99.68±0.47	97.17±2.50	99.46±0.47
7	36	3543	99.25±0.17	90.39±10.8	99.93±0.04	99.60±0.13	98.05±1.25	99.73±0.11
8	113	11158	75.91±2.18	**99.73±0.34**	89.45±3.92	94.48±2.75	92.96±1.55	93.55±2.00
9	63	6140	98.43±0.41	96.26±2.72	98.82±0.70	99.72±0.32	99.75±0.22	**99.79±0.10**
10	33	3245	86.59±1.89	90.64±10.3	96.24±1.25	92.96±1.08	90.81±4.04	**96.34±2.42**
11	11	1057	90.79±2.49	67.42±8.57	92.91±5.61	97.48±1.23	**99.97±0.06**	96.87±2.70
12	20	1907	98.20±1.87	84.30±6.22	96.74±1.02	98.99±1.06	99.60±0.38	**99.93±0.06**
13	10	906	98.12±0.38	**98.25±0.04**	95.11±2.49	97.39±0.63	96.68±2.66	97.83±1.02
14	11	1059	89.39±1.34	83.54±19.8	96.85±0.80	90.71±3.46	**98.57±0.46**	96.15±1.89
15	73	7195	66.18±2.24	**94.17±9.44**	78.33±5.29	82.09±4.34	91.66±4.36	92.51±3.04
16	19	1788	96.36±3.44	85.29±3.84	98.82±0.26	**99.40±0.31**	90.27±7.35	97.57±1.63
OA			88.38±0.38	93.95±0.99	93.66±1.39	95.38±0.66	95.40±0.78	**96.82±0.25**
AA			93.11±0.33	90.00±1.67	95.89±032	96.63±0.15	96.58±0.42	**97.73±0.26**
kappa			0.871±0.004	0.933±0.011	0.929±0.015	0.949±0.007	0.949±0.009	**0.965±0.003**
Computation time (s)			13.1	18.9	22.4	26.5	56.4	42.2

methods, which demonstrates it can effectively extract the spatial features of HSI and improve the separability of ground objects. Nevertheless, the calculation cost of the proposed E2DSSA method is high, which is mainly due to the construction of the adaptive embedded window.

E. Comparisons with other state-of-the-art methods

To highlight the superiority of the proposed NGNMF-E2DSSA method, in this subsection, we compare a few state-of-the-art spectral-spatial methods, i.e. the SuperPCA [55], GF-LFDA [36], IAPs [56] and SSMRPE [17], while the raw data with SVM was set as the baseline method (abbreviated as "SVM"). These state-of-the-art methods can jointly utilize the spectral and spatial information and reduce the dimensions of the HSI data. We downloaded the source code or wrote the code of each compared method and set the parameters optimally as suggested in [17, 36, 55, 56] to generate the results. The embedding window of E2DSSA was set to 81 for all three datasets. For a fair comparison, the reduced feature dimension was set to 35 for these spectral-spatial methods.

Firstly, to fully investigate the performance of the involved DR methods, different number of training samples are used, i.e., varying within {1%, 2%, 3%, 4%, 5%} for Indian Pines, {0.5%, 1%, 1.5%, 2%, 2.5%} for both PaviaU and Salinas datasets per class. Fig. 11 displays and compares the classification accuracies under different training samples on the three datasets. According to Fig. 11, the OA of all DR methods has improved with the increased training samples, mainly because a large training set can provide more information to learn the discriminant features. In addition, some compared methods, such as SuperPCA and GF-LFDA, achieve superior results on some datasets, but poor results on others. The proposed NGNMF-E2DSSA can achieve the highest classification accuracies under all conditions, demonstrating its high effectiveness and robustness, thanks to the enhanced spatial characteristics on the selected band subset and improved discriminability of object features. In order to quantitatively evaluate the performance of each class in different DR methods, 5%, 1% and 1% labeled samples of each class were randomly selected as training samples and the rest were used for testing from the three datasets, respectively. Tables II-IV compare the detailed classification accuracies of the three datasets, with the classification maps shown in Figs. 12-14 for comparison.

In Tables II-IV, NGNMF-E2DSSA achieved the highest classification accuracies in terms of OA, AA, kappa and many class based metrics on the three datasets. SuperPCA transforms within each superpixel region to reduce the data dimension, but it has poor classification performance on smaller ground objects due to the problem of segmentation scale, especially on the PaviaU dataset. The classification accuracy obtained by GF-

Fig. 12. Classification maps on the Indian Pines dataset by SVM (a), SuperPCA (b), GF-LFDA (c), IAPs (d), SSMRPE (e), and NGNMF-E2DSSA (f).

