Tensor Multi-Subspace Representation for Remote Sensing Image Mixed Noise Removal

Jing-Hua Yang, Heng-Chao Li, *Senior Member, IEEE*, Meng Ding, Xi-Le Zhao, *Senior Member, IEEE*, and Wen-Yu Hu

***Abstract*—Remote sensing image (RSI) denoising is an important and fundamental task in RSI processing. Existing denoising methods usually assume that RSI lies in a single matrix or tensor subspace. However, due to the wavelength difference or/and temporal variability, the assumption of a single subspace may not be suitable for RSI. To address this, we propose a tensor multi-subspace representation (TenMSR) for RSI mixed noise removal. To be specific, in this work, we introduce TenMSR to finely characterize the intrinsic tensor multi-subspace structure of RSI. Compared with the single matrix/tensor subspace-based methods, the proposed method can not only precisely describe the wavelength difference or/and temporal variability of RSI but also produce a more compact image distribution in tensor multi-subspace. To mine and preserve the multi-subspace structure, we introduce a nonlinear transform-based 3-D tensor nuclear norm to characterize the tensor low rankness of the multi-subspace representation coefficient. An effective algorithm based on the proximal alternating minimization (PAM) framework is developed to solve the proposed model with theoretical convergence analysis. Extensive experiments show the effectiveness and superiority of the proposed method over existing state-of-the-art single matrix/tensor subspace RSI denoising methods.**

***Index Terms*—Nonlinear transform-based 3-D tensor nuclear norm, proximal alternating minimization (PAM), remote sensing image (RSI) denoising, tensor multi-subspace representation (TenMSR).**

I. Introduction

With the development of remote sensing technology, the remote sensing image (RSI) has received widespread attention and has been widely used in land surface change monitoring, disaster monitoring and assessment, urban planning and management, and agriculture and forestry monitoring [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14]. However, due to the acquisition equipment and transmission environment, the acquired RSI is generally corrupted by various kinds of noises, such as Gaussian noise, impulse noise, and strip noise. The existence of mixed noises seriously reduces the quality of the acquired RSI. Therefore, the RSI denoising is an important preprocessing step and will benefit the subsequent applications, e.g., image classification, image segmentation, and super-resolution. RSI denoising is an ill-posed inverse problem; the heuristic strategy is to study the prior information of RSI.

One of the most widely studied priors is the global low rankness due to the intrinsic correlation and redundancy of RSI along the spectral dimension. Many early denoising techniques view the RSI as a low-rank matrix, e.g., nuclear norm-based denoising methods [15], [16], [17], [18], [19] and low-rank matrix factorization methods [20], [21], [22], [23]. However, these methods [15], [16], [17], [18], [19], [20], [21], [22], [23], [24] are designed for the matrix data. When handling the multidimensional RSI, the matrix denoising methods have to reshape the RSI into a big matrix, which usually destroys the multidimensional structure and leads to performance degradation. To keep the multidimensional structure, one natural approach is to directly manipulate the original RSI.

Recently, many tensor denoising methods based on different tensor decompositions and corresponding tensor ranks are proposed to explore the low-rank tensor prior information of RSI [3], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34]. For instance, based on Tucker decomposition, Renard et al. [31] proposed a Tucker rank minimization method for RSI denoising. Zhang et al. [35] integrated spatial and spectral difference constraints in a low-rank tensor decomposition framework for complex noise removal. A new low Tucker rank nonnegative approximation method was proposed in [36] and demonstrated promising results. Liu et al. [25] used the PARAFAC decomposition to recover the RSI and gave the statistical performance analysis. Due to the promising performance of tensor singular value decomposition, the works [37], [38], [39], [40] utilized the tensor tubal rank to explore the low rankness of the underlying clean spectral image in the transformed Fourier domain for RSI recovery. The aforementioned low-rank tensor methods can achieve the promising denoising performance improvement, while they treat the whole observed noisy data as the input of the algorithm—that brings the high computational burden, especially when the spatial size and spectral number of RSI increase.

To reduce the computational burden, many subspace representation-based methods have been proposed to remove the mixed noise from RSI [41], [42], [43], [44], [45]. The

Received 17 April 2024; revised 10 June 2025 and 28 August 2025; accepted 4 October 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 12401605, Grant 12371456, Grant 62271418, Grant 12201522, and Grant 62266002; in part by the Natural Science Foundation of Sichuan Province under Grant 2024NSFSC1467; in part by Sichuan Science and Technology Program under Grant 2024NSFJQ0038 and Grant 2024NSFSC0038; in part by Jiangxi Provincial Natural Science Foundation under Grant 20252BAC250007; in part by the Postdoctoral Fellowship Program of CPSF under Grant GZC20232198; and in part by China Postdoctoral Science Foundation under Grant 2024M752661. *(Corresponding authors: Heng-Chao Li; Meng Ding.)*

Jing-Hua Yang and Heng-Chao Li are with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China (e-mail: yangjinghua110@126.com; hcli@home.swjtu.edu.cn).

Meng Ding is with the School of Mathematics, Southwest Jiaotong University, Chengdu 611756, China (e-mail: dingmeng56@163.com).

Xi-Le Zhao is with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: xlzhao122003@163.com).

Wen-Yu Hu is with the School of Mathematics and Computer Sciences, Gannan Normal University, Ganzhou 341000, China (e-mail: huwenyu@gnnu.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNNLS.2025.3620863, provided by the authors.

Digital Object Identifier 10.1109/TNNLS.2025.3620863

basic idea is that the RSI across the spectral dimension is strongly correlated, which implies that the clean RSI lies in a low-rank subspace. By decomposing the RSI into two smaller matrices, Sun and Jeon [46] proposed a subspace spatial-spectral low-rank method to improve the denoising performance. Zhuang and Bioucas-Dias [41] proposed a fast hyperspectral denoising (FastHyDe) algorithm by learning the matrix subspace from the Casorati matrix with lower time cost. He et al. [47] developed a unified paradigm to combine the spatial and spectral correlation by learning a global spectral low-rank orthogonal subspace. By exploring the nonlocal self-similarity of representation coefficient matrix, a spectral low-rank subspace method is studied in [42] for RSI denoising. Since the 2-D matricization destroys the spatial information, matrix subspace representation methods can still be further improved. To enhance the ability of subspace representation, Lin et al. [43] learned a basis tensor from the input tensor and proposed a tensor subspace representation method for delivering the multidimensional structure of the RSI. Furthermore, He et al. [48] combined the low-dimensional tensor subspace representation and a nonlocal plug-and-play prior for RSI denoising. By exploiting the global spectral correlation of image, Chen et al. [49] ingeniously incorporated low-rank subspace representation and data-driven coefficient prior for image reconstruction.

In summary, the matrix/tensor subspace-based methods perform well in balancing denoising performance and efficiency; however, they assume that RSI is distributed in a single subspace. This assumption may be too strict for the RSI processing. In fact, the spectra of RSI have obvious differences, especially for the images with hundreds of spectral bands. Besides, due to the collecting variability, the spectral bands of the multitemporal RSI may not be highly correlated. This indicates that it is more reasonable to assume that the RSI lies in multiple subspaces than in a single subspace. Therefore, the multiple subspace representation can accurately describe the internal structure of RSI data and further improve the denoising performance.

Recently, deep learning-based RSI denoising methods primarily use neural networks (e.g., convolutional neural network (CNN), U-Net, and Transformer) to learn the mapping between noisy and clean images and apply a large amount of data to train the model to separate noise while retaining image details [4], [8], [50], [51], [52], [53], [54], [55], [56], [57]. The deep learning-based denoising methods have achieved promising performance, but they encounter several challenges: 1) they require a large amount of noise-free paired training data to learn the denoising network, which is hard to collect in practice; 2) the interpretability and generalization ability of denoising networks may be limited for different data and noise distributions; and 3) the deep learning-based denoising methods often lack the rigorous theoretical analysis of the corresponding algorithm.

A. Motivation

To finely describe the essential structure of the data, we propose a tensor multi-subspace representation (TenMSR) to improve the capability of subspace representation. We denote the RSI as $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where n_1 and n_2 are spatial dimensions and n_3 is the spectral dimension. The matrix subspace methods in [41], [46], and [58] learned the subspace representation from the unfolding matrix of $\mathcal{X}$ along the third dimension. The tensor subspace method in [43] is learned from the permuted tensor $\vec{\mathcal{X}} = \text{Permute}(\mathcal{X}, [1, 3, 2]) \in \mathbb{R}^{n_1 \times n_3 \times n_2}$. However, the matrix/tensor subspace representation methods assume that RSI lies in a single subspace—that ignores spectrum differences in RSI.

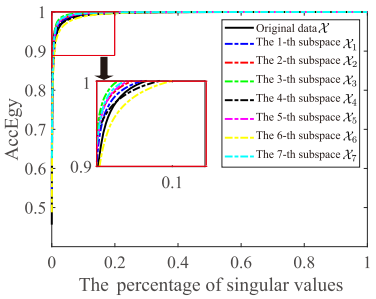

Fig. 1. Comparison of AccEgy of singular values of the original RSI and images belonging to each subspace.

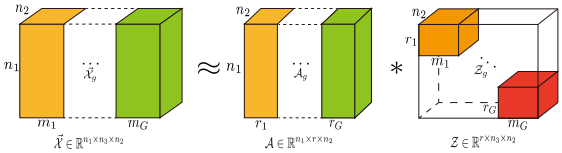

Fig. 2. Illustration of the TenMSR.

