FU, H., SUN, G., ZHANG, A., SHAO, B., REN, J. and JIA, X. 2023. Tensor singular spectral analysis for 3D feature extraction in hyperspectral images. *IEEE transactions on geoscience and remote sensing* [online], 61, article 5403914. Available from: https://doi.org/10.1109/TGRS.2023.3272669

Tensor singular spectral analysis for 3D feature extraction in hyperspectral images.

FU, H., SUN, G., ZHANG, A., SHAO, B., REN, J. and JIA, X.

2023

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3272669

Tensor Singular Spectral Analysis for 3D feature extraction in hyperspectral images

Hang Fu, Genyun Sun, Senior *Member, IEEE*, Aizhu Zhang, *Member, IEEE*,Baojie Shao, Jinchang Ren, *Senior Member, IEEE*, and Xiuping Jia, *Fellow, IEEE*

***Abstract*—Due to the cubic structure of a hyperspectral image (HSI), how to characterize its spectral and spatial properties in three dimensions is challenging. Conventional spectral-spatial methods usually extract spectral and spatial information separately, ignoring their intrinsic correlations. Recently, some 3D feature extraction methods are developed for the extraction of spectral and spatial features simultaneously, although they rely on local spatial-spectral regions and thus ignore the global spectral similarity and spatial consistency. Meanwhile, some of these methods contain huge model parameters which require a large number of training samples. In this paper, a novel Tensor Singular Spectral Analysis (TensorSSA) method is proposed to extract global and low-rank features of HSI. In TensorSSA, an adaptive embedding operation is first proposed to construct a trajectory tensor corresponding to the entire HSI, which takes full advantage of the spatial similarity and improves the adequate representation of the global low-rank properties of the HSI. Moreover, the obtained trajectory tensor, which contains the global and local spatial and spectral information of the HSI, is decomposed by the Tensor singular value decomposition (t-SVD) to explore its low-rank intrinsic features. Finally, the efficacy of the extracted features is evaluated using the accuracy of image classification with a support vector machine (SVM) classifier. Experimental results on three publicly available datasets have fully demonstrated the superiority of the proposed TensorSSA over a few state-of-the-art 2D/3D feature extraction and deep learning algorithms, even with a limited number of training samples.**

***Index Terms*—Hyperspectral image (HSI), 3D feature extraction, TensorSSA, adaptive embedding, trajectory tensor.**

I. Introduction

As a 3D hypercube, hyperspectral images (HSI) contain a 2D spatial scene and a rich 1D spectral profile, which has enabled its ability to detect and identify the minute differences of objects and their changes. HSI is thus widely used in various applications and fields such as mineralogy [1], agriculture [2], load cover classification [3], and target detection [4]. However, raw HSIs often suffer from spectral variations caused by sensor noise and environmental conditions, resulting in poor classification performance [5, 6]. Therefore, effective feature extraction is essential to enhance the separability between different categories in hyperspectral classification.

In the past few decades, a series of feature extraction methods have been developed. Among them, linear transformation models, such as principal component analysis (PCA) [7] and linear discriminant analysis (LDA) [8], have been widely used for spectral feature extraction in HSI. Besides, some manifold learning methods are further developed to analyze the intrinsic features of HSI, improving the separability of the spectral pixels [9, 10]. However, these methods only consider spectral information while ignoring the potential role of spatial information. In recent years, joint spectral-spatial feature extraction methods have received much attention. In most spectral-spatial methods, spectral transform methods are used for spectral feature extraction and spatial methods further extract spatial features, generating the joint spectral-spatial features [11-13]. Besides, some two-branch networks [14-16] are also proposed, which perform feature extraction in the spectral and spatial domains separately, and fuse the features from both branches to improve the classification performance. Actually, the spectral and spatial processing of these methods is an independent process, with a simple fusion of features yielding the final spectral-spatial feature. Unfortunately, the joint dependence of spectral continuity and spatial similarity that is unique to HSI data is often ignored by these methods [17].

To tackle such insufficiency, some 3D spectral-spatial feature extraction methods have been developed due to their potential to extract intrinsic features in high-dimensional data [18]. They can be further divided into three categories: i.e. 3D filter-based methods, tensor-based methods, and deep learning-based methods.

3D filters or operators are usually used to extract spectral-spatial features extraction simultaneously. For example, 3D

Manuscript received xxx; accepted December xxx. Date of publication xxx; date of current version xxx. This work was supported by the National Natural Science Foundation of China (41971292, 42271347), and in part by the Opening Fund of the Key Laboratory of Poyang Lake Wetland and Watershed Research, Jiangxi Normal University, Ministry of Education, under Grant PK2020003. (*Corresponding author: Genyun Sun*.)

Hang Fu and Baojie Shao are with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China.

Genyun Sun is with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China, and also with the Laboratory for Marine Mineral Resources, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China (e-mail: genyunsun@163.com).

Aizhu Zhang is with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China, and also with the Key Laboratory of Poyang Lake Wetland and Watershed Research, Jiangxi Normal University, Ministry of Education, Nanchang 330022, China.

Jinchang Ren is with the School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou 510665, China, and also with the National Subsea Centre, Robert Gordon University, Aberdeen AB10 7AQ, U.K. (e-mail: jinchang.ren@ieee.org).

Xiuping Jia is with the School of Engineering and Information Technology, University of New South Wales at Canberra, Canberra, ACT 2600, Australia. (e-mail: x.jia@adfa.edu.au).

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3272669

morphological profile (3DMP) [19] uses 3D erosion and dilation filters to extract the joint spectral-spatial morphological information of HSI. 3D local binary pattern (3DLBP) [20] characterizes spectral-spatial relationships by encoding a local 3D regular octahedral. In [21], 3DGabor uses a set of 3D complex Gabor wavelet filters with multiple frequencies and orientations to extract joint spatial-spectral features [22], which is further developed in Jia *et al* [23] as a 3D Gabor phase-based coding (3DGPC) for improved efficiency and efficacy. In [24], Tsai *et al.* proposed a 3D gray-level co-occurrence matrix (3DGLCM) for high-order texture analysis. Although the afore-mentioned 3D filter-based methods have achieved some success in exploring the 3D features of the HSIs, there are still some shortages. Firstly, the dimensionality of the obtained 3D features is usually huge, and the discriminability of each module varies from the other [25]. In addition, the processing unit of these methods is a small 3D block, which can only characterize the local structures of the HSI, ignoring the global relevance of the spectral and spatial information.

Due to the inherent low-dimension distribution characteristics of HSI, tensors are used for data analysis as they can effectively explore its low rankness [26, 27]. Currently, tensor-based methods have been widely used in image reconstruction [28], super-resolution [29], and data classification [30] of HSI. Among them, tensor factorization [31], such as CANDECOMP/PARAFAC (CP), Tucker, and higher order singular value decomposition (HOSVD), have achieved superior performance in HSI processing [32-35]. These methods decompose the HSI into sub-tensors, such as rank-1 tensors or kernel tensors, to achieve effective data compression, and synthesis by partial sub-tensors can yield a low-rank approximation to the original HSI. Recently, the tensor singular value decomposition (t-SVD) has been proposed for image restoration [36] and denoising [37]. By defining the tensor tube rank through the Fourier transform, it can complete the operation and description of the entire tensor for a better representation of the low rankness of tensors. It can be inferred that t-SVD has great potential for the global processing of HSI, nevertheless, there are relatively few studies using it for 3D feature extraction.

With the rapid advancement of deep learning [38, 39], they have been successfully applied for 3D feature extraction of HSI [40, 41], especially the deep structural and latent features. As a highly representative 3D deep network, 3D convolutional neural network (3DCNN) [42] applies the 3D convolutional blocks to operate on the original image blocks for extraction of both spectral and local spatial features [43, 44]. However, there is a serious problem associated with the CNN methods in feature representation, i.e. its limited capability in only extracting local spatial and spectral information rather than the global spatial structures of HSI. One of the strategies to overcome this problem is to use graph convolutional networks (GCNs), as GCNs are capable of modeling middle- and long-range spatial relations between samples by means of their graph structure [45]. In [46], the tensor theory is further introduced to GCN to learn a tensor representation of the spatial-spectral features of HSI. Recently, some transformer-based networks have been proposed, capable of exploring global information while reinforcing useful features, such as attention transformation network (AATN) [47] and SpectralFormer [48]. Zhong et al. [49] proposed the spectral-spatial transformer network (SSTN) for exploring 3D features of HSI, achieving superior classification performance. Although the above methods have made some progress, they still face all the problems existing in deep learning, including required large training samples, a huge number of hyper-parameters, and a lack of interpretability of the models.

Recently, Singular Spectral Analysis (SSA), a technique for time series analysis, has proven its capability in hyperspectral feature extraction [50]. SSA acts on the spectral domain and considers both local and global spectral features of pixels through embedding. Similarly, its two-dimensional version (2DSSA) [51, 52] can fuse local and global features of a given band image by 2D embedding windows, well maintaining the global correlation. This makes it feasible to perform global processing of hyperspectral cubes. However, neither SSA nor 2DSSA can extract spectral and spatial features simultaneously. Although Fu *et al.* [53] further proposed a spectral-spatial SSA (1.5DSSA), it only considers local spectral and spatial information and is unable to characterize the global correlation of HSI. It is therefore necessary to explore a new SSA method that can not only extract both spectral and spatial features but also considers the global correlation of the cube.

In this paper, a novel 3D Tensor SSA (TensorSSA) combining the idea of SSA and the advantages of tensors is proposed for characterizing the intrinsic characteristics of the HSI cube. In TensorSSA, the original HSI is adaptively embedded to obtain a trajectory tensor containing global and local spectral-spatial information. Through tensor decomposition and low-rank reconstruction of the trajectory tensor, the low-rank and intrinsic features of HSI can be extracted, which also shows good noise robustness. The major contributions of this approach can be summarized as follows.

1) A novel SSA-based 3D feature extraction method, i.e., TensorSSA is proposed to characterize the global spectral-spatial correlation of HSI. Through adaptive embedding and the t-SVD process, simultaneous extraction of spectral and spatial features can be achieved. Experiments on three public datasets demonstrated that TensorSSA outperforms conventional 2D/3D and deep learning methods even with limited training samples.

2) A new form of HSI data sparsity enhancement, i.e., adaptive embedding operation is developed. It is able to exploit the spatial similarity feature of HSI to demonstrate low-rank properties in all three directions of the trajectory tensor, improving the effectiveness of tensor decomposition. It combined with the corresponding reprojection operation enhances the intra-class similarity while preserving the inter-class differences of the objects.