Fig. 13. Classification maps on the PaviaU dataset by SVM (a), SuperPCA (b), GF-LFDA (c), IAPs (d), SSMRPE (e), and NGNMF-E2DSSA (f).
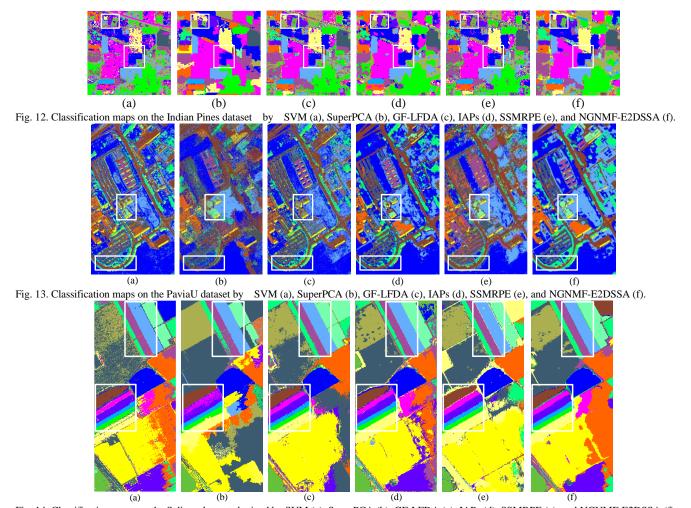

Fig. 14. Classification maps on the Salinas dataset obtained by SVM (a), SuperPCA (b), GF-LFDA (c), IAPs (d), SSMRPE (e), and NGNMF-E2DSSA (f).

LFDA on the Salinas dataset is relatively low. IAPS can quickly extract invariant structural features from images, while its overall classification performance is poor. SSMRPE has a high computation cost because it extracts the manifold structure based on the graph embedding theory. NGNMF-E2DSSA can reduce the spectral dimension and remove the noisy bands simultaneously, and the subsequent E2DSSA can extract and enhance the image features to improve the identification ability of ground objects, which achieves superior classification results. However, since the trajectory matrix is embedded from the whole image, the computation time will increase exponentially with the increased image size, which is also a problem that needs to be further addressed in the future.

The numerical results can be visually compared by inspecting corresponding classification maps in Figs. 12–14. Due to the spectral mixture in hyperspectral images, there are many salt-and-pepper noise in the classification maps. Compared with SVM, GF-LFDA and SSMRPE reduced the classification noise to a certain extent, but they cannot completely remove them. Some misclassified plaques appear in the classification map of SuperPCA due to the superpixel segmentation scale. IAPs basically solves the problem of classification noise inside the land covers, but their performance is limited at the edge of the ground features. Finally, NGNMF-E2DSSA can produce more homogeneous

and smoother classification maps while preserving the edges. Furthermore, it is clear from the classification details in the white boxes that E2DSSA has excellent classification performance on irregularly shaped and small ground objects. The morphology of the ground objects is well preserved with a much higher internal smoothness. This again demonstrates the effectiveness of the extracted features of NGNMF-E2DSSA for HSI classification and land mapping.

F. Comparisons with deep learning methods

In this section, we further compare the proposed NGNMF-E2DSSA with several deep learning methods, including LSTM, 2DCNN, 3DCNN, Morphological Convolutional Neural Networks (MorphCNN), and HybridSN, as provided in [57]. Their classification accuracies are evaluated under spatially disjoint train-test samples based on the Indian Pines and PaviaU datasets (http://dase.grss-ieee.org/index.php), and the results are shown in Tables V and VI. As seen, the proposed NGNMF-E2DSSA achieved the highest classification accuracy on the Indian Pines dataset, while the AA was slightly lower than that of MorphCNN. In addition, the accuracy of NGNMF-E2DSSA was also better than that of LSTM, 2DCNN, 3DCNN and HybridSN on the PU dataset, but lower than that of MorphCNN. NGNMF-E2DSSA can extract the spectral-spatial features of hyperspectral images in a simpler way than the

TABLE V
CLASSIFICATION ACCURACY (%) COMPARISON BETWEEN NGNMF-E2DSSA AND DEEP LEARNING METHODS ON INDIAN PINES WITH SPATIALLY DISJOINT SAMPLES.