Due to the wavelength difference, the image $\vec{\mathcal{X}}$ can be divided into G groups $\vec{\mathcal{X}}_g \in \mathbb{R}^{n_1 \times m_g \times n_2}$ with $\sum_{g=1}^{G} m_g = n_3$. The gth group RSI $\vec{\mathcal{X}}_g$ is highly correlated, but not correlated with RSI in other groups. Fig. 1 plots the accumulation energy ratio (AccEgy) of singular values of the whole image $\vec{\mathcal{X}}$ and each group image $\vec{\mathcal{X}}_g, g = 1, \ldots, G$. The used image is the RSI obtained by Landset8 with seven time nodes, see details in Section V-A. The grouping subspaces are obtained by applying the spectral clustering tool NCut [59] to the spectral dimension of the remote sensing data. One can see that for most of the subspaces, each subspace image needs less singular values than the original image when it achieves the same AccEgy. In other words, the images belonging to one subspace admit a low-rank property more significantly than that of the whole RSI. Therefore, each group image $\vec{\mathcal{X}}_g$ lies in a certain single low-dimensional subspace, and it is more suitable to assume that the RSI is distributed in multiple tensor subspaces along the spectral dimension.

Based on the above analysis, the spectral bands belonging to $\vec{\mathcal{X}}_g$ are related to each other. Using t-product (see Definition 2), $\vec{\mathcal{X}}_g$ can be represented by the basis tensor $\mathcal{A}_g$, and the corresponding representation tensor $\mathcal{Z}_g$ describes the similarity between different bands. Then, the whole image $\vec{\mathcal{X}}$ can be represented by the multibasis tensor and multicoefficient tensor (see Fig. 2). To proceed, in Fig. 3, we present the tensor multi-subspace structure of the hyperspectral image (HSI) PaC and Landset8 datasets in the spectral dimension (the data details are provided in Section V). According to Fig. 3(b) and (d), we observe that the representation tensors of two datasets have an obvious tensor block-diagonal structure, which implies that the above RSIs are distributed in multiple independent subspaces.

To fully mine and preserve the tensor multi-subspace structure, we explore the prior information of the representation tensor. In Fig. 4, we show the singular value curves of representation tensor along different modes. One can see that the curves rapidly decrease and tend to zero. Then, we develop a nonlinear-transform-based 3-D tensor nuclear

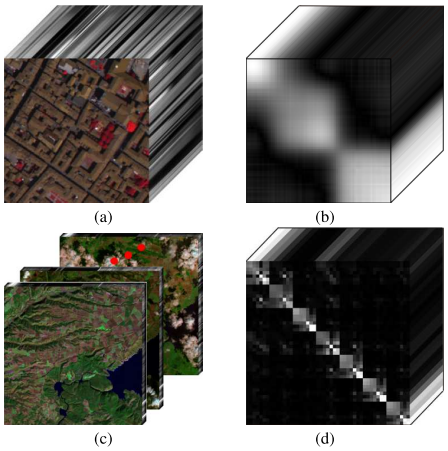

Fig. 3. Illustration of tensor multi-subspace structures on RSIs. (a) PaC dataset. (b) multi-subspace structure of representation tensor on PaC. (c) Landsat8 dataset. (d) multi-subspace structure of representation tensor on Landsat8.

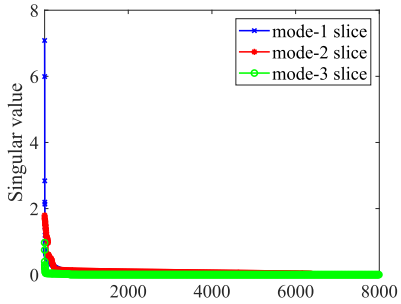

Fig. 4. Singular value curves of representation tensor along different modes.

norm (see Definition 7) to characterize the low rankness of the representation tensor. Therefore, we propose a TenMSR method to effectively remove RSI mixed noise.

B. Contributions

The main contributions of this work are as follows.

1) We design a new TenMSR method to finely characterize the essential structure of RSI. Based on the t-product, TenMSR represents RSI as the multibasis tensor and multicoefficient tensor. The proposed TenMSR can adaptively capture the multiple subspace structures of RSIs by exploring the wavelength difference or temporal variability.
2) Based on the advantage of TenMSR in accurately depicting the internal structure of RSIs, we propose a TenMSR-based model for RSI mixed noise removal. To fully mine the tensor multi-subspace structure, we develop a nonlinear transform-based 3-D tensor nuclear norm (N-3DTNN) to fully and comprehensively characterize the low rankness of multi-subspace representation coefficient tensor.
3) We design an effective algorithm under the proximal alternating minimization (PAM) algorithm framework to solve the proposed model. Theoretically, we prove that the proposed algorithm converges to a critical point. The proposed denoising method is evaluated on extensive remote sensing data, including semi-real and real RSI. Experiments on various datasets and noise settings demonstrate that the proposed method outperforms baselines qualitatively and quantitatively.

The rest of this work is organized as follows. Section II presents some notations and preliminaries used in this work and gives the definition of N-3DTNN. Section III presents the proposed TenMSR denoising model. Section IV shows the proposed PAM algorithm for solving the proposed model with theoretical guarantees. Section V tests the performance of the proposed method on different datasets. Section VI presents some discussions. Section VII summarizes the work.

II. Notations and Definitions

A. Notations

We use symbols $\mathcal{Z}$, $\boldsymbol{Z}$, $\boldsymbol{z}$, and z to denote the tensor, the matrix, the vector, and the scalar, respectively. For a third-order tensor $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we use the MATLAB notations $\mathcal{Z}(i,:,:)$, $\mathcal{Z}(:,j,:)$, and $\mathcal{Z}(:,:,k)$ to denote the ith horizontal, jth lateral, and kth frontal slice, respectively. For convenience, we denote the slice $\mathcal{Z}(:,:,k)$ by $\mathcal{Z}^{(k)}$, the (i,j,k)th entry by $\mathcal{Z}_{i,j,k}$, and the mode-1, mode-2, and mode-3 fibers by $\mathcal{Z}(:,j,k)$, $\mathcal{Z}(i,:,k)$, and $\mathcal{Z}(i,j,:)$, respectively. The tensor Frobenius norm is $\|\mathcal{Z}\|_F = \left(\sum_{i,j,k} |\mathcal{Z}_{i,j,k}|^2\right)^{1/2}$, the tensor infinity norm is $\|\mathcal{Z}\|_\infty = \max_{i,j,k} |\mathcal{Z}_{i,j,k}|$, and the tensor l_1 norm is $\|\mathcal{Z}\|_1 = \sum_{i,j,k} |\mathcal{Z}_{i,j,k}|$. The matrix nuclear norm is $\|\boldsymbol{Z}\|_* = \sum \sigma(\boldsymbol{Z})$, where $\sigma(\boldsymbol{Z})$ is the singular value of $\boldsymbol{Z}$. For $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the permuted tensor can be obtained by $\vec{\mathcal{Z}} := \text{Permute}(\mathcal{Z}, [1,3,2]) \in \mathbb{R}^{n_1 \times n_3 \times n_2}$, where the frontal slice of $\mathcal{Z}$ is transformed into the lateral slice of $\vec{\mathcal{Z}}$. Then, the original data can be obtained by the inverse operation $\mathcal{Z} := \text{inv-Permute}(\vec{\mathcal{Z}}, [1,3,2])$.

B. Generalized Definitions

To better understand the definition of N-3DTNN, we introduce some related definitions.

Definition 1 (Mode-s Matrix-Tensor Product [32]): The mode-s matrix-tensor product of $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}$ and $\boldsymbol{A} \in \mathbb{R}^{m \times n_s}$ yields a tensor $\mathcal{X}$ of size $n_1 \times \cdots \times n_{s-1} \times m \times n_{s+1} \times \cdots \times n_p$, i.e.,

$$\mathcal{X} = \mathcal{Z} \times_s \boldsymbol{A} = \texttt{fold}_s\left(\boldsymbol{A}\boldsymbol{Z}_{(s)}\right)$$

where $\boldsymbol{Z}_{(s)}$ is the mode-s matricization of $\mathcal{Z}$ by $\boldsymbol{Z}_{(s)} = \texttt{unfold}_s(\mathcal{Z})$ and $\texttt{unfold}_s$ is the inverse operator of $\texttt{fold}_s$ operator, i.e., $\mathcal{Z} = \texttt{fold}_s(\texttt{unfold}_s(\mathcal{Z}))$.

Define the discrete Fourier transform (DFT) matrix $\boldsymbol{F}_n = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k, \ldots, \boldsymbol{f}_n] \in \mathbb{C}^{n \times n}$, where $\boldsymbol{f}_k = [\omega^{0\times(k-1)}; \omega^{1\times(k-1)}; \ldots; \omega^{(n-1)\times(k-1)}] \in \mathbb{C}^n$ with $\omega = e^{-(2\pi\sqrt{-1}/n)}$. The DFT of $\mathcal{Z}$ along the sth mode can be obtained by $\mathcal{L}_s = \mathcal{Z} \times_s \boldsymbol{F}_{n_s}$. Similarly, the inverse operator is defined as $\mathcal{Z} = \mathcal{L}_s \times_s \boldsymbol{F}_{n_s}^H$, where $\boldsymbol{F}_{n_s}^H$ is the conjugate transpose of $\boldsymbol{F}_{n_s}$.

Definition 2 (T-Product [60]): The t-product $\mathcal{C} = \mathcal{A} * \mathcal{B}$ of two tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ is a tensor with size $n_1 \times n_4 \times n_3$, whose each tube is computed as

$$\mathcal{C}(i,l,:) = \sum_{j=1}^{n_2} \mathcal{A}(i,j,:) \star \mathcal{B}(j,l,:)$$

where $\star$ is the circular convolution operation of two vectors.

Definition 3 (Conjugate Transpose [39]): For a tensor $\mathcal{Z}$ of size $n_1 \times n_2 \times n_3$, its conjugate transpose $\mathcal{Z}^\top$ $(n_2 \times n_1 \times n_3)$ is obtained by conjugately transposing each frontal slice and then revering the order of transposed frontal slices 2 to n_3.

Definition 4 (Identify Tensor [39]): An identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is given by setting the first frontal slice to be the $n \times n$ identity matrix and other frontal slices to be zeros.