3) A trajectory tensor is constructed and combined with the t-SVD to jointly characterize the global low-rank features of HSI. The arrangement of similar pixels within the trajectory tensor gives it a strong rank-1 property. The t-SVD subsequently solves quickly for singular values in the Fourier

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3272669

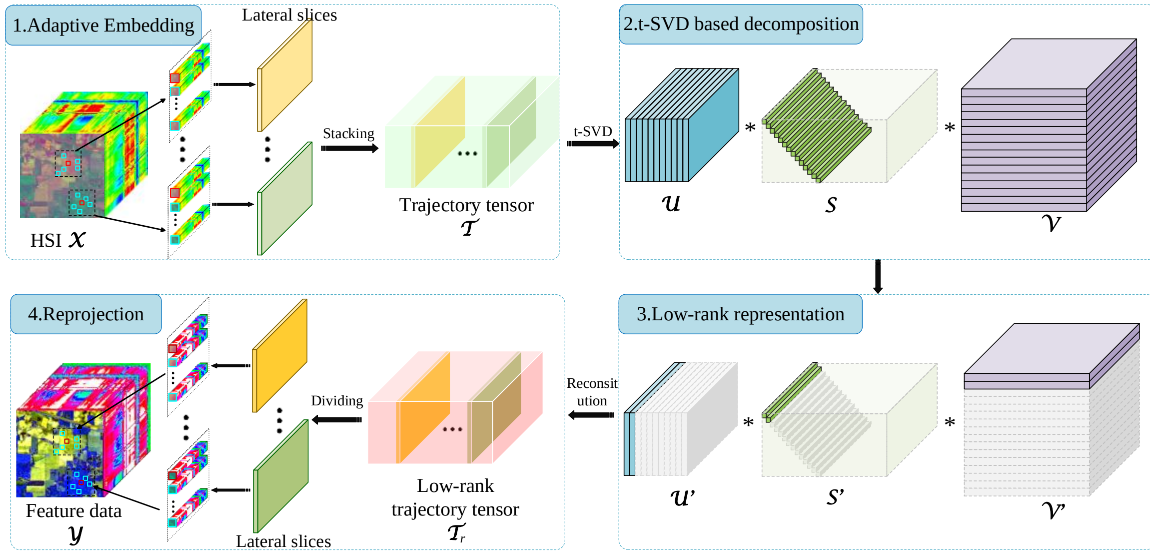


Fig. 1 Illustration of TensorSSA. It contains four stages: 1) Adaptive embedding, 2) t-SVD based decomposition, 3) Low-rank representation, and 4) Reprojection.

transform domain and gives the best approximation of the trajectory tensor by truncation, resulting in low-rank features.

The remainder of this article is organized as follows. The principles of the tensors and SSA are reviewed in Section II. Section III introduces the main steps of TensorSSA. The experimental results and analysis are presented in Section IV. Section V discusses the parameters and characteristics of TensorSSA. The concluding remarks are provided in Section VI.

II. Related Works

A. Notations and Definitions of tensors

We use italic letters to denote scalars (e.g., x and X), boldface lowercase letters for vectors (e.g., $\mathbf{x}$), boldface capital letters for matrices (e.g., $\mathbf{X}$), and calligraphic letters for tensors (e.g., $\mathcal{X}$). $\mathbb{R}$ denotes the real number fields. $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$ denotes a third-order tensor, in which the MATLAB notation $\mathcal{A}(:, :, i)$, $\mathcal{A}(:, i, :)$, and $\mathcal{A}(i, :, :)$ are used for its i-th frontal, lateral, and horizontal slice. Each dimension (way) is called a *mode*. The definitions related to our work are as follows.

Definition 1 (*t-product* [54]): For two third-order tensors $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2\times n_4\times n_3}$. The t-product of $\mathcal{A} * \mathcal{B}$ forms a tensor $\mathcal{C} \in \mathbb{R}^{n_1\times n_4\times n_3}$.

$$\mathcal{C}(i, j, :) = \sum_{k=1}^{n_2} \mathcal{A}(i, k, :) * \mathcal{B}(k, j, :) \tag{1}$$

Definition 2 (*Tensor transpose* [54]): For tensor $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$, transpose tensor $\mathcal{A}^{\mathrm{T}} \in \mathbb{R}^{n_2\times n_1\times n_3}$ is obtained by transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through n_3.

Definition 3 (*Identity Tensor* [54]): The identity tensor $\mathcal{J} \in \mathbb{R}^{n_1\times n_1\times n_3}$ is a tensor whose first frontal slice is the $n_1\times n_1$ identity matrix and all other frontal slices are zero.

Definition 4 (*f-diagonal tensor* [54]): A tensor is called *f*-diagonal if each frontal slice of the tensor is a diagonal matrix.

Definition 5 (*Orthogonal Tensor* [54]): A tensor $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is orthogonal if

$$\mathcal{A}^{\mathrm{T}} * \mathcal{A} = \mathcal{A} * \mathcal{A}^{\mathrm{T}} = \mathcal{J} \tag{2}$$

Definition 6 (t-SVD [54]*)*: As one of high-order SVD, for the tensor $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$, the t-SVD of $\mathcal{A}$ is given by

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V} \tag{3}$$

where $\mathcal{S} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is a rectangular *f*-diagonal tensor. $\mathcal{U} \in \mathbb{R}^{n_1\times n_1\times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2\times n_2\times n_3}$ are orthogonal tensors, and * denotes the t-product.

The tensor tubal rank of $\mathcal{A}$, denoted as $r_{tubal}(\mathcal{A})$, is defined as the number of nonzero singular tubes of $\mathcal{S}$ in the t-SVD factorization, i.e.,

$$r_{tubal}(\mathcal{A}) = \#\{i, \mathcal{S}(i, i, :) \neq 0\} \tag{4}$$

Note that t-SVD is achieved by computing the matrix SVD in the Fourier domain, which can enhance the mathematical solving efficiency [55].

The widely used CP and Tucker decompositions both have several disadvantages. Specifically, the rank-1 component of the CP decomposition is not easy to determine, and the computation of the approximation is numerically unstable for a fixed rank. Tucker decomposition can be seen as a generalization of the CP decomposition, where the truncated decomposition does not give the best fit to the original tensor. In contrast, t-SVD can be easily calculated by solving multiple singular values in the Fourier domain, and gives an optimal approximation of the tensor measured by the Frobenious norm of the difference, as stated in [56].

B. SSA and 2DSSA

Singular spectrum analysis (SSA) and its extensions have been applied to hyperspectral feature extraction successfully, including SSA for spectral pixels and 2DSSA for spatial bands [50, 51]. Both of contain four stages, i.e., embedding, singular value decomposition (SVD), grouping, and reprojection. The detailed steps of SSA are given as follows.

1) *Embedding*: Given a spectral pixel **p**= $[p_1, p_2, \ldots, p_n] \in \mathbb{R}^n$, for a given embedding window of size $l \in [1, n]$, the trajectory matrix **X** can be calculated as

$$\mathbf{X} = \begin{bmatrix} p_1 & p_2 & \cdots & p_{n-l+1} \\ p_2 & p_3 & \cdots & p_{n-l+2} \\ \vdots & \vdots & \ddots & \vdots \\ p_l & p_{l+1} & \cdots & p_n \end{bmatrix} \in \mathbb{R}^{l\times(n-l+1)} \tag{5}$$

Note that the matrix **X** is a Hankel matrix with the same antiangular elements.

2) *SVD*: In this step, SVD is applied to the matrix **X** to obtain the left singular vectors $(u_1, u_2, \ldots, u_l)$ and right singular vector $(v_1, v_2, \ldots, v_l)$, as well as the singular values $(\lambda_1, \lambda_2, \ldots, \lambda_l)$. From this **X** can be written as:

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots \mathbf{X}_l \; (\mathbf{X}_i = \lambda_i u_i v_i) \tag{6}$$

3) *Grouping*: A subset composed of one or more matrices $\mathbf{X}_i$ is selected to obtain the reconstructed matrix $\mathbf{X}_t \in \mathbb{R}^{l\times(n-l+1)}$. In general, these $\mathbf{X}_i$ correspond to large singular values because they usually contain more information.

4) *Reprojection*: The reconstructed matrix $\mathbf{X}_t$ is reprojected to a new spectral vector of $n\times 1$ again, i.e., enhanced pixel. The reprojection is the diagonal averaging in the matrix antidiagonals. More details can be found in [50].

2DSSA has the same operation in SVD and grouping, while differs in embedding and reprojection steps. In 2D embedding stage, a two-dimensional embedding window is defined to construct the trajectory matrix corresponding to the input band image, the obtained trajectory matrix has a structure called HbH, i.e., Hankel by Hankel. Correspondingly, in the reprojection stage, a two-step diagonal averaging process in the matrix antidiagonals in both each block and between blocks is required to reproject the reconstructed matrix to the image size [13, 52].

SSA-based methods, especially 2DSSA, usually utilize regular embedding windows (or extraction scales) to extract local information and cannot be adaptive to the irregular shapes and inconsistent sizes of the ground objects. This characteristic makes SVD less compressible, and the spatial features corresponding to the maximum singular value lose more details and edge information [57]. In addition, 2DSSA mainly acts on the spatial domain, ignoring the full use of spectral information to model the 3D structure of the HSI. However, considering the characteristic that SSA can model the global through the local, the current problem is expected to be solved if the SSA method is extended to 3D structures.

III. The Proposed TensorSSA Method

To extract 3D features of HSI, we designed a new SSA-based 3D feature extraction framework, called TensorSSA. It contains four steps: 1) Adaptive embedding, 2) T-SVD based decomposition, 3) Low-rank representation, and 4) Reprojection. The overall architecture of our method is shown in Fig. 1. In TensorSSA, an input HSI data is denoted by a 3D tensor $\mathcal{X} \in \mathbb{R}^{W\times H\times B}$, in which W, H, and B represent the wide, the height, and the number of bands, respectively. The detailed process of the proposed TensorSSA is given as follows.

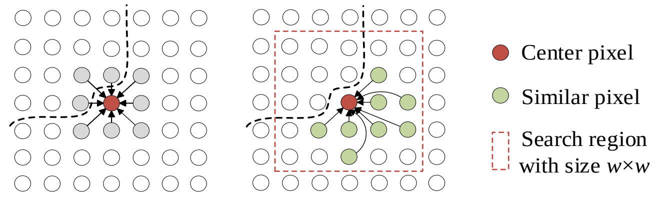


Fig. 2 Comparison of embedding. (a) 2DSSA embedding. (b) adaptive embedding of TensorSSA.

A. Adaptive Embedding

The first step in the construction of a 3D SSA model is to embed spectral and spatial information jointly, where the use of spatial information is particularly important. Most spatial methods including 2DSSA usually use fixed rectangle windows for feature extraction as shown in Fig. 2(a). Obviously, this is not appropriate for various objects with irregular shapes and inconsistent sizes. In other words, the regions used for spatial feature extraction should be adaptive to individual spatial structures of the HSI [52]. Nevertheless, spatial self-similarity is common in HSI, and this similarity has been well characterized as low rankness in tensors [56, 58, 59]. Motivated by this, a novel spatial similarity-based adaptive embedding window is proposed in TensorSSA, which is more flexible than 2DSSA as shown in Fig. 2(b).