Class	Train	Test	LSTM	2DCNN	3DCNN	HybridSN	MorphCNN	NGNMF-E2DSSA
1	29	25	89.60	73.64	48.18	82.66	**92.27**	72.00
2	762	675	82.22	83.12	85.12	82.17	84.05	**91.26**
3	435	404	64.15	81.98	77.22	76.73	79.34	**95.30**
4	146	99	55.35	45.39	50.11	33.33	**52.14**	45.45
5	232	274	89.27	89.11	80.28	81.14	**91.66**	91.24
6	394	354	96.39	95.02	89.81	97.36	95.74	**99.15**
7	16	2	0.00	0.00	0.00	0.00	0.00	0.00
8	235	250	99.2	99.96	95.96	96.53	100.0	**100.00**
9	10	10	76.00	26.66	**77.78**	66.66	44.44	50.00
10	470	503	81.51	77.44	77.90	74.35	80.77	**95.23**
11	1424	1065	80.40	89.40	82.73	79.18	88.54	**91.74**
12	328	282	76.31	87.72	82.64	71.04	**88.46**	83.69
13	132	80	**97.25**	95.28	89.72	96.25	87.64	96.25
14	728	545	94.13	98.94	98.31	91.68	98.82	**99.63**
15	291	99	**90.71**	82.02	55.17	45.45	69.44	64.65
16	57	44	**94.09**	82.00	82.50	84.09	84.00	88.64
OA			83.55	87.25	83.60	80.86	87.45	**91.91**
AA			79.16	75.48	73.38	72.41	**77.33**	75.89
kappa			0.813	0.855	0.814	0.782	0.858	**0.908**

TABLE VI
CLASSIFICATION ACCURACY (%) COMPARISON BETWEEN NGNMF-E2DSSA AND DEEP LEARNING METHODS ON PAVIAU WITH SPATIALLY DISJOINT SAMPLES.

Class	Train	Test	LSTM	2DCNN	3DCNN	HybridSN	MorphCNN	NGNMF-E2DSSA
1	548	6304	82.63	92.40	85.66	89.74	**94.52**	87.58
2	540	18146	78.74	96.84	95.88	81.78	**96.58**	91.24
3	392	1815	60.73	65.48	68.11	82.88	**85.08**	80.77
4	524	2912	97.10	93.25	97.02	83.66	**96.89**	96.70
5	265	1113	99.28	98.02	98.90	**99.94**	99.25	99.46
6	532	4572	65.94	80.52	68.85	72.43	93.92	**94.90**
7	375	981	84.95	89.29	73.09	**96.16**	84.69	95.41
8	514	3364	88.89	92.50	95.21	92.80	96.62	**98.87**
9	231	795	**98.29**	93.80	93.54	94.04	97.05	94.72
OA			80.38	91.23	89.43	84.18	**95.35**	92.05
AA			84.06	89.01	86.25	88.16	**93.50**	93.30
kappa			0.743	0.892	0.856	0.791	**0.936**	0.895

traditional 2D/3D convolution methods, and the obtained features have stronger interpretability without being too abstract. The main reason why the proposed method does not have a significant advantage over MorphCNN model is that the latter embeds the feature extraction MP method into the deep network, which effectively improves the feature extraction capability. Such deep models embedded in traditional methods are also an interesting trend of current research. This also inspires us to carry out subsequent related research.

IV. CONCLUSION

In this paper, a novel dimensionality reduction method, namely NGNMF-E2DSSA, is proposed to extract interpretable low-dimensional features in HSI. The NGNMF is used to reduce the dimension and preserve the relevant original spectral information, where E2DSSA is used to enhance the spatial context and structural information of each selected band. The obtained low-dimensional features can significantly improve the classification performance of HSI.

Compared with existing DR methods, experimental results on three publicly available HSI datasets have fully demonstrated that the proposed NGNMF-E2DSSA can achieve superior classification results, and shows good smoothness and continuity in the classification maps. In addition, the extracted low-dimensional features are obtained from the space of the original data without data transformation or projection, hence a higher interpretability of the features.

While E2DSSA can adaptively extract the local features of different objects whilst taking into account the global features, the use of a large embedding window will increase the computational cost, especially on large-size images. This problem also exists in conventional 2DSSA. In our future work, we will focus on how to combine other techniques, such as image partition, to improve the computational efficiency of E2DSSA and further benefit more efficient the HSI classification.