Definition 5 (Mode-s TNN [38]): The mode-s tensor nuclear norm of $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is

$$\|\mathcal{Z}\|_{\text{TNN}_s} = \Sigma_{i=1}^{n_s} \left\| (\mathcal{L}_s)_s^{(i)} \right\|_* \tag{1}$$

where $\mathcal{L}_s = \mathcal{Z} \times_s \boldsymbol{F}_{n_s}$ and $(\mathcal{L}_s)_s^{(i)}$ is the ith slice of $\mathcal{L}_s$ along the sth mode.

Definition 6 (3DTNN [38]): Let $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, and then, three-direction tensor nuclear norm (3DTNN) is defined as

$$\|\mathcal{Z}\|_{\text{3DTNN}} = \sum_{s=1}^{3} \alpha_s \|\mathcal{Z}\|_{\text{TNN}_s} \tag{2}$$

where α_s $(s = 1, 2, 3)$ are the nonnegative weights satisfying $\sum_{s=1}^{3} \alpha_s = 1$.

The above 3DTNN depends on the linear transform, i.e., DFT, to explore the low rankness of the underlying tensor. However, for exploiting the low-rank structure of the data, a linear transform may not always be effective. Inspired by [61], we define the nonlinear transform-based tensor nuclear norm along different modes.

Definition 7 (N-3DTNN): For $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, N-3DTNN as

$$\begin{aligned} \|\mathcal{Z}\|_{\text{N-3DTNN}} &= \sum_{s=1}^{3} \alpha_s \|\phi(\mathcal{Z})\|_{\text{TNN}_s} \\ &= \sum_{s=1}^{3} \sum_{i=1}^{n_k} \alpha_s \left\| \phi\left((\mathcal{L}_s)_s^{(i)} \right) \right\|_* \end{aligned}$$

where ϕ is the element-wise nonlinear transform. Since $\mathcal{L}_s = \mathcal{Z} \times_s \boldsymbol{F}_{n_s}$, the above nonlinear transform is composed of Fourier transform along different modes and the element-wise nonlinear transform on slices of the transformed tensor.

Compared with 3DTNN, N-3DTNN can achieve a better low-rank approximation of the underlying tensor. Under the framework of TenMSR, we introduce N-3DTNN to characterize the low rankness of representation tensor, thus effectively preserving the image details and textures of RSI, see Section VI for more details.

III. Proposed TenMSR Model

A. Problem Formulation

Given the observed RSI $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, due to the corruption by various noises (including Gaussian noise, impulse noise, dead lines, and stripes), the data degradation process can be mathematically formulated as

$$\mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N} \tag{3}$$

where $\mathcal{X}$ is the underlying RSI; $\mathcal{S}$ is the mixture of impulse noise, dead lines, and stripes; and $\mathcal{N}$ denotes the Gaussian noise. The goal of RSI denoising is to restore the underlying clean image $\mathcal{X}$ from the noisy data $\mathcal{Y}$.

B. Proposed Model

As aforementioned, due to the wavelength difference of collected spectral images, the RSI $\vec{\mathcal{X}}$ can be divided into multiple groups $\{\vec{\mathcal{X}}_g\}_{g=1}^{G}$. Each $\vec{\mathcal{X}}_g$ admits a high global correlation; hence, the bands of $\vec{\mathcal{X}}_g$ lie in a low-tensor-rank subspace. By the t-product, $\vec{\mathcal{X}}_g$ can be represented as the tensor linear representation of the tensor basis, i.e.,

$$\vec{\mathcal{X}}_g = \mathcal{A}_g * \mathcal{Z}_g, \quad g = 1, \ldots, G \tag{4}$$

where $\vec{\mathcal{X}}_g \in \mathbb{R}^{n_1 \times m_g \times n_2}$ is the gth group data, $\mathcal{A}_g \in \mathbb{R}^{n_1 \times r_g \times n_2}$ denotes the tensor basis with rank r_g, and $\mathcal{Z}_g \in \mathbb{R}^{r_g \times m_g \times n_2}$ is the tensor representation coefficient (see Fig. 2). Considering all groups, $\vec{\mathcal{X}}$ can be rewritten as $\vec{\mathcal{X}} = [\vec{\mathcal{X}}_1, \vec{\mathcal{X}}_2, \ldots, \vec{\mathcal{X}}_G]$. According to (4), we have

$$\begin{aligned} \vec{\mathcal{X}} &= \left[\vec{\mathcal{X}}_1, \vec{\mathcal{X}}_2, \ldots, \vec{\mathcal{X}}_G \right] \\ &= [\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_G] * \begin{bmatrix} \mathcal{Z}_1 & & \\ & \ddots & \\ & & \mathcal{Z}_G \end{bmatrix} \end{aligned} \tag{5}$$

where $\mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_G$ represent the tensor representation coefficients of different subspaces. They are distributed on the diagonal of the tensor, thus forming a block-diagonal tensor. Based on the above discussion, the RSI is distributed in multiple subspaces rather than a single subspace.

In this work, we propose a new TenMSR for RSI mixed noise removal. Define $\mathcal{A} = [\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_G] \in \mathbb{R}^{n_1 \times r \times n_2}$ with $r = \sum r_g$ and the block-diagonal tensor

$$\mathcal{Z} = \begin{bmatrix} \mathcal{Z}_1 & & \\ & \ddots & \\ & & \mathcal{Z}_G \end{bmatrix} \in \mathbb{R}^{r \times n_3 \times n_2}$$

then (5) can be equivalently reformulated as

$$\vec{\mathcal{X}} = \mathcal{A} * \mathcal{Z}.$$

Here, $\mathcal{Z}$ is the multiple subspace representation coefficient and depicts the correlation of RSI belonging to different subspaces. In the proposed method, we use the pre-denoised data by [38] as the dictionary $\mathcal{A}$, then each $\mathcal{Z}_g$ characterizes the similarity between bands in the gth subspace and the membership of bands to the subspace. Therefore, according to the block-diagonal structure of $\mathcal{Z}$, we can adaptively learn the number of subspaces by the number of blocks, and the dimension of each subspace can be estimated by the dimension of the corresponding basis tensor $\mathcal{A}_g$. Naturally, $\mathcal{Z}$ can cluster all bands $\vec{\mathcal{X}}(:, j, :)$ $(j = 1, 2, \ldots, n_3)$ into G groups, and the images belonging to one certain group $\vec{\mathcal{X}}_g$ are highly correlated between each other and more closely distributed. Therefore, we consider the low rankness of tensor representation $\mathcal{Z}$ to characterize the similarity of data in each subspace. We can use the mode-s TNN to explore the low rankness of $\mathcal{Z}$. However, there exist two limitations. First, one certain mode-s TNN only considers the low rankness along the sth dimension. Second, the linear transform-based TNN may not effectively exploit the low rankness of tensor data.

To accurately mine the tensor multi-subspace structure, we introduce N-3DTNN to characterize the low rankness of the tensor representation coefficient and propose the following tensor multi-subspace low-rank representation for RSI denoising:

$$\min_{\mathcal{Z}, \mathcal{S}, \mathcal{L}_k} \sum_{s=1}^{3} \sum_{i=1}^{n_s} \lambda_{1s} \left\| \phi\left((\mathcal{L}_s)_s^{(i)} \right) \right\|_* + \lambda_2 \left\| \vec{\mathcal{S}} \right\|_1$$

$$+\frac{1}{2}\left\|\vec{\mathcal{Y}}-\mathcal{A}*\mathcal{Z}-\vec{\mathcal{S}}\right\|_F^2$$
$$\text{s.t. } \mathcal{Z}=\mathcal{L}_s\times_s \boldsymbol{F}_{n_s}^H \tag{6}$$

where λ_{1s} $(s=1,2,3)$ and λ_2 are regularization parameters, ϕ is the element-wise nonlinear transform, and $(\mathcal{L}_s)_s^{(i)}$ is the ith slice of $\mathcal{L}_s$ along the sth mode. The first term aims to explore the low rankness of representation tensor of TenMSR. The second term is to constrain the sparse noise. The third term aims to learn the tensor multi-subspace structure of RSIs along the spectral dimension and the Frobenius norm is used to constrain Gaussian noise. Note that the proposed TenMSR can adaptively learn the inherent multi-subspace structure of RSIs, rather than a simple combination of traditional single subspace learning methods.

IV. Optimization Technique

In this section, we propose an effective algorithm to solve our model (6) and establish its convergence analysis.

A. Proposed Algorithm

By using the half quadratic splitting (HQS) [62], [63] technique and introducing the variables $\mathcal{U}=\mathcal{Z}$ and $\mathcal{W}_s=\phi(\mathcal{L}_s)$, we can rewrite the optimization problem as

$$\begin{aligned}\min_{\mathcal{W}_s,\mathcal{Z},\mathcal{S},\mathcal{L}_s,\mathcal{U}}\ &\sum_{s=1}^{3}\left\{\sum_{i=1}^{n_s}\lambda_{1s}\left\|(\mathcal{W}_s)_s^{(i)}\right\|_*+\frac{\beta_{3s}}{2}\left\|\mathcal{W}_s-\phi\left(\mathcal{L}_s\right)\right\|_F^2\right\}\\&+\frac{\beta_1}{2}\left\|\mathcal{U}-\mathcal{Z}\right\|_F^2+\frac{\beta_2}{2}\left\|\vec{\mathcal{Y}}-\mathcal{A}*\mathcal{U}-\vec{\mathcal{S}}\right\|_F^2\\&+\lambda_2\left\|\vec{\mathcal{S}}\right\|_1+\frac{\beta_4}{2}\left\|\mathcal{Z}-\mathcal{L}_s\times_s\boldsymbol{F}_{n_s}^H\right\|_F^2\end{aligned}\tag{7}$$

where β_1, β_2, β_{3s}, and β_4 are penalty parameters. To solve the optimization problem (7), we develop an effective PAM algorithm [64], [65]. Then, defining the objective function in (7) as $f(\mathcal{W}_s,\mathcal{Z},\mathcal{S},\mathcal{L}_s,\mathcal{U})$, the above variables can be iteratively updated as follows, (8), as shown at the bottom of the page, where t denotes the iteration number and ρ is a positive constant. Next, we show the details for solving all subproblems.