For a certain pixel $\mathbf{x}_i \in \mathbb{R}^{B\times 1} (i = 1, 2, \ldots, WH)$ to be processed, the $w\times w$ size search region centered on it is firstly determined. Then, we obtain an adaptive embedding window with L pixels (including $\mathbf{x}_i$), by searching the (L-1) pixels with high similarity to $\mathbf{x}_i$ and then constructing a matrix $\mathbf{M}_i \in \mathbb{R}^{L\times B}$. The spectral similarity metric used here is the classical Normalized Euclidean distance (NED), which is a simple and effective measure of spectral similarity and is insensitive to the data scale [60]. Once the matrix $\mathbf{M}_i$ is twisted as a lateral slice $\mathcal{T}(:, i, :)$ of the trajectory tensor, the entire HSI $\mathcal{X}$ can be transformed into the trajectory tensor $\mathcal{T}$ as follows:

$$\mathcal{T} = \sum_{i=1}^{WH} \mathcal{T}(:, i, :) = \{\mathbf{M}_1, \ldots, \mathbf{M}_i, \ldots, \mathbf{M}_{WH}\} \in \mathbb{R}^{L\times WH\times B} \tag{7}$$

The obtained trajectory tensor $\mathcal{T}$ has several characteristics: Firstly, it contains both spectral and spatial information corresponding to the entire HSI. Secondly, its frontal slices are quasi-Hankel matrices [61], where only a part of the elements are the same in the matrix antidiagonals, which is completely different from the 2DSSA trajectory matrix. More importantly, the trajectory tensor $\mathcal{T}$ has the property of low rank due to the pixel similarity on *mode*-1and the high correlation among the spectral bands on *mode*-3. Note that the search region determines the size of the embedding spatial domain in this stage, and L can be set to any size smaller than $w\times w$. Details of the parameters are analyzed in section V. A.

B. T-SVD based decomposition

The obtained trajectory tensor $\mathcal{T}$ preserves and enhances the low-rank characteristics of $\mathcal{X}$, which reflects the main 3D spectral-spatial characteristics of HSI. In this stage, t-SVD is used to decompose the trajectory tensor to obtain the intrinsic

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3272669

characteristics of HSI. For the input tensor $\mathcal{T}$, the Fourier transform is firstly applied on *mode*-3, which is described in Eq. (8) as follows:

$$\mathcal{D} = fft\ (\mathcal{T}, [], 3) \quad (8)$$

where $fft\ (\cdot, [], 3)$ denotes the discrete Fourier transformation (DFT) along the third mode of a 3-way tensor. $\mathcal{D}$ represents the transformation tensor of $\mathcal{T}$.

Then, SVD is applied on the frontal slice of tensor $\mathcal{D}$ to obtain the singular vectors and singular values according to Eqs. (9) and (10) as follows:

$$[\mathbf{U}_i, \mathbf{S}_i, \mathbf{V}_i] = SVD\ (\mathcal{D}(:, :, i)\ ).\ i=1, \ldots, B \quad (9)$$

$$\overline{\mathcal{U}}\ (:, :, i) = \mathbf{U}_i;\ \ \overline{\mathcal{S}}\ (:, :, i) = \mathbf{S}_i;\ \ \overline{\mathcal{V}}\ (:, :, i) = \mathbf{V}_i \quad (10)$$

where $\mathbf{U}_i$, $\mathbf{S}_i$, and $\mathbf{V}_i$ represent the left singular vector matrix, singular value matrix, and right singular vector matrix of $\mathcal{D}(:, :, i)$, respectively. These matrices form the tensor $\overline{\mathcal{U}} \in \mathbb{R}^{L\times L\times B}$, $\overline{\mathcal{S}} \in \mathbb{R}^{L\times WH\times B}$, and $\overline{\mathcal{V}} \in \mathbb{R}^{WH\times WH\times B}$, respectively.

Finally, the obtained tensors are transformed to the real number domain from the Fourier domain by the fast inverse discrete Fourier transformation (IFFT), which is defined as Eq. (11).

$$\mathcal{U} = ifft(\overline{\mathcal{U}}, [], 3);\ \ \mathcal{S} = ifft(\overline{\mathcal{S}}, [], 3);\ \ \mathcal{V} = ifft(\overline{\mathcal{V}}, [], 3) \quad (11)$$

where $ifft(\cdot, [], 3)$ denotes the IFFT along the *mode*-3 of a tensor. The obtained $\mathcal{U}$ and $\mathcal{V}$ are orthogonal singular tensors and $\mathcal{S}$ is the singular tuples. The t-SVD decomposition of the tensor $\mathcal{T}$ is illustrated in Fig. 1(2).

C. Low-rank representation

The feature tensors $\mathcal{U}$, $\mathcal{S}$ and $\mathcal{V}$ obtained by t-SVD decomposition are low rank in different modes, while they also contain some useless information. For effective low-rank representation, we further truncate these feature tensors to approximate the original trajectory tensor $\mathcal{T}$ optimally, i.e., truncated t-SVD [54]. To this end, the ideal tensor tube rank r_{tubal} is defined, which satisfies Eq. (12)

$$r_{tubal}\ (\mathcal{T}) << \min\ (L, WH) \quad (12)$$

Then, to further exploit the low-rank and sparse characteristic of the third-order tensor, the original $\mathcal{U}$, $\mathcal{S}$, and $\mathcal{V}$ are simplified through the tubal rank interception as shown in Eq. (13) as follows:

$$\begin{aligned} \mathcal{U}' &= \mathcal{U}\ (: , 1: r_{tubal}, :) \\ \mathcal{S}' &= \mathcal{S}\ (1: r_{tubal}, 1: r_{tubal}, :) \\ \mathcal{V}' &= \mathcal{V}\ (: , 1: r_{tubal}, :) \end{aligned} \quad (13)$$

After tubal rank simplification, the new trajectory tensor $\mathcal{T}_r$ is obtained according to Eq. (14):

$$\mathcal{T}_r = \mathcal{U}' * \mathcal{S}' * \mathcal{V}'^{\mathrm{T}} \in \mathbb{R}^{L\times WH\times B} \quad (14)$$

where $*$ denotes the t-product here. Tensor $\mathcal{T}_r$ is a low-rank tensor, and it can be regarded as the rank-r approximation of the original $\mathcal{T}$. In this stage, the parameter r_{tubal} determines the amount of information used for tensor reconstruction.

D. Reprojection

The obtained reconstruction tensor $\mathcal{T}_r$ contains redundant information in the *mode*-2 direction, and its frontal slice is no longer the necessarily quasi-Hankel matrix. Therefore, it is necessary to reduce the tensor redundancy through averaging, essentially the same as the diagonal average of SSA. To this end, the low-rank trajectory tensor $\mathcal{T}_r$ is reprojected to a new tensor of size $W\times H\times B$, defined as the feature data of HSI.

Each lateral slice of tensor $\mathcal{T}_r$ is firstly squeezed into a matrix and each matrix corresponds to a processed spectral pixel with its similar pixels. Then, these pixels are relocated to their original spatial positions. Considering that pixels in certain positions are selected multiple times during the embedding process, the average value of these pixels is taken as the final pixel value for multiple pixel values at the same position. It can be written as Eq. (13):

$$\mathcal{Y} \leftarrow Rp\ (\mathcal{T}_r) \quad (13)$$

where $Rp(\cdot)$ denotes the reprojection operation. Tensor $\mathcal{Y} \in \mathbb{R}^{W\times H\times B}$ is the obtained feature data corresponding to $\mathcal{X}$.

In HSI, the homogeneous area of the image usually has more similar pixels, while the non-homogeneous area, such as the edges, has fewer similar pixels. For TensorSSA, the number of selections of similar pixels in the homogeneous area is large while that of the heterogeneous area is small in TensorSSA. Therefore, the average degree of the homogeneous area is higher during the reprojection which improves the spectral consistency within the ground object, while heterogeneous regions are lower which preserves the differences between the classes of ground objects. The obtained feature data has several characteristics: in the spectral domain, the main spectral discrimination trends are retained and spectral oscillations (noise) are eliminated; in the spatial domain, intra-class variability is reduced and inter-class separability is improved while noise is eliminated. Section V. B gives more presentations and discussions. The code of this work is available at https://github.com/Hang-Fu/TensorSSA.

IV. Experimental Results and Analysis

A. Datasets

Three hyperspectral datasets, including the Indian Pines (IP), Pavia University (PU), and MUUFL Gulfport (MG) are used to evaluate the performance of our proposed method.

1) IP [1]: The well-known IP dataset covers Northwest Indiana, USA, which was acquired by the AVIRIS sensor with a spectral range from 0.4 to 2.5μm. It has a scene with 145 × 145 pixels with a spatial resolution of 20 m per pixel and 220 spectral bands. In the experiment, the number of bands is reduced to 200 by removing 20 water absorption bands.

2) PU [1]: The PU dataset was captured through a Reflective Optics Spectrographic Imaging System (ROSIS), flying over the city of Pavia, Italy, with a spectral range from 0.43 to 0.86μm. It contains 103 bands of size 610 × 340 pixels with a spatial resolution of 1.3 m per pixel.

3) MG [2]: It was acquired over the campus of the University of Southern Mississippi Gulf Park, Long Beach, Mississippi. It originally contains 72 bands while due to noise, the first four and last four bands are omitted, bringing about an image with 64 bands. The spatial resolution is 1 m per pixel. The

[1] https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

[2] https://zenodo.org/record/1186326

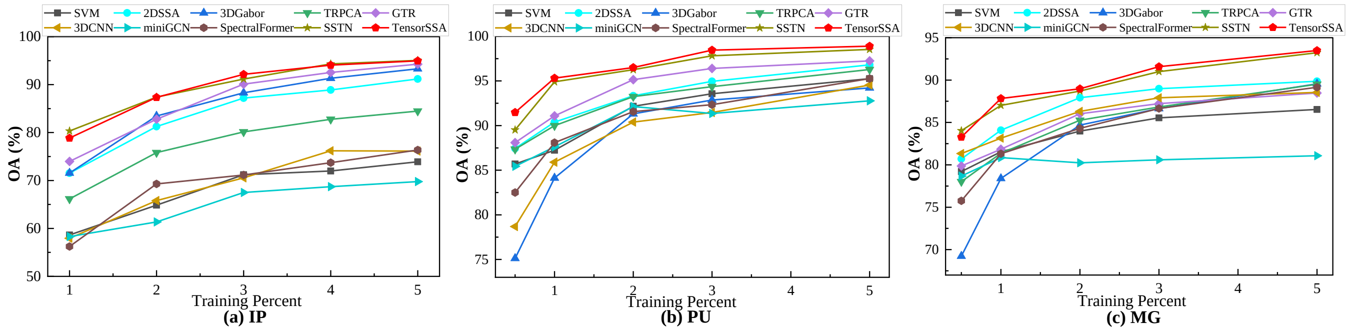


Fig. 3 OA obtained by different methods with different training percentages over (a) IP, (b) PU, and (c) MG datasets.

hyperspectral scene has 325 × 337 pixels, and the provided ground truth map includes 11 classes [62].

B. *Experimental Setup*

The classification tasks are used to validate the effectiveness of TensorSSA. As to evaluation metrics, five objective quality indexes, i.e., the producer accuracy (PA) of each class, overall accuracy (OA), average accuracy (AA), kappa coefficient (Kappa), and running time (s), are utilized in the following experiments. To avoid systematic errors and biased estimation, all experiments were conducted five times independently, both the average values and standard deviations are listed in the experiments.