REFERENCES

[1] M. Imani, and H. Ghassemian, "An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges," *Inf. Fusion*, vol. 59, pp. 59-83, Jul. 2020.

[2] W. Chen, Z. Yang, J. Ren, J. Cao, N. Cai, H. Zhao, and P. Yuen, "MIMN-DPP: Maximum-information and minimum-noise determinantal point processes for unsupervised hyperspectral band selection," *Pattern Recognit.*, vol. 102, pp. 107213, Jun. 2020.

[3] X. Huang, and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173-4185, Dec. 2008.

[4] X. Xu, J. Li, S. Li, and A. Plaza, "Subpixel component analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5564-5579, Apr. 2019.

[5] K. Siebels, K. Goïta, and M. Germain, "Estimation of mineral abundance from hyperspectral data using a new supervised neighbor-band ratio unmixing approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 6754-6766, Mar. 2020.

[6] Y. Yan, J. Ren, J. Tschannerl, H. Zhao, B. Harrison, and F. Jack, "Nondestructive phenolic compounds measurement and origin discrimination of peated barley malt using Near-Infrared hyperspectral imagery and machine learning," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-15, May. 2021

[7] W. Zhao, and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544-4554, Apr. 2016.

[8]   D. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," *Proc. AMS Math Challenges Lecture*, pp. 1-32, Jan. 2000.

[9]   G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.

[10]   X. Jia, B. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676-697, Feb. 2013.

[11]   H. Huang, and M. Yang, "Dimensionality reduction of hyperspectral images with sparse discriminant embedding," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5160-5169, Apr. 2015.

[12]   J. Tschannerl, J. Ren, P. Yuen, G. Sun, H. Zhao, Z. Yang, Z. Wang, and S. Marshall, "MIMR-DGSA: Unsupervised hyperspectral band selection based on information theory and a modified discrete gravitational search algorithm," *Inf. Fusion*, vol. 51, pp. 189-200, Nov. 2019.

[13]   J. Zabalza, J. Ren, M. Yang, Y. Zhang, J. Wang, S. Marshall, and J. Han, "Novel Folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS-J. Photogramm. Remote Sens.*, vol. 93, pp. 112-122, Jul. 2014.

[14]   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 Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10394-10409, Jan. 2021.

[15]   X. Li, J. Pan, Y. He, and C. Liu, "Bilateral filtering inspired locality preserving projections for hyperspectral images," *Neurocomputing*, vol. 164, pp. 300-306, Sep. 2015.

[16]   W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185-1198, Oct. 2012.

[17]   H. Huang, G. Shi, H. He, Y. Duan, and F. Luo, "Dimensionality reduction of hyperspectral imagery based on spatial–spectral manifold learning," *IEEE T. Cybern.*, vol. 50, no. 6, pp. 2604-2616, Mar. 2020.

[18]   X. Zheng, Y. Yuan, and X. Lu, "Dimensionality reduction by spatial–spectral preservation in selected bands," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5185-5197, Jun. 2017.

[19]   B. Zhang, Z. C. Chen, L. F. Zheng, Q. X. Tong, Y. N. Liu, Y. D. Yang, and Y. Q. Xue, "Object detection based on feature extraction from hyperspectral imagery and convex cone projection transform," *J. Infrared Millim. Waves*, vol. 23, no. 6, pp. 441-+, Dec. 2004.

[20]   S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88-102, Jul. 2016.

[21]   H. Sun, J. Ren, H. Zhao, G. Sun, W. Liao, Z. Fang, and J. Zabalza, "Adaptive distance-based band hierarchy (ADBH) for effective hyperspectral band selection," *IEEE Trans Cybern.*, early access, Mar. 2020, doi: 10.1109/TCYB.2020.2977750.

[22]   Q. Wang, Q. Li, and X. Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5028-5039, Jul. 2021.

[23]   J. Feng, J. Jiao, F. Liu, T. Sun, and X. Zhang, "Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images," *Pattern Recognit.*, vol. 51, pp. 295-309, Mar. 2016.

[24]   L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879-893, Aug. 2012.

[25]   M. D. Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542-546, Dec. 2011.

[26]   X. Kang, C. Li, S. Li, and H. Lin, "Classification of hyperspectral images by Gabor filtering based deep network," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1166-1178, Apr. 2018.