1) The $\mathcal{W}_s$-subproblem is

$$\min_{\mathcal{W}_s}\sum_{i=1}^{n_s}\lambda_1\left\|(\mathcal{W}_s)_s^{(i)}\right\|_*+\frac{\beta_{3s}}{2}\left\|\mathcal{W}_s-\phi\left(\mathcal{L}_s^t\right)\right\|_F^2+\frac{\rho}{2}\left\|\mathcal{W}_s-\mathcal{W}_s^t\right\|_F^2.$$

The above subproblem can be decomposed into the following n_s problems:

$$\min_{\mathcal{W}_s}\lambda_1\left\|\mathcal{W}_s^{(i)}\right\|_*+\frac{\beta_{3s}+\rho}{2}\left\|\mathcal{W}_s^{(i)}-\mathcal{H}_s^{(i)}\right\|_F^2$$

where $\mathcal{H}_s^{(i)}=(\beta_{3s}\phi((\mathcal{L}^t)^{(i)})+\rho\mathcal{W}_{(i)}^t)/(\beta_{3s}+\rho)$. According to the singular value thresholding operator [66], we have

$$\mathcal{W}_s^{(i)}=\boldsymbol{U}\Gamma_{\frac{\lambda_1}{\beta_{3s}+\rho}}(\Sigma)\boldsymbol{V}^\top \tag{9}$$

where $\boldsymbol{U}\Sigma\boldsymbol{V}^\top$ is the singular values decomposition of $\mathcal{H}_s^{(i)}$ and $\Gamma_{\lambda_1/(\beta_{3s}+\rho)}(\Sigma)=\mathrm{diag}(\max(\sigma_j-\lambda_1/(\beta_{3s}+\rho),0))$, in which σ_j denotes the jth singular value.

2) The $\mathcal{Z}$-subproblem is

$$\min_{\mathcal{Z}}\frac{\beta_1}{2}\left\|\mathcal{U}^t-\mathcal{Z}\right\|_F^2+\frac{\beta_4}{2}\left\|\mathcal{Z}-\mathcal{L}_s^t\times_s\boldsymbol{F}_{n_s}^H\right\|_F^2+\frac{\rho}{2}\left\|\mathcal{Z}-\mathcal{Z}^t\right\|_F^2.$$

This is a least-squares problem with the following closed-form solution:

$$\mathcal{Z}^{t+1}=\frac{\beta_1\mathcal{U}^t+\beta_4\mathcal{L}_s^t\times_s\boldsymbol{F}_{n_s}^H+\rho\mathcal{Z}^t}{\beta_1+\beta_4+\rho}.\tag{10}$$

3) The $\mathcal{S}$-subproblem is

$$\min_{\mathcal{S}}\lambda_2\left\|\vec{\mathcal{S}}\right\|_1+\frac{\beta_2}{2}\left\|\vec{\mathcal{Y}}-\mathcal{A}*\mathcal{U}^t-\vec{\mathcal{S}}\right\|_F^2+\frac{\rho}{2}\left\|\vec{\mathcal{S}}-\vec{\mathcal{S}}^t\right\|_F^2.$$

Then, we have

$$\min_{\mathcal{S}}\lambda_2\left\|\vec{\mathcal{S}}\right\|_1+\frac{\beta_2+\rho}{2}\left\|\vec{\mathcal{S}}-\frac{\beta_2\left(\vec{\mathcal{Y}}-\mathcal{A}*\mathcal{U}^t\right)+\rho\vec{\mathcal{S}}^t}{\beta_2+\rho}\right\|_F^2.$$

The subproblem can be calculated by the soft shrinkage operator

$$\vec{\mathcal{S}}^{t+1}=\mathrm{shrink}\left(\frac{\beta_2\left(\vec{\mathcal{Y}}-\mathcal{A}*\mathcal{U}^t\right)+\rho\vec{\mathcal{S}}^t}{\beta_2+\rho},\frac{\lambda_2}{\beta_2+\rho}\right)\tag{11}$$

where $[\mathrm{shrink}(\mathcal{D},\xi)]_{i,j,k}=\mathrm{sgn}(\mathcal{D}_{i,j,k})\max(|\mathcal{D}_{i,j,k}|-\xi,0)$, and sgn denotes the sign function, which is defined as follows:

$$\mathrm{sgn}(x)=\begin{cases}1, & \text{if } x>0\\0, & \text{if } x=0\\-1, & \text{if } x<0.\end{cases}$$

4) The $\mathcal{L}_s$-subproblem is

$$\begin{aligned}\min_{\mathcal{L}_s}\sum_{k=1}^{3}&\frac{\beta_{3s}}{2}\left\|\mathcal{W}_s^{t+1}-\phi\left(\mathcal{L}_s\right)\right\|_F^2+\frac{\beta_4}{2}\left\|\mathcal{Z}^{t+1}-\mathcal{L}_s\times_s\boldsymbol{F}_{n_s}^H\right\|_F^2\\&+\frac{\rho}{2}\left\|\mathcal{L}_s-\mathcal{L}_s^t\right\|_F^2.\end{aligned}$$

The subproblem can be equivalently transformed into

$$\min_{\boldsymbol{L}_{s(s)}}\sum_{s=1}^{3}\frac{\beta_{3s}}{2}\left\|\boldsymbol{W}_{s(s)}^{t+1}-\phi\left(\boldsymbol{L}_{s(s)}\right)\right\|_F^2$$

$$\begin{cases}\mathcal{W}_s^{t+1}=\arg\min\limits_{\mathcal{W}_s}\left\{Q_1\left(\mathcal{W}_s|\mathcal{W}_s^t\right)=f\left(\mathcal{W}_s,\mathcal{Z}^t,\mathcal{S}^t,\mathcal{L}_s^t,\mathcal{U}^t\right)+\frac{\rho}{2}\left\|\mathcal{W}_s-\mathcal{W}_s^t\right\|_F^2\right\}\\\mathcal{Z}^{t+1}=\arg\min\limits_{\mathcal{Z}}\left\{Q_2\left(\mathcal{Z}|\mathcal{Z}^t\right)=f\left(\mathcal{W}_s^{t+1},\mathcal{Z},\mathcal{S}^t,\mathcal{L}_s^t,\mathcal{U}^t\right)+\frac{\rho}{2}\left\|\mathcal{Z}-\mathcal{Z}^t\right\|_F^2\right\}\\\mathcal{S}^{t+1}=\arg\min\limits_{\mathcal{S}}\left\{Q_3\left(\mathcal{S}|\mathcal{S}^t\right)=f\left(\mathcal{W}_s^{t+1},\mathcal{Z}^{t+1},\mathcal{S},\mathcal{L}_s^t,\mathcal{U}^t\right)+\frac{\rho}{2}\left\|\mathcal{S}-\mathcal{S}^t\right\|_F^2\right\}\\\mathcal{L}_s^{t+1}=\arg\min\limits_{\mathcal{L}_s}\left\{Q_4\left(\mathcal{L}_s|\mathcal{L}_s^t\right)=f\left(\mathcal{W}_s^{t+1},\mathcal{Z}^{t+1},\mathcal{S}^{t+1},\mathcal{L}_s,\mathcal{U}^t\right)+\frac{\rho}{2}\left\|\mathcal{L}_s-\mathcal{L}_s^t\right\|_F^2\right\}\\\mathcal{U}^{t+1}=\arg\min\limits_{\mathcal{U}}\left\{Q_5\left(\mathcal{U}|\mathcal{U}^t\right)=f\left(\mathcal{W}_s^{t+1},\mathcal{Z}^{t+1},\mathcal{S}^{t+1},\mathcal{L}_s^{t+1},\mathcal{U}\right)+\frac{\rho}{2}\left\|\mathcal{U}-\mathcal{U}^t\right\|_F^2\right\}\end{cases}\tag{8}$$

$$+\frac{\beta_4}{2}\left\|\boldsymbol{L}_{s(s)}-\boldsymbol{F}_s\boldsymbol{Z}_{(s)}^{t+1}\right\|_F^2+\frac{\rho}{2}\left\|\boldsymbol{L}_{s(s)}-\boldsymbol{L}_{s(s)}^t\right\|_F^2$$
$$=\min_{\boldsymbol{L}_{s(s)}}\sum_{s=1}^{3}\frac{\beta_{3s}}{2}\left\|\phi\left(\boldsymbol{L}_{s(s)}\right)-\boldsymbol{W}_{s(s)}^{t+1}\right\|_F^2+\frac{\beta_4+\rho}{2}\left\|\boldsymbol{L}_{s(s)}-\boldsymbol{M}_s\right\|_F^2 \tag{12}$$

where $\boldsymbol{M}_s = (\beta_4\boldsymbol{F}_{n_s}\boldsymbol{Z}_{(s)}^{t+1}+\rho\boldsymbol{L}_{s(s)}^t)/(\beta_4+\rho)$. We solve the above subproblem by the Newton method.