In order to evaluate the proposed method, we compare it with 9 state-of-the-art algorithms, including conventional 2DSSA [51], 3DGabor [63], two tensor methods, i.e., tensor robust principal component analysis (TRPCA) [64] and GTR [34], and four deep learning methods, i.e., 3DCNN [65], miniGCN [45], Morphological Convolutional Neural Networks (MorphCNN) [66], SpectralFormer [48] and SSTN [49]. The details of the compared methods are listed as follows.

1) Support vector machine (SVM) classifier [67] with radial basis function (RBF) kernel on raw HSI data as baseline spectral method, in which fivefold cross-validation is utilized to determine hyper-parameters.

2) 2DSSA [51] as the spatial method, whose embedding window and reconstruction parameters of 2DSSA are 5×5 and 1, respectively.

3) 3DGabor as the 3D operator method, KPCA is firstly used for dimension reduction, and then a set of Gabor wavelets with parameters ρ = [0.5, 0.25, 0.125, 0.0625] and $\vartheta = 0$, $\phi = 0$ are used [63], and SVM used as classifier.

4) TRPCA [64] as the tensor-based feature extraction method, in which block size is set as 24×24 and the number of iterations is 500, SVM is utilized as the classifier.

5) GTR as the tensor classification method and the optimal parameters from [34] are used.

6) 3DCNN, the 3D deep network consists of two 3D convolution blocks (a 3D convolution layer, a batch re-normalization layer, and a ReLU function) and two full connection layers [65].

7) miniGCN, this lightweight GCN uses the training parameters of [45]. Note that PU data needs to be chunked to prevent out-of-memory.

8) MorphCNN [66] combines morphological operators and convolutional kernels, in which B/4 dilations, B/4 erosions and 3 × 3 kernels are used for three datasets (B is the number of bands).

9) SpectralFormer [48] as the first transformer network for HSI, the patch-wise version is used and the patch size is 7×7. The batch size is 64 and epochs are 300 for three datasets.

10) SSTN, the batch size is 32 and epochs are 100 for three datasets. The other hyperparameters remain consistent with [49].

11) For the proposed TensorSSA, the optimal parameters are given in section V. A for three datasets, and the SVM classifier is also used for classification.

All conventional methods are implemented in MATLAB 2021a, and the networks are implemented using the Tensorflow and Pytorch frameworks in PyCharm on Windows 10 machines with an NVIDIA GeForce RTX 3060 GPU.

C. *Classification Results with random Samples*

It is important to analyze the classification performance obtained using randomly selected and varying training sets. Accordingly, the classification results of TensorSSA and the other seven compared methods using random samples are presented for the IP, PU, and MG datasets. The details are given as follows.

1) *Performance with Different Training Percentages*: Fig. 3 shows the OAs obtained by different methods for three datasets using different training percentages. Specifically, the randomly selected training samples vary within {1%, 2%, 3%, 4%, 5%} for IP and {0.5%, 1%, 2%, 3%, 5%} for PU and MG datasets, the remaining samples are used for testing.

According to Fig. 3, with the increase of training samples, the accuracy of all comparison methods has been improved to a certain extent. Among them, the proposed TensorSSA method utterly outperforms all other 9 methods, even when the sample size is small, e.g., 1% from PU, and MG. The state-of-the-art SSTN achieved sub-optimal results on three datasets, outperforming TensorSSA in some cases. The performance of TensorSSA is comparable to that of the most advanced deep learning methods. In addition, the performance of the other 8 methods on different data varies. For example, GTR has achieved accuracy second only to TensorSSA and SSTN on IP and PU datasets, while classification performance was poor on the MG dataset. This proves the robustness of the TensorSSA method again.

2) *Quantitative Evaluation*: In order to detail evaluate the superiority of the proposed method, the quantitative results of PA of each class, OA, AA, and Kappa are listed in Tables I–III. We can see that TensorSSA achieves the highest classification

TABLE I
CLASSIFICATION RESULTS OBTAINED BY DIFFERENT METHODS FOR THE IP DATASET (2% TRAINING PERCENTAGE)

Class	Samples	SVM	2DSSA	3DGabor	TRPCA	GTR	3DCNN	miniGCN	MorphCNN	SpectralFormer	SSTN	TensorSSA
Alfalfa	46	6.67(2.23)	38.52(1.28)	48.89(6.67)	57.04(11.4)	60.00(15.6)	4.44(2.22)	14.07(3.39)	12.59(5.59)	41.67(6.94)	50.00(20.3)	**93.20(3.97)**
Corn-N	1428	58.54(10.1)	74.77(3.15)	79.51(3.45)	72.69(3.00)	75.98(1.18)	58.99(4.36)	66.26(8.43)	75.48(8.54)	64.52(4.96)	81.48(3.41)	**86.16(0.74)**
Corn-M	830	45.47(7.59)	79.70(7.01)	78.64(7.56)	61.58(7.01)	82.37(1.33)	46.95(2.63)	41.16(12.6)	67.94(5.93)	56.23(9.89)	**94.46(1.11)**	89.68(1.60)
Corn	237	16.09(4.81)	69.97(5.00)	63.36(9.63)	42.96(13.6)	72.27(10.7)	16.52(5.18)	23.56(7.62)	17.82(5.05)	30.42(6.91)	**86.88(6.53)**	74.22(3.09)
Grass-P	483	67.58(1.05)	72.59(4.99)	75.62(7.31)	79.35(2.12)	**88.94(2.32)**	43.76(1.18)	62.16(1.52)	62.79(10.8)	71.15(2.06)	86.17(4.56)	84.73(2.50)
Grass-T	730	88.16(5.73)	95.80(3.18)	84.80(2.41)	91.52(6.14)	86.62(4.89)	86.99(1.56)	89.42(3.61)	76.60(3.86)	80.46(1.27)	**96.62(1.73)**	95.82(1.28)
Grass-PM	28	30.86(27.8)	96.30(3.70)	77.78(6.42)	80.25(8.55)	86.42(14.0)	16.05(13.0)	77.78(12.8)	69.14(7.71)	22.22(19.3)	48.71(10.3)	**93.83(4.28)**
Hay-W	478	90.17(3.15)	90.10(9.00)	93.59(3.56)	97.44(1.11)	88.82(0.96)	**98.01(1.50)**	68.16(15.0)	73.86(6.55)	88.34(6.72)	85.97(1.47)	96.35(3.28)
Oats	20	17.54(13.3)	43.86(8.04)	52.36(22.9)	33.33(8.04)	42.11(9.12)	14.04(10.9)	31.58(13.9)	33.33(10.9)	12.28(12.2)	50.00(30.6)	**100.0(0.00)**
Soybean-N	972	57.25(11.5)	74.61(2.67)	78.29(2.41)	74.61(7.52)	74.30(3.53)	56.20(5.62)	63.48(2.92)	78.36(5.71)	72.82(2.96)	**89.43(1.84)**	81.27(10.8)
Soybean-M	2455	71.02(9.94)	88.40(2.22)	91.60(3.21)	79.93(3.87)	84.88(5.68)	76.42(0.52)	50.98(5.31)	**95.47(2.06)**	74.91(4.16)	90.07(6.01)	90.01(2.04)
Soybean-C	593	28.97(3.74)	49.51(4.84)	74.64(9.08)	48.14(3.930	69.65(7.25)	37.35(4.80)	40.16(7.76)	48.31(2.95)	35.42(5.59)	76.80(9.73)	**80.18(8.14)**
Wheat	205	92.33(3.18)	96.00(2.65)	89.33(7.75)	93.83(3.01)	93.83(1.89)	71.83(16.9)	93.83(4.65)	87.83(4.25)	82.49(17.3)	**98.30(1.69)**	97.32(1.04)
Woods	1265	94.49(1.53)	93.41(3.60)	90.66(4.11)	93.62(1.80)	96.15(1.42)	93.14(1.61)	88.65(7.55)	89.35(5.58)	87.37(2.91)	89.53(6.28)	**98.95(1.12)**
Buildings	386	20.90(1.65)	63.84(13.5)	66.93(8.21)	41.45(4.80)	73.72(17.4)	35.71(5.16)	35.01(5.61)	58.55(8.41)	45.45(3.94)	**85.49(6.65)**	79.13(13.2)
Stone	93	45.42(6.72)	90.48(6.25)	93.04(4.58)	41.03(17.6)	91.94(9.35)	44.32(6.34)	80.59(16.8)	7.69(3.96)	54.81(25.5)	93.34(2.44)	**98.89(0.01)**
OA		64.86(1.11)	81.25(0.25)	83.44(0.82)	75.83(0.52)	82.78(1.62)	65.81(0.76)	61.35(0.35)	76.76(0.57)	69.28(1.05)	88.39(0.15)	**89.11(0.66)**
AA		51.97(1.64)	76.12(1.00)	77.46(1.77)	68.05(2.02)	79.25(1.11)	50.05(3.11)	57.93(0.52)	59.69(1.61)	57.54(0.49)	85.45(7.86)	**89.98(0.74)**
Kappa×100		59.59(0.92)	78.58(0.28)	81.06(0.94)	72.32(0.50)	80.32(1.81)	60.51(0.89)	56.31(0.44)	72.55(0.74)	64.89(1.20)	86.30(0.26)	**87.60(0.74)**
time(s)		**5.61**	11.29	13.26	11.36	9.24	287.87	16.17	19.21	349.11	30.91	14.10

TABLE II
CLASSIFICATION RESULTS OBTAINED BY DIFFERENT METHODS FOR THE PU DATASET (1% TRAINING PERCENTAGE)