[27]   X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666-2677, Jul. 2014.

[28]   X. Kang, P. Duan, and S. Li, "Hyperspectral image visualization with edge-preserving filtering and principal component analysis," *Inf. Fusion*, vol. 57, pp. 130-143, May. 2020.

[29]   J. Zabalza, J. Ren, J. Zheng, J. Han, H. Zhao, S. Li, and S. Marshall, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4418-4433, Feb. 2015.

[30]   H. Fu, G. Sun, J. Ren, A. Zhang, and X. Jia, "Fusion of PCA and Segmented-PCA domain multiscale 2-D-SSA for effective spectral-spatial feature extraction and data classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1-14, Nov. 2022.

[31]   G. Sun, H. Fu, J. Ren, A. Zhang, J. Zabalza, X. Jia, and H. Zhao, "SpaSSA: Superpixelwise adaptive SSA for unsupervised spatial-spectral feature extraction in hyperspectral image," *IEEE T. Cybern.*, early access, Sep. 2021, doi: 10.1109/TCYB.2021.3104100.

[32]   X. Kang, S. Li, L. Fang, and J. A. Benediktsson, "Intrinsic image decomposition for feature extraction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2241-2253, Oct. 2015.

[33]   S. Li, Q. Hao, X. Kang, and J. A. Benediktsson, "Gaussian pyramid based multiscale feature fusion for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3312-3324, Aug. 2018.

[34]   X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140-7151, Sep. 2017.

[35]   N. Falco, J. A. Benediktsson, and L. Bruzzone, "Spectral and spatial classification of hyperspectral images based on ICA and reduced morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6223-6240, Jul. 2015.

[36]   H. Zhang, W. Liu, and H. Lv, "Spatial-spectral joint classification of hyperspectral image with locality and edge preserving," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2240-2250, May. 2020.

[37]   Z. Wu, J. Sun, Y. Zhang, Y. Zhu, J. Li , A. Plaza, J. Benediktsson, and Z. Wei. "Scheduling-Guided Automatic Processing of Massive Hyperspectral Image Classification on Cloud Computing Architectures," *IEEE Trans Cybern.*, vol. 51, no. 7, pp. 3588-3601, July 2021.

[38]   Z. Wu, J. Sun, Y. Zhang, Z. Wei and J. Chanussot. "Recent Developments in Parallel and Distributed Computing for Remotely Sensed Big Data Processing," *Proc. IEEE*, vol. 109, no. 8, pp. 1282-1305, Aug. 2021.

[39]   S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179-204, Oct. 2021.

[40]   S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep Learning for Hyperspectral Image Classification: An Overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690-6709, Apr. 2019.

[41]   M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. Plaza, J. Li, and F. Pla, "Capsule networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2145-2160, Oct. 2019.

[42]   M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral–spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740-754, Aug. 2019.

[43]   X. Zhang, G. Sun, X. Jia, L. Wu, A. Zhang, J. Ren, H. Fu, and Y. Yao, "Spectral-spatial self-attention networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, early access, Aug. 2021, doi: 10.1109/TGRS.2021.3102143.

[44]   H. Liang, and Q. Li, "Hyperspectral imagery classification using sparse representations of convolutional neural network features," *Remote Sens.*, vol. 8, no. 2, Feb. 2016.

[45]   N. He, M. E. Paoletti, J. M. Haut, L. Fang, S. Li, A. Plaza, and J. Plaza, "Feature extraction with multiscale covariance maps for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 755-769, Aug. 2019.

[46]   A. Sellami, M. Farah, I. Riadh Farah, and B. Solaiman, "Hyperspectral imagery classification based on semi-supervised 3-D deep neural network and adaptive band selection," *Expert Syst. Appl.*, vol. 129, pp. 246-259, Sep. 2019.

[47]   Y. Chen, K. Zhu, L. Zhu, X. He, P. Ghamisi, and J. A. Benediktsson, "Automatic design of convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7048-7066, Apr. 2019.

[48]   L. Ji, L. Wang, and X. Geng, "An automatic bad band pre-removal method for hyperspectral imagery," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4985-4994, Oct. 2019.

[49]   C. Ricotta, "From the Euclidean distance to compositional dissimilarity: What is gained and what is lost," *Acta Oecol.-Int. J. Ecol.*, vol. 111, pp. 103732, Aug. 2021.