5) The $\mathcal{U}$-subproblem is

$$\min_{\mathcal{U}}\frac{\beta_1}{2}\left\|\mathcal{U}-\mathcal{Z}^{t+1}\right\|_F^2+\frac{\beta_2}{2}\left\|\vec{\mathcal{Y}}-\mathcal{A}*\mathcal{U}-\vec{\mathcal{S}}^{t+1}\right\|_F^2+\frac{\rho}{2}\left\|\mathcal{U}-\mathcal{U}^t\right\|_F^2.$$

The solution satisfies the following equation:

$$\left(\mathcal{A}^\top*\mathcal{A}+\frac{\beta_1+\rho}{\beta_2}\mathcal{I}\right)*\mathcal{U}=\left(\mathcal{A}^\top*\vec{\mathcal{Y}}-\mathcal{A}^\top*\vec{\mathcal{S}}^{t+1}\right)+\frac{\beta_1}{\beta_2}\mathcal{Z}^{t+1}+\frac{\rho}{\beta_2}\mathcal{U}^t. \tag{13}$$

Algorithm 1 TenMSR for RSI Mixed Noise Removal

Input: the observed tensor $\mathcal{X}$, dictionary $\mathcal{A}$, parameters λ_{1s}, λ_2, β_1, β_2, β_{3s}, β_4, $\rho = 1.2$, and $\beta_{max} = 10^{10}$.

Output: the denoising result.

1: Initialize $\mathcal{W}_s^0$, $\mathcal{Z}^0$, $\mathcal{S}^0$, $\mathcal{L}_s^0$, $\mathcal{U}^0$, maximum iterations $T_{\max} = 500$.
2: **While** $\frac{\|\mathcal{Z}^{t+1}-\mathcal{Z}^t\|_F}{\|\mathcal{Z}^t\|_F} > 10^{-4}$ and $t \le T_{\max}$ **Do**
3: Updating $\mathcal{W}_s$ via (9);
4: Updating $\mathcal{Z}$ via (10);
5: Updating $\mathcal{S}$ via (11);
6: Updating $\mathcal{L}_s$ via (12);
7: Updating $\mathcal{U}$ via (13);
8: **End**
9: **Return**: inv-Permute$(\mathcal{A}*\mathcal{Z}^{t+1}, [1,3,2])$.

The solving algorithm is summarized in Algorithm 1, which will be named as the TenMSR algorithm in the sequel. The dictionary $\mathcal{A}$ can be estimated by some preprocessing approaches (see Section V). Moreover, the tensor multi-subspace structure can be automatically divided based on the learned representation tensor $\mathcal{Z}$.

B. Computational Complexity

The proposed algorithm involves multiple variables $\mathcal{W}_s$, $\mathcal{Z}$, $\mathcal{S}$, $\mathcal{L}_s$, and $\mathcal{U}$. For $\mathcal{W}_s \in \mathbb{R}^{r\times n_3\times n_2}$-subproblem, since DFT and inverse DFT cost $\mathcal{O}(rn_3n_2^2)$ and computing n_2 singular value decompositions on $r\times n_3$ costs $\mathcal{O}(r^2n_2n_3)$, the computational cost of $\mathcal{W}_s$-subproblem is $\mathcal{O}(r^2n_2n_3+rn_3n_2^2)$. For $\mathcal{Z}\in\mathbb{R}^{r\times n_3\times n_2}$-subproblem, the computational cost is $\mathcal{O}(rn_2^2n_3)$. For $\mathcal{S}\in\mathbb{R}^{n_1\times n_3\times n_2}$-subproblem, the main cost is to calculate $\mathcal{A}*\mathcal{U}$, which costs $\mathcal{O}(rn_1n_2n_3)$. For $\mathcal{L}_s\in\mathbb{R}^{r\times n_3\times n_2}$-subproblem, the cost of Newton method is $\mathcal{O}(rn_2n_3)$. For $\mathcal{U}\in\mathbb{R}^{r\times n_3\times n_2}$-subproblem, the main computation depends on the t-product at the cost of $\mathcal{O}(rn_1n_2n_3)$. Therefore, the computational complexity of each iteration of the proposed algorithm is $\mathcal{O}(r^2n_2n_3+rn_2^2n_3+rn_1n_2n_3)$.

C. Convergence Analysis

We establish the theoretical convergence of the proposed algorithm in the following theorem.

Theorem 1: For the sequence $\{\mathcal{W}_s,\mathcal{Z},\mathcal{S},\mathcal{L}_s,\mathcal{U}\}$ generated from Algorithm 1, it can globally converge to a critical point of Problem (7).

According to the framework in [64], we need to verify the following three conditions.

1) $f(\mathcal{W}_s^t,\mathcal{Z}^t,\mathcal{S}^t,\mathcal{L}_s^t,\mathcal{U}^t)$ is a proper lower semi-continuous function.
2) $f(\mathcal{W}_s^t,\mathcal{Z}^t,\mathcal{S}^t,\mathcal{L}_s^t,\mathcal{U}^t)$ is a Kurdyka–Łojasiewicz (KŁ) function and satisfies the Kproperty.
3) $(\mathcal{W}_s^t,\mathcal{Z}^t,\mathcal{S}^t,\mathcal{L}_s^t,\mathcal{U}^t)$ satisfies the decrease condition and relative error condition. Therefore, we show the details of each condition.

The proof can be found in the Supplementary Materials.

Remark 1: We clarify the effect of parameters on the convergence of the proposed algorithm. Among the parameters in our algorithm, the proximal parameter ρ plays an important role in the sufficient descent and relative error conditions of the convergence theorem. Specifically, in the sufficient descent condition, the parameter ρ should satisfy $\rho \ge L/2$, where L is the gradient Lipschitz constant [64], to ensure that the objective function value decreases monotonically. In the relative error condition, ρ controls the accuracy of the solution of the subproblem, and small ρ ensures a small approximation error. Therefore, to guarantee the convergence of the proposed algorithm, the parameter ρ should be small, but $\rho \ge L/2$. The results shown in Section VI-D demonstrate the numerical convergence of our algorithm.

V. Numerical Experiments

In this section, we evaluate the performance of the proposed TenMSR on semi-real RSI (including multitemporal RSI and HSIs) and real RSI. The test data are normalized to $[0,1]$ band by band. All numerical experiments are tested in MATLAB R2018b with an Intel[1] Core[2] i7-8700M CPU with 3.70 GHz and 32 GB of RAM on a desktop.

Baselines: We select ten RSIs denoising methods, including the following.

1) *Four Original Space Representation-Based Methods:* Low-rank matrix recovery (LRMR) [24], Kronecker-basis-representation-based tensor sparsity measure (KBR) [67], low-rank tensor recovery (LRTR) [37], and 3-D tensor nuclear norm (3DTNN) [38].
2) *Four Single Subspace Representation-Based Methods:* Fast HSI denoising based on low-rank and sparse representations (FastHyDe) [68], spectral-spatial L_0 gradient regularized low-rank tensor factorization (LRTFL0) [69], nonlocal meets global (NGmeet) [58], and Eigen-image2Eigenimage (E2E) [8].
3) *Two Deep Learning Methods:* Three-dimensional spatial–spectral attention Transformer (TDSAT) [70] and unsupervised feature fusion-guided network (UFFGNet) [5].

Parameter Settings: The proposed TenMSR contains regularization parameters λ_{1k} $(k=1,2,3)$ and λ_2 and penalties parameters β_1, β_2, β_{3k} $(k=1,2,3)$, β_4, and ρ. In our experiments, we empirically set the regularization parameters

[1]Registered trademark.
[2]Trademarked.

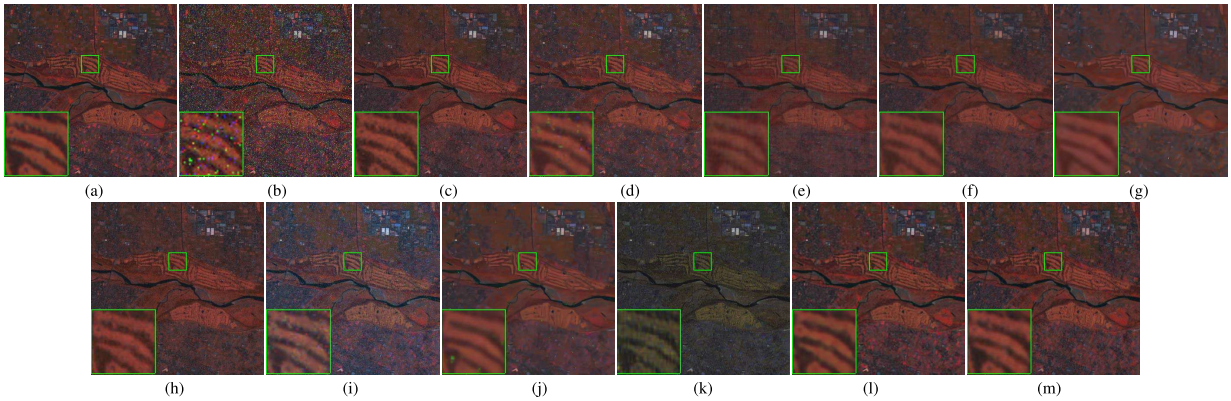

Fig. 5. Denoising results on Sentinel2A dataset under Gaussian noise $\sigma = 0.05$ and impulse noise $\nu = 0.05$. (a) Original. (b) Noisy. (c) LRMR. (d) KBR. (e) LRTR. (f) 3DTNN. (g) FastHyDe. (h) LRTFL0. (i) NGmeet. (j) E2E. (k) TDSAT. (l) UFFGNet. (m) TenMSR.

in the range of $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$. β_1 is selected from $[1, 10]$ with increment of 1. We empirically choose β_{3k} $(k = 1, 2, 3)$ and β_4 from the set $\{10^{-4}, 10^{-3}, 10^{-2}, 5 * 10^{-2}, 10^{-1}, 5 * 10^{-1}\}$. In addition, β_2 and ρ are set to 0.0001 in all experiments. The nonlinear transform ϕ = Tanh is chosen in our experiments. We use the estimated data by 3DTNN [38] as the dictionary $\mathcal{A}$. The parameters of baselines are adjusted according to the rules suggested in the corresponding papers, which can be found in the Supplementary Materials.