Class	Samples	SVM	2DSSA	3DGabor	TRPCA	GTR	3DCNN	miniGCN	MorphCNN	SpectralFormer	SSTN	TensorSSA
Asphalt	6631	53.66(10.6)	71.23(11.4)	58.14(1.81)	68.22(5.04)	44.36(18.9)	27.13(12.3)	51.42(5.34)	95.60(0.95)	86.70(7.07)	**98.42(0.07)**	84.67(0.75)
Meadows	18649	77.58(16.8)	**96.97(5.25)**	96.36(4.81)	80.00(14.3)	52.42(22.6)	70.91(33.2)	72.12(6.94)	**98.35(0.97)**	96.88(1.18)	95.09(0.71)	91.52(14.7)
Gravel	2099	58.30(8.98)	65.58(2.94)	32.12(21.1)	68.12(19.3)	30.18(7.98)	38.30(20.2)	84.00(18.0)	72.44(6.29)	59.91(5.59)	**97.62(0.64)**	89.45(11.3)
Trees	3064	83.57(13.1)	81.04(7.40)	21.33(6.22)	88.47(12.7)	90.68(11.6)	75.83(15.2)	87.84(3.95)	94.89(1.21)	91.53(7.00)	**99.07(0.45)**	97.31(2.19)
Painted-MS	1345	27.85(16.2)	23.64(8.24)	19.41(2.92)	46.41(5.12)	84.81(21.9)	38.82(21.4)	**99.58(0.73)**	90.06(4.56)	98.54(0.43)	99.47(0.44)	67.51(0.73)
Bare Soil	5029	94.49(2.07)	97.69(1.88)	97.93(0.35)	95.56(2.51)	**99.93(0.09)**	97.76(2.61)	92.23(2.44)	94.68(1.34)	68.15(7.08)	95.36(0.12)	99.77(0.17)
Bitumen	1330	91.16(2.61)	93.52(2.46)	92.30(1.60)	95.06(0.61)	**99.42(0.61)**	93.44(0.65)	92.50(5.19)	73.51(5.45)	76.52(1.85)	96.53(0.86)	95.34(0.04)
SB-Bricks	3682	81.99(15.8)	77.02(9.67)	76.68(13.7)	77.02(14.6)	**91.80(2.16)**	75.67(13.7)	72.31(5.50)	89.59(3.70)	84.70(1.30)	82.40(2.74)	83.40(11.8)
Shadows	947	99.73(0.47)	95.66(5.84)	54.74(14.1)	99.32(0.23)	97.97(1.08)	97.02(2.24)	99.86(0.23)	76.81(7.06)	96.36(0.54)	**98.01(1.13)**	97.02(0.23)
OA		87.23(0.44)	90.39(0.21)	84.13(1.05)	89.99(0.94)	91.08(0.95)	85.90(1.78)	87.62(1.32)	93.71(0.40)	88.09(3.30)	94.90(0.12)	**95.89(0.33)**
AA		74.26(1.46)	78.04(1.75)	61.00(1.93)	79.80(4.09)	76.84(2.24)	68.22(2.77)	83.54(2.19)	87.32(0.24)	84.37(3.15)	**95.78(0.16)**	94.55(1.68)
Kappa×100		81.73(0.55)	86.11(0.25)	75.81(1.70)	85.64(1.48)	87.09(1.40)	79.26(2.82)	82.52(1.85)	91.63(0.52)	84.03(4.45)	93.18(0.13)	**93.65(0.30)**
time(s)		**2.85**	22.84	10.75	42.08	5.04	539.68	126.09	49.55	562.69	70.33	18.23

TABLE III
CLASSIFICATION RESULTS OBTAINED BY DIFFERENT METHODS FOR THE MG DATASET (1% TRAINING PERCENTAGE)

Class	Samples	SVM	2DSSA	3DGabor	TRPCA	GTR	3DCNN	miniGCN	MorphCNN	SpectralFormer	SSTN	TensorSSA
Trees	23246	92.79(0.97)	94.69(1.01)	89.53(0.90)	93.79(1.08)	**96.06(0.28)**	94.12(1.06)	92.60(0.18)	95.37(1.12)	95.26(0.04)	95.40(0.03)	95.40(0.55)
Grass	4270	62.86(3.09)	69.83(4.35)	68.73(1.76)	65.79(2.54)	75.59(2.14)	63.77(2.92)	70.74(0.66)	75.74(4.60)	67.15(0.23)	83.10(3.66)	**83.10(4.75)**
MG-surface	6882	68.59(2.76)	76.65(5.37)	70.42(1.94)	73.59(3.90)	**82.31(1.26)**	73.02(4.88)	73.50(1.14)	78.22(3.12)	66.73(0.21)	80.18(0.02)	80.18(3.52)
Dirt/sand	1826	74.27(11.3)	72.27(8.27)	74.76(6.63)	71.52(12.2)	67.37(2.92)	72.27(3.27)	67.48(2.97)	67.77(9.29)	78.12(0.06)	78.62(0.43)	**78.62(3.78)**
Road	6687	87.66(2.60)	88.85(2.28)	73.26(0.98)	88.56(1.94)	81.14(0.38)	83.93(3.04)	81.60(1.00)	90.85(1.77)	88.98(0.08)	91.79(0.87)	**91.79(2.03)**
Water	466	81.42(2.32)	81.78(5.08)	58.64(14.0)	**83.08(3.37)**	2.46(1.52)	31.16(10.5)	67.46(2.83)	55.53(17.3)	14.05(0.49)	82.36(3.83)	82.36(3.83)
B-shadow	2233	67.60(7.48)	60.47(4.53)	58.02(3.42)	67.48(2.55)	7.80(1.09)	68.52(5.36)	**78.56(1.32)**	72.88(10.4)	72.43(0.23)	67.33(2.86)	67.33(2.86)
Buildings	6240	79.82(2.44)	82.62(2.08)	84.38(2.40)	79.29(4.27)	82.40(1.71)	78.57(2.92)	74.23(0.18)	87.94(1.61)	66.54(0.24)	88.97(2.17)	**88.97(2.17)**
Sidewalk	1385	38.17(10.0)	37.98(3.64)	24.43(2.79)	52.22(5.56)	5.52(5.23)	32.97(8.14)	51.39(2.82)	**59.08(8.36)**	46.04(0.44)	48.36(13.7)	48.36(13.7)
Y-curb	183	21.36(8.01)	13.08(7.74)	3.13(1.69)	**23.02(11.9)**	12.33(0.23)	4.74(3.68)	19.11(3.47)	14.18(8.81)	7.21(0.95)	6.63(3.45)	6.63(3.45)
Cloth-P	269	54.39(11.3)	67.29(15.7)	18.30(4.02)	48.25(7.83)	46.12(2.50)	27.07(8.14)	71.76(2.66)	29.20(6.82)	69.10(0.38)	80.95(14.5)	**80.95(14.5)**
OA		81.54(0.52)	84.09(0.24)	78.39(0.50)	83.15(0.90)	80.85(0.32)	81.36(0.63)	81.84(0.27)	86.42(0.58)	81.31(0.13)	87.02(0.41)	**87.85(0.27)**
AA		66.27(1.08)	67.77(2.20)	56.69(1.88)	67.87(1.62)	53.70(1.06)	57.29(1.43)	68.04(0.72)	66.07(2.60)	61.06(0.30)	82.88(1.26)	**83.06(1.57)**
Kappa×100		75.53(0.69)	78.90(0.26)	70.94(0.70)	77.62(1.21)	76.96(0.49)	75.17(0.88)	75.59(0.37)	81.77(0.81)	75.13(0.18)	83.32(0.53)	**83.86(0.38)**
time(s)		7.32	14.57	18.01	28.67	**3.32**	638.92	66.75	56.19	491.52	40.26	12.45

accuracy in terms of three metrics and most classes, especially in terms of IP and MG datasets. TensorSSA has the highest OA and Kappa but slightly lower AA than SSTN. Take Table I as an example for analysis, firstly, the OA of TensorSSA is improved by 24.25%, 7.86%, 5.67%, 13.28%, 6.33%, 23.3%, 27.76%, 2.18%, 19.83%, and 0.72% compared with SVM, 2DSSA, 3DGabor, TRPCA, GTR, 3DCNN, miniGCN, MorphCNN, SpectralFormer, and SSTN, which is unexpected accuracy improvements. Secondly, the PA of 14 of the 16 classes exceeded 80%, and that of 9 classes exceeded 90%, especially for the classes with fewer samples, such as “Alfalfa”, “Grass-PM” and “Oats”. Besides, in Tables II and III, the proposed method also has the same superior performance.

In terms of comparison methods, compared with the classic SVM, the two feature extraction methods 2DSSA and 3DGabor have higher classification accuracies due to the utilization of spatial context and texture features. Unfortunately, the accuracy of 3DGabor on PU and MG datasets is decreased, maybe because the texture features obtained by Gabor wavelets kernel neglect some small ground objects. 3DCNN, miniGCN, and SpectralFormer are weaker than SVM in some cases, mainly because of the small number of samples to train the network adequately. MorphCNN has better classification performance on PU and MG but poorer on IP datasets due to its ability to effectively identify irregular and broken features. SSTN has better performance because it fully exploits the deep spectral-spatial features. Two tensor-based methods, TRPCA and GTR can both extract low-rank information, but their overall performance is limited. TensorSSA has optimal performance on all datasets due to the exploration and utilization of effective 3D spectral-spatial features.

3) Qualitative Evaluation: The classification maps yielded by all methods alongside false color images and GT on three datasets are shown in Figs. 4–6, respectively. Note that, in our

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3272669

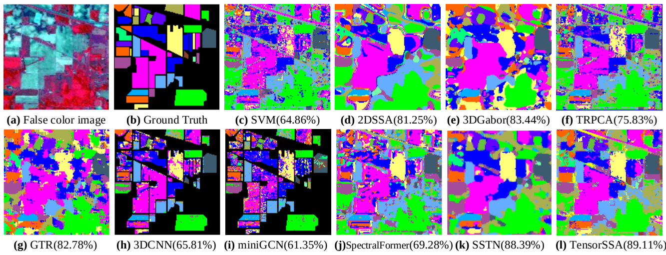


Fig. 4 (a) False color image from IP dataset. (b) Ground truth. Classification maps obtained by (c) SVM, (d) 2DSSA, (e) 3DGabor, (f) TRPCA, (g) GTR, (h) 3DCNN, (i) miniGCN, (j) SpectralFormer, (k) SSTN, and (l) TensorSSA.

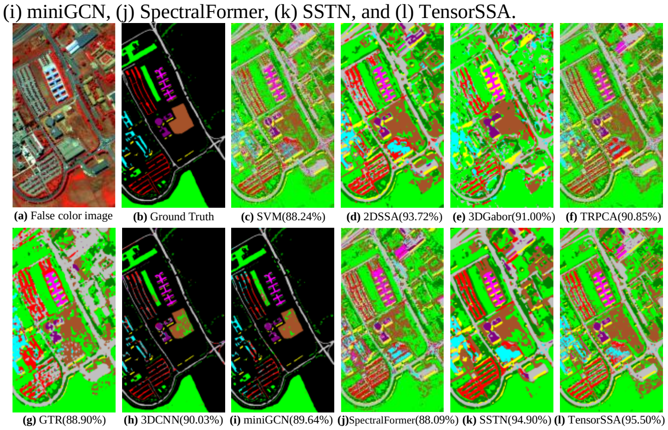


Fig. 5 (a) False color image from PU dataset. (b) Ground truth. Classification maps obtained by (c) SVM, (d) 2DSSA, (e) 3DGabor, (f) TRPCA, (g) GTR, (h) 3DCNN, (i) miniGCN, (j) SpectralFormer, (k) SSTN, and (l) TensorSSA.