[50] B. Huang, L. Ge, G. Chen, M. Radenkovic, X. Wang, J. Duan, and Z. Pan, "Nonlocal graph theory based transductive learning for hyperspectral image classification," *Pattern Recognit.*, vol. 116, pp. 107967, Aug. 2021.

[51] L. Zhuang, X. Fu, M. Ng, and J. Bioucas-Dias, "Hyperspectral image denoising based on global and nonlocal low-rank factorizations," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, pp. 1-17, Jan. 2021.

[52] Z. Wu, W. Zhu, J. Chanussot, Y. Xu, S. Osher. "Hyperspectral Anomaly Detection via Global and Local Joint Modeling of Background," *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3858-3869, July 2019.

[50] F. Melgani, and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778-1790, Nov. 2004.

[51] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1519-1531, Oct. 2016.

[52] 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 Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581-4593, Jun. 2018.

[53] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791-3808, Jan. 2020.

[54] S. Shabbir, M. Ahmad, S. Roy, D. Hong, X. Wu, J. Yao, A. Khan, M. Mazzara, S. Distefano, and J. Chanussot, "Hyperspectral image classification - traditional to deep models: A survey for future prospects," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, early access, Dec. 2021, doi: 10.1109/JSTARS.2021.3133021.

[55] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.

Hang Fu received the B.Sc. degrees in geomatics engineering from China University of Petroleum (East China), Qingdao, China, in 2019. He is currently pursuing the PhD degree in computer technology and resource information engineering at China University of Petroleum (East China), Qingdao, China.

His research interests include feature extraction and hyperspectral classification.

Aizhu Zhang (Member, IEEE) received the B.Sc., M.Sc. and Ph.D. degrees from China University of Petroleum (East China), Qingdao, China, in 2011, 2014, and 2017, respectively.

She is currently a Lecturer with China University of Petroleum (East China), Qingdao, China. Her research interests are in artificial intelligence, pattern recognition, city remote sensing, and wetland remote sensing.

Genyun Sun (Member, IEEE) received the B.Sc. degree from Wuhan University, Wuhan, China, in 2003 and Ph.D. degree in Institute of Remote Sensing Applications, Chinese Academy of Sciences in 2008.

He is currently a Professor with China University of Petroleum (East China), Qingdao, China. His research interests cover remote sensing image processing, including hyperspectral and high resolution remote sensing, and intelligent optimization.

Jinchang Ren (Senior Member, IEEE) received the B.Eng., M.Eng. and D.Eng. degrees from the Northwestern Polytechnical University, Xi'an, China in 1992, 1997 and 2000, respectively, and the Ph.D. degree from the University of Bradford, Bradford, U.K., in 2019.

He is currently a Professor with the National Subsea Centre, Robert Gordon University, Aberdeen, U.K. His research interests include image processing, computer vision, machine learning, and big data analytics.

Dr. Ren acts as Associate Editor for several international journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) and *Journal of the Franklin Institute*.

Xiuping Jia (Fellow, IEEE) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982 and the Ph.D. degree in Electrical Engineering from The University of New South Wales, Australia, in 1996.

Since 1988, she has been with the School of Engineering and Information Technology, The University of New South Wales at Canberra, Australia, where she is currently an Associate Professor. She has authored or coauthored more than 270 referred papers, including over 160 journal papers, addressing various topics including data correction, feature reduction and image classification using machine learning techniques. She has coauthored of the remote sensing textbook titled *Remote Sensing Digital Image Analysis* [Springer-Verlag, 3rd (1999) and 4th eds. (2006)]. Her research interests include remote sensing, hyperspectral image processing and spatial data analysis.

Dr. Jia is a Subject Editor of the *Journal of Soils and Sediments* and an Associate Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Zhaojie Pan received the B.Sc. degree in geomatics engineering from China University of Petroleum (East China), Qingdao, China, in 2021. He is currently working toward the M.Sc. degree at China University of Petroleum (East China), Qingdao, China.

His research interests include hyperspectral classification and deep learning.

Hongzhang Ma received the M.Sc. degree in atmospheric physics science from Nanjing Institute of Meteorology, Nanjing, China, in 2005 and the Ph.D. degree in cartography and geography information system from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2011.

He is currently an Associate Professor with the College of Science, China University of Petroleum (East China), Qingdao, China. His research interests include microwave remote sensing of soil moisture and application technology of multisource remote sensing data.