Evaluation Metrics: The peak signal-to-noise ratio (PSNR) (dB) and the structural similarity index (SSIM) [71] are commonly used metrics to evaluate the quality of recovered results in image processing tasks. Specifically, PSNR is a measure of the difference between the original image and the reconstructed image, and SSIM is a metric that assesses the structural similarity between two images. By calculating the mean of PSNR (MPSNR) and mean of SSIM (MSSIM) values of all bands, we can obtain the metrics of whole spectral images. Higher (M)PSNR and (M)SSIM values indicate better recovered image quality.

A. Experiments on Multitemporal RSIs

To evaluate the effectiveness of the proposed TenMSR in removing mixed noises, we perform the experiments on two multitemporal remote sensing data.[3] Due to the difference between spectrums in different periods, it is appropriate to assume that the multitemporal RSI lies in the multiple subspaces. We set the images corrupted by *Gaussian noise* and *salt and pepper noise*, i.e., all bands of images are corrupted by zero-mean Gaussian noise with different variances σ and salt and pepper noise with different proportions ν.

The first experiment is obtained by Sentinel2A. The Sentinel2A data contain the spectral image of four time nodes (i.e., 20181024, 20181103, 20181113, and 20181203[4]), and the size of image of each time node is $400 \times 400 \times 4$.

Table I lists the quantitative results by all methods on Sentinel2A data corrupted by different Gaussian noise and salt and pepper noise levels. Compared with baselines, the proposed TenMSR achieves better performance with respect to the MPSNR and MSSIM values while also remaining competitive in running time. Fig. 5 shows the visual effects of recovered images under Gaussian noise $\sigma = 0.05$ and salt and pepper noise $\nu = 0.05$. These results show that all methods can suppress the most mixed noise. Nevertheless, the images restored by LRMR, KBR, LRTR, 3DTNN, and TDSAT remain a small amount noise. FastHyDe, LRTFL0, and E2E oversmooth the edges and lose some detailed information. The reason is that the methods based on the original space and single subspace representation cannot fully exploit the essential structure of multitemporal RSIs. In comparison, by mining the multi-subspace structure of data, the proposed method can effectively preserve the sharp edges and smooth regions of RSIs.

For the second experiment, the used multitemporal image captured by Landsat8 from seven time nodes (i.e., 20180619, 20180721, 20180806, 20180923, 20181009, 20181025, and 20181116). The image of each time node is of size $400 \times 400 \times 7$.

Table II presents the MPSNR and MSSIM values recovered by all methods. Again, the proposed method achieves the best performance with the acceptable time cost in most cases for different noise levels, which shows the robustness of our method. Fig. 6 shows the recovery results of the Landsat8 data under Gaussian noise $\sigma = 0.10$ and salt and papper noise $\nu = 0.10$. Compared with baselines, TenMSR can effectively reduce the mixed noise and keep the detail sharpness and structures. The quantitative and visual results on multitemporal images prove the effectiveness of the multiple tensor subspace low-rank representation of the proposed TenMSR.

Fig. 7 shows the PSNR and SSIM curves of all bands of Landsat8 under two cases $\sigma = 0.05$ and $\nu = 0.05$, and $\sigma = 0.1$ and $\nu = 0.1$, respectively. Clearly, our TenMSR obtains the highest PSNR and SSIM values in most bands.

To verify the effectiveness of the proposed TenMSR in processing multi-subspace data, Fig. 8 shows the affinity matrix $\tilde{\boldsymbol{Z}}$ learned by the proposed method on the Landset8 dataset with different noise, which contains the multiple subspace division structure. Here, the affinity matrix can be obtained by the representation tensor $\mathcal{Z}$, i.e., $\tilde{\boldsymbol{Z}} = 1/(2n_3)\Sigma_{k=1}^{n_3}(|\mathcal{Z}^{(k)}| + |(\mathcal{Z}^{(k)})^\top|)$. One can see that the representation coefficients have a clear block-diagonal structure under different mixed noise levels. Each block represents a subspace, and Fig. 8 implies that the Landset8 dataset is distributed in seven subspaces. Therefore, the proposed TenMSR can finely mine the tensor multi-subspace structure of the data—that is beneficial to deal with the noise removal problem.

[3]https://theia.cnes.fr/atdistrib/rocket/#/signin

[4]The number "20181203" means the time December 3, 2018.

TABLE I

PERFORMANCE [MPSNR, MSSIM, AND TIME (SECONDS)] OF DATA RECOVERY ON SYNTHETIC DATASET SENTINEL2A WITH DIFFERENT NOISE LEVELS. THE **BEST** VALUES ARE HIGHLIGHTED BY BOLD FRONTS

Gaussian	Salt and pepper	Index	Noisy	LRMR	KBR	LRTR	3DTNN	FastHyDe	LRTFL0	NGmeet	E2E	TDSAT	UFFGNet	TenMSR
$\sigma = 0.05$		MPSNR	14.43	29.49	27.25	27.48	31.15	26.81	30.72	25.02	26.60	25.46	30.64	**32.61**
		MSSIM	0.12	0.6870	0.6916	0.6408	0.8174	0.7148	0.7612	0.5392	0.7100	0.5715	0.7843	**0.8434**
		Time	-	21.70	281.60	19.21	18.37	0.80	243.20	133.20	9.20	**0.65**	97.38	11.89
$\sigma = 0.10$	$\nu = 0.10$	MPSNR	13.63	26.32	26.98	27.43	29.35	26.74	27.74	27.09	26.43	23.77	29.03	**30.73**
		MSSIM	0.09	0.4935	0.6788	0.6302	0.7363	0.7087	0.5818	0.7068	0.6994	0.4556	0.7262	**0.7582**
		Time	-	20.40	220.60	19.84	19.68	0.70	217.80	119.70	9.24	**0.65**	98.43	17.86
$\sigma = 0.20$		MPSNR	11.34	22.29	26.55	25.38	26.36	26.48	24.28	26.05	25.99	21.43	24.86	**28.57**
		MSSIM	0.05	0.2705	0.6600	0.4290	0.4794	**0.6840**	0.3639	0.6313	0.6750	0.2979	0.5658	0.6521
		Time	-	21.90	217.30	31.63	40.14	0.80	192.10	121.90	9.25	**0.65**	98.19	11.82
$\sigma = 0.05$	$\nu = 0.05$	MPSNR	17.17	31.32	26.13	30.80	31.44	28.85	30.97	26.21	28.83	26.80	31.25	**33.13**
		MSSIM	0.22	0.7649	0.5910	0.7593	0.8271	0.7496	0.7788	0.5645	0.7416	0.6446	0.8031	**0.8551**
		Time	-	19.50	123.20	19.60	21.66	0.70	199.20	100.70	9.22	**0.63**	97.95	12.05
	$\nu = 0.20$	MPSNR	11.58	25.82	23.09	27.04	30.64	23.15	29.63	23.13	22.98	23.65	30.34	**31.66**
		MSSIM	0.05	0.4909	0.6191	0.6228	0.7981	0.6586	0.7034	0.6339	0.6485	0.4530	0.7837	**0.8043**
		Time	-	24.60	248.50	24.61	24.91	0.80	265.70	149.60	9.22	**0.64**	97.29	17.51
	$\nu = 0.30$	MPSNR	9.87	22.42	20.31	26.46	30.40	19.96	27.97	19.92	20.20	22.34	29.68	**30.86**
		MSSIM	0.03	0.3106	0.5753	0.5985	0.7799	0.5808	0.6208	0.5597	0.5976	0.3599	0.7591	**0.7871**
		Time	-	79.60	486.80	48.51	44.44	2.00	380.50	4009.80	9.23	**0.64**	97.24	34.07

TABLE II

PERFORMANCE [MPSNR, MSSIM, AND TIME (SECONDS)] OF DATA RECOVERY ON SYNTHETIC LANDSAT8 DATASET WITH DIFFERENT NOISE LEVELS. THE **BEST** VALUES ARE HIGHLIGHTED BY BOLD FRONTS

Gaussian	Salt and pepper	Index	Noisy	LRMR	KBR	LRTR	3DTNN	FastHyDe	LRTFL0	NGmeet	E2E	TDSAT	UFFGNet	TenMSR
$\sigma = 0.05$		MPSNR	14.04	28.66	24.88	25.73	29.63	24.15	21.97	22.47	24.73	23.98	29.38	**31.46**
		MSSIM	0.17	0.7641	0.5958	0.6456	0.8316	0.6306	0.6326	0.4940	0.6576	0.6279	0.7820	**0.8430**
		Time	-	66.30	355.00	61.79	67.43	**1.00**	356.60	101.70	9.05	1.46	99.47	49.12
$\sigma = 0.10$	$\nu = 0.10$	MPSNR	13.28	26.69	24.97	24.84	27.22	24.05	21.50	22.64	24.61	23.48	27.26	**28.99**
		MSSIM	0.13	0.6629	0.6125	0.5696	0.7343	0.6213	0.5738	0.5729	0.6490	0.6022	0.7222	**0.7535**
		Time	-	64.40	1136.50	63.70	75.46	**1.00**	454.60	117.30	9.14	1.45	98.57	132.13
$\sigma = 0.20$		MPSNR	11.13	23.44	23.99	21.13	23.56	23.77	20.64	22.41	24.13	22.39	25.05	**25.98**
		MSSIM	0.08	0.4727	0.5500	0.2960	0.4472	0.5940	0.4663	0.5482	**0.6150**	0.5429	0.6049	0.5997
		Time	-	64.40	1010.80	78.5610	111.65	**1.10**	448.70	120.90	9.13	1.45	98.79	50.32
$\sigma = 0.05$	$\nu = 0.05$	MPSNR	16.78	29.15	27.09	26.08	30.04	26.97	22.04	24.49	26.76	24.72	29.33	**31.99**
		MSSIM	0.29	0.7817	0.6708	0.6655	0.8420	0.7265	0.6388	0.5606	0.7291	0.6566	0.7992	**0.8549**
		Time	-	63.30	324.70	124.83	108.02	**0.90**	448.50	103.90	12.79	1.86	100.10	81.20
	$\nu = 0.20$	MPSNR	11.17	26.92	21.09	24.95	28.81	20.95	21.79	20.04	21.13	23.56	28.10	**30.53**
		MSSIM	0.08	0.6948	0.4982	0.5994	0.8080	0.5269	0.6162	0.4811	0.5432	0.6027	0.7468	**0.8111**
		Time	-	106.30	1454.60	185.90	88.00	**1.20**	488.20	177.40	9.11	1.46	98.71	94.14
	$\nu = 0.30$	MPSNR	9.46	24.33	18.22	23.92	27.56	18.22	21.56	17.82	18.34	23.56	27.63	**29.01**
		MSSIM	0.05	0.5649	0.3995	0.5287	0.7279	0.4408	0.5939	0.4016	0.4487	0.5901	0.7288	**0.7400**
		Time	-	150.10	1297.50	186.14	156.84	**1.80**	620.30	1325.20	9.13	1.56	99.02	60.52