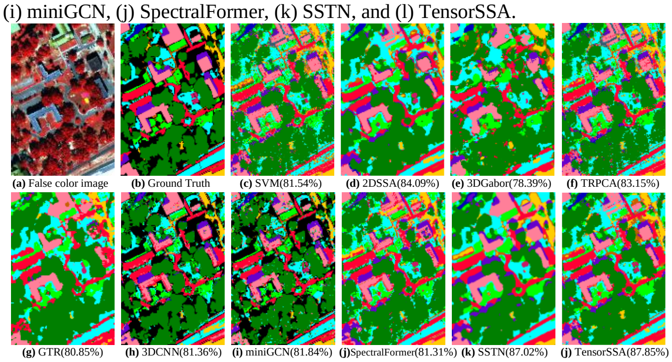


Fig. 6 (a) False color image from MG dataset. (b) Ground truth. Classification maps obtained by (c) SVM, (d) 2DSSA, (e) 3DGabor, (f) TRPCA, (g) GTR, (h) 3DCNN, (i) miniGCN, (j) SpectralFormer, (k) SSTN, and (l) TensorSSA.

experiments, the background is not included in the classification map for 3DCNN and miniGCN methods for some reason. As shown in Figs. 4–6, SVM, as well as miniGCN, appear obvious classification noise inside the land covers. 2DSSA, 3DGabor, and SpectralFormer can eliminate spot-like misclassification, but they cannot reasonably distinguish the boundaries of features and ignore some smaller features (such as Road). 3DCNN uses fixed patch blocks for feature extraction, and serious misclassification will occur in small or strip features. TRPCA has some misclassified plaques. As for GTR, it has lost many morphological features of ground features and misclassified seriously at the edges. The SSTN has curved feature edges. The proposed TensorSSA eliminates the internal noise and preserves the details of ground objects, tiny ground objects such as roads. In general, our method can obtain the classification map that is closest to the actual ground object distribution.

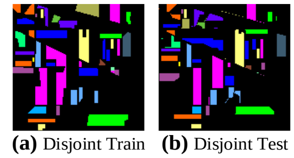


Fig. 7 Spatially disjoint training and testing samples of DIP.

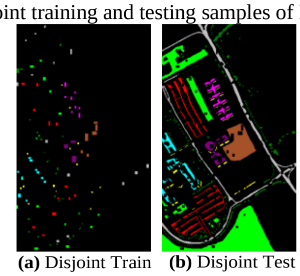


Fig. 8 Spatially disjoint training and testing samples of DPU.

4) Analysis of Running Time: Tables I–III also give the running time of all methods. As we can see, SVM and GTR have the highest running efficiency, while 2DCNN, 3DGabor, and TRPCA take more time due to feature extraction. 3DCNN mainly conducts model training based on patch blocks, which take the longest time. MiniGCN will take a long time on a larger dataset because it needs to calculate the adjacency matrix. Patch-wise SpectralFormer uses a large patch size to explore global features, leading to a long running time. The training time of SSTN is long and takes up a lot of time spent. On the contrary, TensorSSA is faster than most deep learning methods and two conventional Methods, i.e., 2DSSA, TRPCA, but slower than SVM, GTR and 3DGabor mainly because it takes more time to decompose the trajectory tensor, which is also what needs to be improved in the future.

D. Classification Results over Disjoint Samples

To further verify the effectiveness of the proposed method, the disjoint training samples have been considered. Compared with random sampling, disjoint samples usually acquire more realistic classification results and introduce certain challenges [13, 40]. In this section, two more challenging datasets that are publicly available from the GRSS DASE website [3], Disjoint IP (DIP) and Disjoint PU (DPU) are used for evaluation. The spatial disjoint training and testing samples for two datasets are shown in Fig. 7 and Fig. 8. Besides, some classical machine learning and representative deep learning methods available in [40] [4] are added as compared methods, including MLR, MLP, RNN, LSTM, GRU, 2DCNN, state-of-the-art hybrid spectral CNN (HybridSN) and morphological CNN (MorphCNN). Accordingly, all compared methods are divided into two groups: conventional methods and deep learning methods. The quantitative results in terms of OA, AA, and Kappa are given in Tables IV-V.

As seen in Tables IV-V, the proposed TensorSSA method outperforms all conventional and most deep learning methods, i.e., the highest OA and Kappa on DIP, second only to MorphCNN on DUP, achieving satisfactory classification accuracy. In the conventional methods group, three classifiers SVM, MLR, and GTR are mainly based on spectral features for classification, ignoring the potential role of spatial information, leading to limited classification performance. While 2DSSA

[3] http://dase.grss-ieee.org

[4] https://github.com/AnkurDeria/HSI-Traditional-to-Deep-Models

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3272669

TABLE VI
OA% (RUNNING TIME (S)) OF PARAMETERS L AND $w \times w$ ON THREE DATASETS

w×w	L of the IP dataset				L of the PU dataset				L of the MU dataset			
	9	25	49	81	9	25	49	81	9	25	49	81
5×5	80.11(3.32)	—	—	—	**95.52(18.42)**	—	—	—	86.75(4.45)	—	—	—
7×7	78.40(4.31)	86.03(8.19)	—	—	94.35(24.78)	96.34(54.91)	—	—	86.75(6.02)	**87.69(11.48)**	—	—
9×9	78.49(5.77)	86.09(9.66)	86.75(11.87)	—	94.38(34.55)	96.31(65.70)	96.10(98.37)	—	86.38(8.48)	87.02(13.80)	86.21(18.30)	—
11×11	78.00(7.55)	83.57(11.28)	**89.41(14.01)**	87.52(18.41)	94.96(45.73)	95.95(77.02)	96.56(101.8)	96.33(290.8)	86.48(11.34)	86.60(16.81)	86.50(21.06)	85.69(27.61)
13×13	77.17(9.51)	83.03(13.21)	88.60(15.83)	90.68(19.63)	94.23(59.80)	96.44(93.13)	96.40(114.8)	—	86.56(15.16)	86.48(20.69)	86.21(26.22)	86.32(31.46)

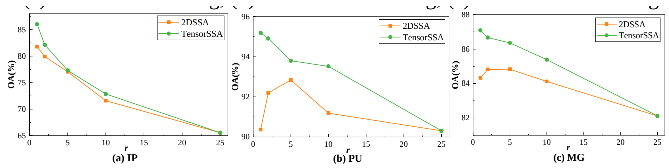


Fig. 10 Accuracy comparison of r_{tubal} and r between TensorSSA (L=25) and 2DSSA (5×5). OA obtained on (a) IP with 2% training, (b) PU with 1% training, (c) MG with 1% training.

and TRPCA have achieved higher accuracy mainly because the extracted spatial features make up for the lack of spectral information. A similar situation also appears in the deep learning method group. Compared with spectral deep models, including MLP, RNN, LSTM, GRU and miniGCN, 2DCNN, SpectralFormer and 3DCNN respectively extract spatial and spectral-spatial features of HSI for classification, achieving higher classification accuracy. Moreover, HybridSN, MorphCNN and SSTN exploit deep intrinsic features using a mixture of convolutional kernels, filter transforms and attention mechanisms to achieve the best classification performance in some cases. Generally, the performance of deep learning methods is superior to conventional methods mainly because their multi-layer network architecture can extract more deep semantic features for classification. TensorSSA effectively characterizes and extracts the spectral-spatial low-rank features of images based on spatial similarity and t-SVD. Compared with the deep model, our method can achieve better classification performance without training and construction of a multi-layer network.

V. DISCUSSION

A. Parameter Analysis

There are mainly two parameters, i.e., the embedding window size L and the low-rank component r_{tubal} in TensorSSA. The sensitivity analysis of the two parameters is as follows.

1) L: This parameter, as well as the search region determines the degree of utilization of spatial information. Accordingly, we design the experiments to evaluate the L of TensorSSA on three datasets.

First, we evaluated the L under a fixed search region (11×11 as an example), and 2DSSA is added for comparison to reinforce the claim of superiority and robustness of TensorSSA at the same time. According to the experience [13, 51], four embedding window sizes are selected, i.e., {3×3, 5×5, 7×7, 9×9} for 2DSSA and {9, 25, 49, 81} for TensorSSA to facilitate comparison. The OAs obtained by the SVM classifier are shown in Fig. 9. We can find that TensorSSA is superior to 2DSSA in all sizes of L, which testifies that compared with the local regular window of 2DSSA, the adaptive spatial similarity information used in TensorSSA are more effective. Simultaneously, the robustness and universality of L in TensorSSA are better. 2DSSA has different optimal windows for different datasets, while TensorSSA achieves almost the highest accuracy within a certain range window, namely L=25 and 49.

In addition, we further analyzed the effect of L in relation to the search region size on OAs and running time, with L taken as {9, 25, 49, 81} and w×w taken as {5×5, 7×7, 9×9, 11×11, 13×13}. The results are listed in Table VI. As can be seen from Table VI, an increase in both w×w and L results in a significant increase in the calculation. Optimal classification accuracy is usually concentrated between w×w=7×7 to 11×11 and L = 25 to 49. Due to the differences in spatial resolution and feature morphology of the different datasets, the corresponding optimal parameter combinations also differ. Therefore, considering the processing efficiency, the optimal parameter combinations w×w=11×11 and L = 49 for IP, w×w=5×5 and L = 9 for PU, w×w=7×7 and L = 25 for MG are set respectively.

2) r_{tubal}: It determines the amount of information used to reconstruct the trajectory tensor. To select the optimal r_{tubal}, another set of experiments was designed on three datasets, in which r_{tubal} and r is set to vary within {1, 2, 5, 10, 25} for TensorSSA and 2DSSA, respectively. Fig. 10 shows the OA with different parameters.

As can be seen in Fig. 10, TensorSSA achieves higher accuracy than 2DSSA under different low-rank components. In the case where all components are used for reconstruction, i.e., r_{tubal} = 25, both have the same accuracy because it is equivalent to classifying the original image. Additionally, TensorSSA has the highest accuracy in the case of rank-1. As the rank of the reconstruction component increases, the accuracy decreases, which also testifies that its information is more concentrated in low rank. In contrast, 2DSSA has different optimal ranks for different data, indicating that its main information is relatively scattered, which may be caused by the difference of various features in a fixed window. Therefore, the parameters r_{tubal} is fixed to 1 on all three datasets in the experiments.

B. Feature map comparison

In this subsection, we display the extracted feature maps by TensorSSA and its comparison with 2DSSA. In Fig. 9, two band images of IP including the normal image (band-20) and noise image (band-2) are selected for the performance comparison between TensorSSA and 2DSSA. The original image and ground truth image are also given as benchmarks. Spatial details are represented by different color boxes.

As seen in Fig. 11, 2DSSA ignores most of the spatial structure of the ground objects, leading to a blurred image and degraded boundaries. It is mainly because the embedding of 2DSSA has the same number of selections for each pixel, thus the average processing for each pixel in the reprojection step is the same, which usually ignores the difference in features in non-homogeneous areas. In contrast, TensorSSA can enhance

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3272669

the spatial structure of the image and extract effective spatial features even in the presence of severe noise, mainly thanks to the use of spatial similarity information and spectral information. Besides, based on the contrast of image details (different colored boxes), compared to the original image and 2DSSA image, TensorSSA can preserve the edge and structural features of ground objects and enhance the interclass differences of different ground objects, which are consistent with the ground truth image. This has demonstrated again that the 3D spectral-spatial features extracted by TensorSSA are very effective.