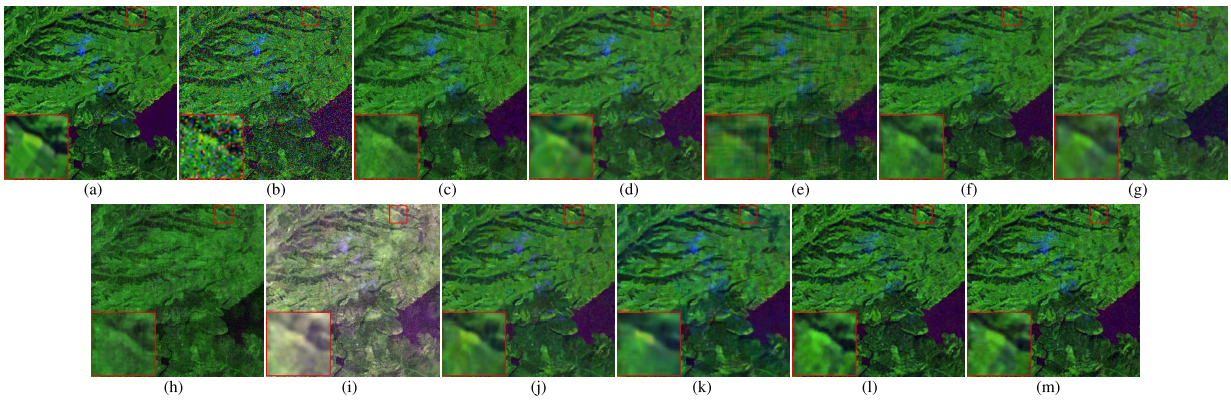

Fig. 6. Denoising results on Landsat8 dataset under Gaussian noise $\sigma = 0.10$ and impulse noise $\nu = 0.10$. (a) Original. (b) Noisy. (c) LRMR. (d) KBR. (e) LRTR. (f) 3DTNN. (g) FastHyDe. (h) LRTFL0. (i) NGmeet. (j) E2E. (k) TDSAT. (l) UFFGNet. (m) TenMSR.

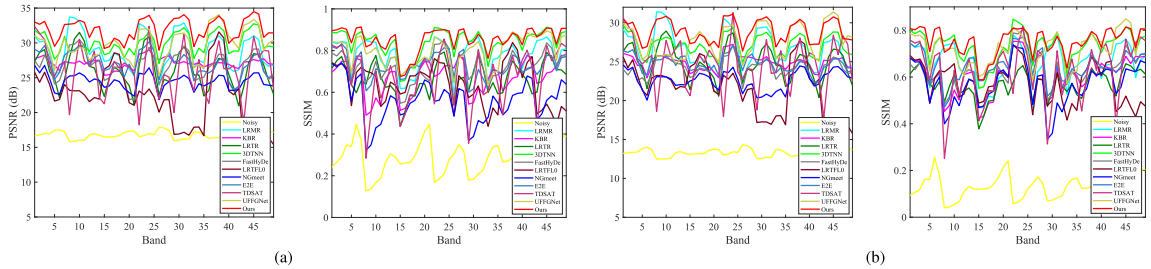

Fig. 7. Performance (PSNR and SSIM values) of each band by different methods on Landset8 dataset. (a) $\sigma = 0.05$ and $\nu = 0{:}05$. (b) $\sigma = 0.10$ and $\nu = 0.10$.

B. Experiments on HSIs

As discussed above, due to the multitemporal property, the TenMSR can finely characterize the distribution of multitemporal RSI, which is better than the single tensor subspace assumption.

Actually, for the HSI, usually containing hundreds of frequency bands, we can also use the TenMSR to finely deliver

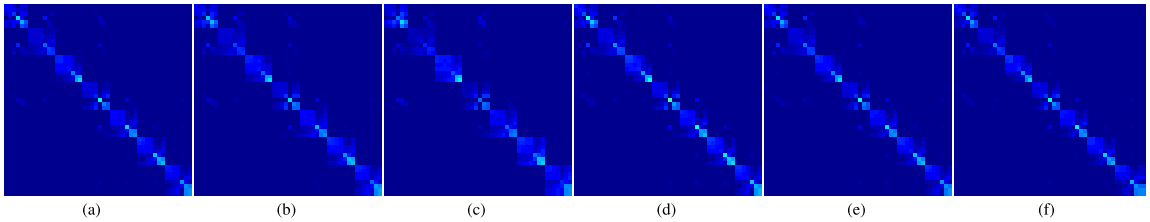

Fig. 8. multi-subspace structure on Landset8 dataset under different Gaussian noise and impulse noise settings. (a) $\sigma = 0.05$ and $\nu = 0.10$. (b) $\sigma = 0.10$ and $\nu = 0.10$. (c) $\sigma = 0.20$ and $\nu = 0.10$. (d) $\sigma = 0.05$ and $\nu = 0.05$. (e) $\sigma = 0.05$ and $\nu = 0.20$. (f) $\sigma = 0.05$ and $\nu = 0.30$.

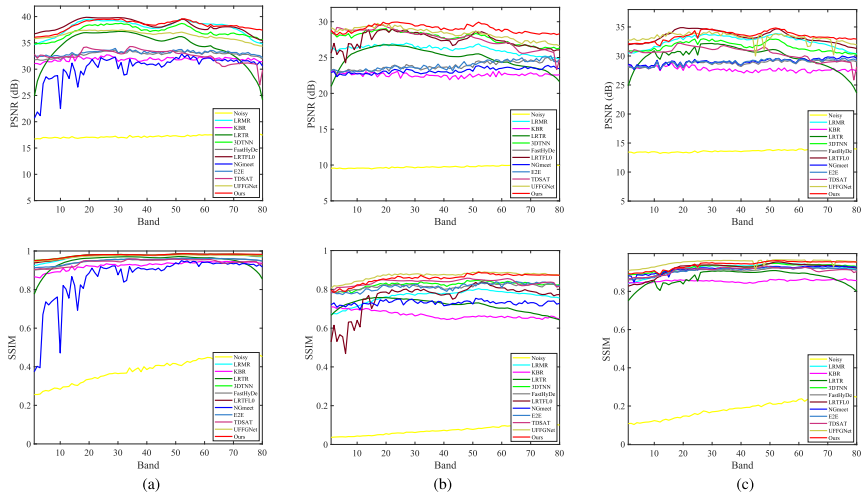

Fig. 9. Performance (PSNR and SSIM) of each band by different methods on the PaC dataset. (a) Case 1. (b) Case 3. (c) Case 5.

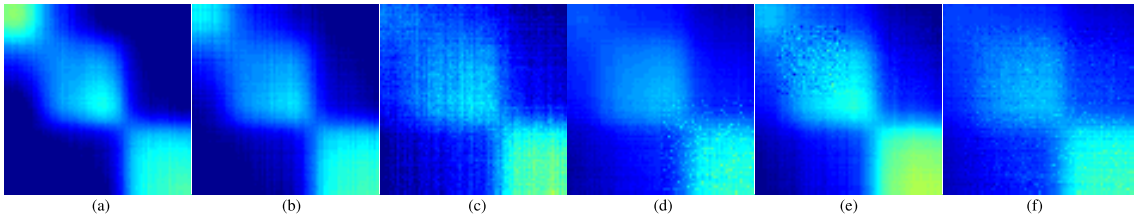

Fig. 10. multi-subspace structure of PaC dataset under different cases. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4. (e) Case 5. (f) Case 6.

the intrinsic structure of HSI due to the wavelength difference along the spectral dimension.

In this section, we conduct experiments on two representative datasets to evaluate the noise removal performance of our TenMSR.

The data include hyperspectral Pavia City Center data (PaC) and HYDICE Washington DC Mall data (WDC).[5] The PaC dataset with size $200 \times 200 \times 80$ uses the subscene of the image, which is acquired by the ROSIS sensor over Pavia, northern Italy. The WDC dataset ($256 \times 256 \times 191$) is the subimage of HSI obtained by an airborne hyperspectral data flight line over the Washington DC Mall. To simulate mixed noise scenarios, we conduct the following different noise cases.

Cases 1–3 *(Gaussian Noise + Salt and Pepper Noise)*: All bands are added to the mixtures of zero-mean Gaussian noise and impulse noise. For Cases 1–3, the variances σ of Gaussian noise and the proportion ν of salt and pepper noise are set to $\sigma = \nu = 0.05, 0.10, 0.20$.

Case 4 *(Gaussian Noise + Salt and Pepper Noise + Dead Lines)*: All bands are added to the mixtures of zero-mean Gaussian noise, salt and pepper noise, and dead lines. The variance of Gaussian noise is set as 0.10 and the percentages of salt and pepper noise are 0.10 in all bands. In addition, about 25% of bands are corrupted by dead lines and the width of deadlines is randomly generated from 1 to 3.

Case 5 *(Gaussian Noise + Salt and Pepper Noise + Stripes)*: All bands are added to the mixtures of zero-mean Gaussian noise, salt and pepper noise, and stripes. The settings of Gaussian noise and salt and pepper noise are the same as Case 4. In addition, about 15% of bands are corrupted by stripes and the number of stripes is from 10 to 20.