C. Ablation study

The proposed TensorSSA contains two critical parts for boosting the classification performance compared to 2DSSA, i.e., the adaptive embedding based on spatial similarity, and t-SVD based decomposition. Here, we used all three datasets to verify the validity of these two operations. In the embedding step, the no embedding, i.e., raw HSI, and 2D embedding mode of 2DSSA was used for comparison, while in the decomposition step, band-by-band SVD of 2DSSA and Tucker decomposition with rank (r, r, r) = (15, 15, 15) corresponding to the highest accuracy were used for comparison. 2%, 1%, and 1% training samples were selected from the IP, PU, and MG datasets respectively, and the SVM classifier is used for classification. The corresponding classification results are shown in Table VII. Note that each row represents the performance of a different combination of embedding and decomposition methods.

As shown in Table I, the combination of adaptive embedding and t-SVD decomposition, i.e., TensorSSA, achieves optimal performance. Compared with raw data and 2D embedding, the adaptive embedding approach achieved higher classification accuracy on all three datasets, thanks to the spatial similarity boosting the low rank of the trajectory tensor. As for the decomposition step, from the accuracy in Table VII and the comparison of singular values in Fig. 4, we see that t-SVD has a better performance than SVD in terms of feature compression. While SVD acts mainly on the extraction of spatial domain information, t-SVD is based on a compact operation along three dimensions and therefore yields a very efficient feature representation. Tucker decomposition performs a matrix expansion of the 3D tensor in different directions via the Tucker rank. Although its decomposition in the trajectory tensor is better than the raw data, its overall classification accuracy is still limited (rows 2, 5, and 8 of Table VII). On the one hand, it has poor performance for data compression and reconstruction, i.e. most of the information is concentrated on the first ten or twenty components; on the other hand, its decomposition destroys the original spectral and spatial intrinsic features, and the selection of more components in the reconstruction can make larger size data face the problem of memory overflow (e.g. MG dataset). Tucker is more suitable for the processing of raw HSI than trajectory tensor. In contrast, t-SVD, with its fast solving and approximation capabilities (described in section II.A), can approximate the entire tensor under rank-1. This advantage is further amplified by the adaptive embedding in Tensor SSA, which collaboratively accomplishes the extraction of low-rank spectral-spatial features for the entire HSI. Overall, t-SVD is superior to Tucker decomposition in this task, and the resulting Tensor SSA also achieves far better performance than 2DSSA.

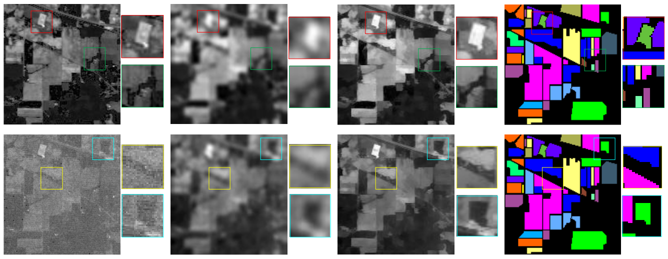
D. Computational complexity Analysis

The computational complexity of the TensorSSA, including computational cost and memory requirements, is further analyzed in this subsection. In terms of computational cost, the main computation of TensorSSA is the truncated t-SVD [68], which requires the calculation of a trajectory tensor of $L \times WH \times B$, and the complexity is $\mathcal{O}(LWHB\log B + L^2 WHB)$, compared with the complexity $\mathcal{O}[L^2(W\text{-}L+1)(H\text{-}L+1)B+L^3B]$ of 2DSSA. As for memory requirements, the complexity of TensorSSA is $\mathcal{O}(LWHB)$ while $\mathcal{O}[L(W\text{-}L+1)(H\text{-}L+1)]$ for 2DSSA. Nevertheless, with a smaller embedding window L (Table VI), TensorSSA spends less running time yet achieves higher classification accuracies than 2DSSA on PU and MG datasets according to Tables I-III.

E. Analysis between TensorSSA and classification

The combination of TensorSSA plus classifier is used in this paper to jointly accomplish the task of ground object classification. The 3D feature extraction of TensorSSA can enhance the performance of classification by improving intra-class consistency and removing noise, etc. However, the feature extraction does not benefit from the prior knowledge of the samples in the classification task. TensorSSA is an unsupervised feature extraction method that performs feature enhancement on the entire HSI rather than the training samples with labels. This enhancement is effective for the differentiation of ground objects with large class differences, but also runs the risk of removing the differences between two similar ground objects, leading to the phenomenon of misclassification. Supervised feature extraction methods have been shown to further improve classification performance [69, 70], and the

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3272669



(a) Original image **(b)** 2DSSA image **(c)** TensorSSA image **(d)** Ground truth image

Fig. 11 Comparison of feature images. (a) Original image; (b) 2DSSA image; (c) TensorSSA image; and (d) Ground truth image.

TABLE VII
OAs (%) OF DIFFERENT COMBINATIONS OF EMBEDDING AND DECOMPOSITION ON THREE DATASETS

Embedding stage			Decomposition stage			Accuracy (OA)		
Raw HSI	2D embedding	Adaptive embedding	SVD	Tucker	t-SVD	IP (2%)	PU (1%)	MG (1%)
✓			✓			74.24	64.13	52.38
✓				✓		64.55	77.08	71.55
✓					✓	81.35	62.96	53.19
	✓		✓			80.15	88.19	84.26
	✓			✓		70.17	89.16	—
	✓				✓	81.33	90.15	84.30
		✓	✓			85.34	95.15	86.99
		✓		✓		74.65	91.29	—
		✓			✓	**88.86**	**95.89**	**87.83**

inclusion of sample prior information in TensorSSA is also worth exploring.

F. *Analysis for search region size*

Generally, it is difficult to determine the best search region size for different datasets, because it usually varies depending on the spatial resolution of the images, the size and morphology of the ground objects. According to the OAs in Table VI, the IP dataset with a low spatial resolution (20 m) has an optimal window of 13×13, while the higher resolution PU (1.3 m) and MG (1 m) data have an optimal window between 7×7 and 9×9. The search region needs to be embedded with abundant information about the neighbor pixels and thus increases with decreasing spatial resolution. Furthermore, larger sizes and regularly shaped features also require larger search regions (e.g. IP) for multiple similarity information extraction, whereas irregularly striped and smaller features (e.g. PU) can accomplish their goals in smaller search regions. Despite the above analysis, it is currently not possible to determine the optimal search region through a paradigm for a given dataset, which needs to be further addressed.

VI. Conclusion

Due to the 3D inherent of HSI, it is desirable to find an effective method for the simultaneous extraction of spectral and spatial features. In conventional spectral-spatial methods, the spectral and spatial features are usually extracted separately. Although the existing 3D methods further consider the spectral and spatial correlation, they pay insufficient attention to global features, and their performance is also limited by the training samples. To solve these problems, this paper proposed a new TensorSSA method for 3D feature extraction of HSI, which is suitable for classification under limited samples.

Adaptive embedding operation considers the spatial self-similarity in HSI and constructs a trajectory tensor containing global spectral-spatial features. T-SVD and truncated t-SVD can jointly extract the low-rank intrinsic characteristics of trajectory tensor, and eliminate the influence information such as noise. Reprojection operation further improves intra-class similarity and maintains inter-class difference while transforming the reconstruction tensor into a feature image. The final features have good low rank, robustness, and representativeness, leading to higher classification accuracy with limited training samples.

Experimental carried out on IP, PU, and MG datasets have demonstrated that 1) the extracted features have an image enhancement and denoising effect. 2) the classification accuracies of TensorSSA are superior to current 3D methods and most deep learning methods under both random and spatially disjoint training samples. 3) TensorSSA can preserve the shape structure of fine ground objects and their irregular boundaries in the classification maps.

For future work, we will consider the rapid implementation of TensorSSA and its extension to large-scale hyperspectral data applications such as UAV hyperspectral wetland classification and fine identification of forest species.

References

[1] N. Fox, A. Parbhakar-Fox, J. Moltzen, S. Feig, K. Goemann, and J. Huntington, “Applications of hyperspectral mineralogy for geoenvironmental characterisation,” *Minerals Engineering,* vol. 107, pp. 63-77, 2017.

[2] 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 Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-15, 2021.

[3] W. Lu, J. Qi, and H. Feng, “Urban functional zone classification based on self-supervised learning: A case study in Beijing, China,” vol. 10, 2022-November-17, 2022.

[4] A. Zare, C. Jiao, and T. Glenn, “Discriminative Multiple Instance Hyperspectral Target Characterization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 40, no. 10, pp. 2342-2354, 2018.

[5] A. Zare, and K. C. Ho, “Endmember Variability in Hyperspectral Analysis: Addressing Spectral Variability During Spectral Unmixing,” *IEEE Signal Processing Magazine,* vol. 31, no. 1, pp. 95-104, 2014.

[6] S. Mei, Q. Bi, J. Ji, J. Hou, and Q. Du, “Spectral Variation Alleviation by Low-Rank Matrix Approximation for Hyperspectral Image Analysis,” *IEEE Geoscience and Remote Sensing Letters,* vol. 13, no. 6, pp. 796-800, 2016.

[7] 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 Journal of Photogrammetry and Remote Sensing,* vol. 93, pp. 112-122, 2014.

[8] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, “Classification of Hyperspectral Images With Regularized Linear Discriminant Analysis,” *IEEE Transactions on Geoscience and Remote Sensing,* vol. 47, no. 3, pp. 862-873, 2009.

[9] D. Hong, N. Yokoya, and X. X. Zhu, “Learning a Robust Local Manifold Representation for Hyperspectral Dimensionality Reduction,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 10, no. 6, pp. 2960-2975, 2017.

[10] Y. Fang, H. Li, Y. Ma, K. Liang, Y. Hu, S. Zhang, and H. Wang, “Dimensionality Reduction of Hyperspectral Images Based on Robust Spatial Information Using Locally Linear Embedding,” *IEEE Geoscience and Remote Sensing Letters,* vol. 11, no. 10, pp. 1712-1716, 2014.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3272669

[11] J. Xia, M. D. Mura, J. Chanussot, P. Du, and X. He, "Random Subspace Ensembles for Hyperspectral Image Classification With Extended Morphological Attribute Profiles," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 53, no. 9, pp. 4768-4786, 2015.

[12] S. Li, Q. Hao, X. Kang, and J. A. Benediktsson, "Gaussian Pyramid Based Multiscale Feature Fusion for Hyperspectral Image Classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 11, no. 9, pp. 3312-3324, 2018.

[13] 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 Transactions on Geoscience and Remote Sensing*, pp. 1-14, 2020.

[14] G. Sun, X. Zhang, X. Jia, J. Ren, A. Zhang, Y. Yao, and H. Zhao, "Deep Fusion of Localized Spectral Features and Multi-scale Spatial Features for Effective Classification of Hyperspectral Images," *International Journal of Applied Earth Observation and Geoinformation,* vol. 91, pp. 102157, 2020.

[15] C. Mu, Q. Zeng, Y. Liu, and Y. Qu, "A Two-Branch Network Combined With Robust Principal Component Analysis for Hyperspectral Image Classification," *IEEE Geoscience and Remote Sensing Letters,* pp. 1-5, 2020.

[16] J. Yang, Y. Zhao, and J. C. Chan, "Learning and Transferring Deep Joint Spectral–Spatial Features for Hyperspectral Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, no. 8, pp. 4729-4742, 2017.

[17] L. O. Jimenez, and D. A. Landgrebe, "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 28, no. 1, pp. 39-54, 1998.

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

[19] B. Hou, T. Huang, and L. Jiao, "Spectral–Spatial Classification of Hyperspectral Data Using 3-D Morphological Profile," *IEEE Geoscience and Remote Sensing Letters,* vol. 12, no. 12, pp. 2364-2368, 2015.

[20] S. Jia, J. Hu, J. Zhu, X. Jia, and Q. Li, "Three-Dimensional Local Binary Patterns for Hyperspectral Imagery Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, no. 4, pp. 2399-2413, 2017.

[21] L. He, J. Li, A. Plaza, and Y. Li, "Discriminative Low-Rank Gabor Filtering for Spectral–Spatial Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, no. 3, pp. 1381-1395, 2017.

[22] T. C. Bau, S. Sarkar, and G. Healey, "Hyperspectral Region Classification Using a Three-Dimensional Gabor Filterbank," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 48, no. 9, pp. 3457-3464, 2010.

[23] S. Jia, L. Shen, J. Zhu, and Q. Li, "A 3-D Gabor Phase-Based Coding and Matching Framework for Hyperspectral Imagery Classification," *IEEE Transactions on Cybernetics,* vol. 48, no. 4, pp. 1176-1188, 2018.

[24] F. Tsai, and J. Lai, "Feature Extraction of Hyperspectral Image Cubes Using Three-Dimensional Gray-Level Cooccurrence," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 51, no. 6, pp. 3504-3513, 2013.

[25] J. Zhu, J. Hu, S. Jia, X. Jia, and Q. Li, "Multiple 3-D Feature Fusion Framework for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 56, no. 4, pp. 1873-1886, 2018.

[26] Q. Li, and D. Schonfeld, "Multilinear Discriminant Analysis for Higher-Order Tensor Data Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 36, no. 12, pp. 2524-2537, 2014.

[27] N. Qi, Y. Shi, X. Sun, J. Wang, B. Yin, and J. Gao, "Multi-Dimensional Sparse Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 40, no. 1, pp. 163-178, 2018.

[28] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-Local Meets Global: An Integrated Paradigm for Hyperspectral Denoising." pp. 6861-6870.

[29] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral Images Super-Resolution via Learning High-Order Coupled Tensor Ring Representation," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 31, no. 11, pp. 4747-4760, 2020.

[30] G. Zhao, B. Tu, H. Fei, N. Li, and X. Yang, "Spatial-spectral classification of hyperspectral image via group tensor decomposition," *Neurocomputing,* vol. 316, pp. 68-77, 2018.

[31] R. Dian, S. Li, L. Fang, T. Lu, and J. M. Bioucas-Dias, "Nonlocal Sparse Tensor Factorization for Semiblind Hyperspectral and Multispectral Image Fusion," *IEEE Transactions on Cybernetics,* vol. 50, no. 10, pp. 4469-4480, 2020.

[32] Y. Xu, Z. Wu, J. Chanussot, P. Comon, and Z. Wei, "Nonlocal Coupled Tensor CP Decomposition for Hyperspectral and Multispectral Image Fusion," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 58, no. 1, pp. 348-362, 2020.

[33] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral Computational Imaging via Collaborative Tucker3 Tensor Decomposition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 31, no. 1, pp. 98-111, 2021.

[34] J. Liu, Z. Wu, L. Xiao, J. Sun, and H. Yan, "Generalized Tensor Regression for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 58, no. 2, pp. 1244-1258, 2020.

[35] R. Li, Z. Pan, Y. Wang, and P. Wang, "The correlation-based tucker decomposition for hyperspectral image compression," *Neurocomputing,* vol. 419, pp. 357-370, 2021.

[36] J. Lin, T.-Z. Huang, X.-L. Zhao, T.-H. Ma, T.-X. Jiang, and Y.-B. Zheng, "A novel non-convex low-rank tensor approximation model for hyperspectral image restoration," *Applied Mathematics and Computation,* vol. 408, pp. 126342, 2021.

[37] H. S. Khaleel, S. V. Mohd Sagheer, M. Baburaj, and S. N. George, "Denoising of Rician corrupted 3D magnetic resonance images using tensor-SVD," *Biomedical Signal Processing and Control,* vol. 44, pp. 82-95, 2018.

[38] W. Lu, C. Tao, H. Li, J. Qi, and Y. Li, "A unified deep learning framework for urban functional zone extraction based on multi-source heterogeneous data," *Remote Sensing of Environment,* vol. 270, pp. 112830, 2022.

[39] L. Zhang, and L. Zhang, "Artificial Intelligence for Remote Sensing Data Analysis: A review of challenges and opportunities," *IEEE Geoscience and Remote Sensing Magazine,* vol. 10, no. 2, pp. 270-294, 2022.

[40] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 158, pp. 279-317, 2019.

[41] 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, 2021.

[42] Y. Li, H. Zhang, and Q. Shen, "Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network," *Remote Sensing,* vol. 9, no. 1, 2017.

[43] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 56, no. 2, pp. 847-858, 2018.

[44] X. Tang, F. Meng, X. Zhang, Y. M. Cheung, J. Ma, F. Liu, and L. Jiao, "Hyperspectral Image Classification Based on 3-D Octave Convolution With Spatial–Spectral Attention Network," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 59, no. 3, pp. 2430-2447, 2021.

[45] 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,* pp. 1-13, 2020.

[46] J. Y. Yang, H. C. Li, Z. C. Li, and T. Y. Ma, "Spatial-Spectral Tensor Graph Convolutional Network for Hyperspectral Image Classification." pp. 2222-2225.

[47] C. Pu, H. Huang, and Y. Li, "Aggregated-Attention Transformation Network for Hyperspectral Image Classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 15, pp. 5674-5688, 2022.

[48] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "SpectralFormer: Rethinking Hyperspectral Image Classification With Transformers," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 60, pp. 1-15, 2022.

[49] Z. Zhong, Y. Li, L. Ma, J. Li, and W. S. Zheng, "Spectral–Spatial Transformer Network for Hyperspectral Image Classification: A Factorized Architecture Search Framework," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 60, pp. 1-15, 2022.

[50] J. Zabalza, J. Ren, Z. Wang, S. Marshall, and J. Wang, "Singular Spectrum Analysis for Effective Feature Extraction in Hyperspectral Imaging," *IEEE Geoscience and Remote Sensing Letters,* vol. 11, no. 11, pp. 1886-1890, 2014.

[51] 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 Transactions on Geoscience and Remote Sensing,* vol. 53, no. 8, pp. 4418-4433, 2015.

[52] G. Sun, H. Fu, J. Ren, A. Zhang, J. Zabalza, X. Jia, and H. Zhao, "SpaSSA: Superpixelwise Adaptive SSA for Unsupervised Spatial-Spectral Feature

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3272669

Extraction in Hyperspectral Image," *IEEE Transactions on Cybernetics*, pp. 1-12, 2021.

[53] H. Fu, G. Sun, J. Zabalza, A. Zhang, J. Ren, and X. Jia, "A Novel Spectral-Spatial Singular Spectrum Analysis Technique for Near Real-Time In Situ Feature Extraction in Hyperspectral Imaging," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2214-2225, 2020.

[54] Z. Zhang, and S. Aeron, "Exact Tensor Completion Using t-SVD," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1511-1526, 2017.

[55] C. Lu, J. Tang, S. Yan, and Z. Lin, "Nonconvex Nonsmooth Low Rank Minimization via Iteratively Reweighted Nuclear Norm," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 829-839, 2016.

[56] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel Methods for Multilinear Data Completion and De-noising Based on Tensor-SVD." pp. 3842-3849.

[57] H. Fu, A. Zhang, G. Sun, J. Ren, X. Jia, Z. Pan, and H. Ma, "A Novel Band Selection and Spatial Noise Reduction Method for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-13, 2022.

[58] L. Zhuang, X. Fu, M. K. Ng, and J. M. Bioucas-Dias, "Hyperspectral Image Denoising Based on Global and Nonlocal Low-Rank Factorizations," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1-17, 2021.

[59] L. Zhang, L. Song, B. Du, and Y. Zhang, "Nonlocal Low-Rank Tensor Completion for Visual Data," *IEEE Transactions on Cybernetics*, vol. 51, no. 2, pp. 673-685, 2021.

[60] C. Ricotta, "From the euclidean distance to compositional dissimilarity: What is gained and what is lost," *Acta Oecologica*, vol. 111, pp. 103732, 2021.

[61] N. Golyandina, A. Korobeynikov, A. Shlemov, and K. Usevich, "Multivariate and 2D Extensions of Singular Spectrum Analysis with theRssaPackage," *Journal of Statistical Software*, vol. 67, no. 2, 2015.

[62] X. Du, and A. Zare, "Technical Report: Scene Label Ground Truth Map for MUUFL Gulfport Data Set," 2017. University of Florida, Gainesville, FL, Tech. Rep. 20170417. Available: http://ufdc.ufl.edu/IR00009711/00001.

[63] S. Jia, Z. Zhan, M. Zhang, M. Xu, Q. Huang, J. Zhou, and X. Jia, "Multiple Feature-Based Superpixel-Level Decision Fusion for Hyperspectral and LiDAR Data Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 1437-1452, 2021.

[64] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor Robust Principal Component Analysis with a New Tensor Nuclear Norm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 925-938, 2020.

[65] X. Yang, Y. Ye, X. Li, R. Y. K. Lau, X. Zhang, and X. Huang, "Hyperspectral Image Classification With Deep Learning Models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5408-5423, 2018.

[66] S. K. Roy, R. Mondal, M. E. Paoletti, J. M. Haut, and A. Plaza, "Morphological Convolutional Neural Networks for Hyperspectral Image Classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 8689-8702, 2021.

[67] C. Chih-Chung, and L. Chih-Jen, "Libsvm: a library for support vector machines," *ACM Trans. Intell. Syst. Technol., vol. 2(3)*, pp. 1-27, 01/01, 2011.

[68] X. Wang, L. Gu, H.-w. Lee, and G. Zhang, "Quantum tensor singular value decomposition*," *Journal of Physics Communications*, vol. 5, no. 7, pp. 075001, 2021.

[69] H. Huang, G. Shi, H. He, Y. Duan, and F. Luo, "Dimensionality Reduction of Hyperspectral Imagery Based on Spatial–Spectral Manifold Learning," *IEEE Transactions on Cybernetics*, vol. 50, no. 6, pp. 2604-2616, 2020.

[70] 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.

Hang Fu received the B.Sc. degree 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.

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.

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.

Baojie Shao received the B.Sc. degree in geomatics engineering from China University of Petroleum (East China), Qingdao, China, in 2020. She is currently working toward the M.Sc. degree at China University of Petroleum (East China), Qingdao, China.

Her research interests include large-scale water extraction and Google Earth Engine.

This article has been accepted for publication in IEEE Transactions on Geoscience and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2023.3272669

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.