Case 6 *(Gaussian Noise + Salt and Pepper Noise + Dead Lines + Stripes)*: All bands are added to the mixtures of zero-mean Gaussian noise, salt and pepper noise, dead lines, and stripes. The distributions of Gaussian noise and salt and pepper noise are the same as Case 4. In addition, the distribution of deadlines is the same as Case 4, and the distribution of strip contamination is the same as Case 5.

Tables III and IV list the restoration results by all methods in terms of MPSNR, MSSIM, and Time values for PaC and WDC datasets. It is clear that under different noise settings, the proposed TenMSR achieves the better performance than comparison methods in most cases. The original space representation-based method 3DTNN and deep learning method UFFGNet obtain suboptimal results. This suggests

[5]https://rslab.ut.ac.ir/data

TABLE III

PERFORMANCE [MPSNR, MSSIM, AND TIME (SECONDS)] OF DATA DENOISING BY DIFFERENT METHODS ON SIMULATED PAC DATASET. THE **BEST** VALUES ARE HIGHLIGHTED BY BOLD FRONTS

Case	Index	Noisy	LRMR	KBR	LRTR	3DTNN	FastHyDe	LRTFL0	NGmeet	E2E	TDSAT	UFFGNet	TenMSR
Case 1	MPSNR	17.20	37.97	31.74	34.32	37.16	32.91	38.37	30.44	33.11	32.93	36.41	**38.39**
	MSSIM	0.38	0.9735	0.9224	0.9466	0.9771	0.9481	0.9747	0.8653	0.9501	0.9428	0.9729	**0.9779**
	Time	-	52.50	208.30	214.13	75.61	**0.40**	155.60	27.60	2.56	0.60	30.63	46.52
Case 2	MPSNR	13.64	32.61	27.64	30.17	32.83	28.73	33.38	29.06	28.84	30.84	33.76	**33.98**
	MSSIM	0.19	0.9229	0.8525	0.8819	0.9406	0.9069	0.9225	0.9166	0.9092	0.9113	**0.9524**	0.9437
	Time	-	53.20	449.50	27.71	37.40	**0.40**	144.40	31.00	2.59	0.57	30.96	26.81
Case 3	MPSNR	9.73	26.20	22.56	24.98	27.76	23.83	27.26	23.15	23.86	28.05	28.30	**29.08**
	MSSIM	0.07	0.7639	0.6672	0.7181	0.8226	0.8161	0.7584	0.7268	0.8138	0.8386	**0.8703**	0.8576
	Time	-	52.60	399.00	121.61	63.05	**0.40**	249.80	31.80	2.58	0.56	46.47	99.05
Case 4	MPSNR	13.64	32.43	27.73	29.46	31.98	28.90	**33.25**	29.22	29.02	30.88	32.62	32.62
	MSSIM	0.19	0.9210	0.8495	0.8648	0.9289	0.9051	0.9209	0.9152	0.9075	0.9079	**0.9443**	0.9346
	Time	-	113.30	584.00	123.43	109.09	1.10	177.20	639.30	2.58	**0.61**	31.38	99.17
Case 5	MPSNR	13.65	32.53	27.74	29.77	31.59	28.77	33.35	29.10	28.93	30.90	33.09	**33.44**
	MSSIM	0.19	0.9218	0.8553	0.8701	0.9247	0.9066	0.9217	0.9168	0.9097	0.9087	**0.9518**	0.9390
	Time	-	72.60	588.70	96.11	58.30	**0.60**	195.70	50.60	2.64	0.61	31.61	75.99
Case 6	MPSNR	13.53	32.13	27.74	28.88	30.61	28.84	**33.18**	29.17	29.10	30.87	32.98	31.62
	MSSIM	0.18	0.9199	0.8518	0.8513	0.9061	0.9042	0.9192	0.9137	0.9081	0.9076	**0.9463**	0.9205
	Time	-	82.30	574.10	105.80	51.34	**0.50**	165.70	101.80	2.97	0.61	31.48	87.13

TABLE IV

PERFORMANCE [MPSNR, MSSIM, AND TIME (SECONDS)] OF DATA DENOISING BY DIFFERENT METHODS ON SIMULATED WDC DATASET. THE **BEST** VALUES ARE HIGHLIGHTED BY BOLD FRONTS

Case	Index	Noisy	LRMR	KBR	LRTR	3DTNN	FastHyDe	LRTFL0	NGmeet	E2E	TDSAT	UFFGNet	TenMSR
Case 1	MPSNR	16.86	37.55	32.26	33.51	37.10	32.67	34.39	31.76	32.69	31.47	35.17	**38.79**
	MSSIM	0.34	0.9650	0.9207	0.9294	0.9597	0.9274	0.9534	0.8834	0.9300	0.8944	0.9623	**0.9747**
	Time	-	118.00	859.70	123.40	156.70	**0.80**	477.50	48.60	4.04	3.08	44.21	352.67
Case 2	MPSNR	13.35	33.62	27.67	28.53	31.72	28.12	32.63	28.88	27.99	30.08	31.14	**34.52**
	MSSIM	0.17	0.9093	0.8367	0.7701	0.8576	0.8636	0.9251	0.8900	0.8655	0.8660	0.9123	**0.9354**
	Time	-	119.10	2068.70	239.74	295.68	**1.50**	488.90	51.80	4.04	3.34	44.43	561.74
Case 3	MPSNR	9.48	25.28	22.22	20.05	27.19	23.18	28.52	22.91	22.66	27.89	28.57	**28.74**
	MSSIM	0.07	0.7153	0.6434	0.3315	0.7561	0.7434	0.8221	0.6895	0.7453	0.7944	**0.8499**	0.8114
	Time	-	118.70	1904.80	272.84	226.16	**1.00**	591.90	51.80	4.01	2.48	44.84	363.36
Case 4	MPSNR	13.32	33.49	27.60	28.08	31.07	28.19	32.48	28.94	28.08	30.04	32.44	**33.67**
	MSSIM	0.17	0.9085	0.8285	0.7605	0.8469	0.8627	0.9199	0.8875	0.8686	0.8653	**0.9323**	0.9269
	Time	-	154.80	2867.20	284.93	316.79	**1.00**	650.60	239.10	4.09	2.75	45.68	350.63
Case 5	MPSNR	13.33	33.53	27.52	28.19	31.24	28.10	32.44	28.78	27.97	30.05	33.26	**33.96**
	MSSIM	0.17	0.9078	0.8329	0.7583	0.8431	0.8636	0.9192	0.8882	0.8651	0.8656	**0.9419**	0.9261
	Time	-	200.30	2461.20	319.95	398.54	**1.40**	514.00	53.90	4.01	2.32	44.59	447.57
Case 6	MPSNR	13.31	**33.40**	27.47	27.74	30.50	28.17	32.57	28.94	28.08	30.00	32.19	32.64
	MSSIM	0.16	0.9066	0.8269	0.7487	0.8315	0.8613	0.9231	0.8878	0.8673	0.8644	**0.9295**	0.9083
	Time	-	170.70	2278.40	162.20	213.02	**1.20**	1083.40	53.40	4.82	2.34	43.90	553.40

that the representation ability of tensor multi-subspace is stronger than that of the original space representation. Besides, UFFGNet utilizes the high-quality guidance image to extract semantic features of images, thereby enhancing the restoration results. In contrast, without the guidance image, TenMSR achieves a competitive performance by exploring the tensor multi-subspace prior of HSI. Fig. 9 presents the PSNR and SSIM values across all bands of the recovered results on the PaC dataset under Cases 1, 3, and 5. Obviously, our method achieves the highest PSNR values in most bands. This also illustrates the effectiveness of the proposed TenMSR for RSI mixed noise removal.

To show the ability of our method in exploiting the multiple tensor subspace relationship embedded in HSI, Fig. 10 displays the structure of representations learned by the proposed TenMSR on PaC dataset under Cases 1–6. We observe that all of the representation coefficients show obvious group effects and have clear block-diagonal structures, especially for Cases 1 and 2. From the block-diagonal structures of representation coefficients, one can conclude that the PaC dataset is distributed in three tensor subspaces, which is consistent with the above discussion. Fig. 10 verifies the ability and superiority of the proposed tensor in exploring the multiple tensor subspace representation for HSI.

Figs. 11 and 12 show the pseudo-clolor images of recovery results of PaC and WDC datasets by different methods under different cases, respectively. From these recovered images, we see that TenMSR achieves the promising performance than the compared methods. To be specific, in Case 5 of PaC dataset, the recovered results of LRMR, LRTR, 3DTNN, and TDSAT still have residual noise or stripes. FastHyDe, LRTFL0, NGmeet, E2E, and UFFGNet recover the coarse structure of the images but blur the image details. In comparison, the images obtained by our method are closest to the original ones. Similarly, from the zoomed-in area of WDC in Case 3, one can observe that the proposed TenMSR outperforms the compared methods in removing mixed noise and recovering image details and textures. More experimental results are shown in the Supplementary Materials.

C. Experiments on Real RSIs

We also test the performance of all methods on real RSIs, including Urban data and GaoFen-5 data.[6] Since there is no clean data as the References, we evaluate the effectiveness using the visual effects of recovered images. The parameters involved in all methods are adjusted to achieve the best visual performance.

The Urban dataset is one of the most widely used RSI, which is collected by the HYDICE sensor. The spatial size of Urban dataset is 307×307 and contains 210 spectral bands. This image is heavily contaminated with Gaussian noise and stripes. Fig. 13 shows the recovered images of the original data and denoising results by different methods. In order to conduct a more detailed comparison, we zoom in on a specific region in each subfigure, indicated by a red box. Obviously, the recovered images of KBR, LRTR, and 3DTNN

[6]https://rslab.ut.ac.ir/data

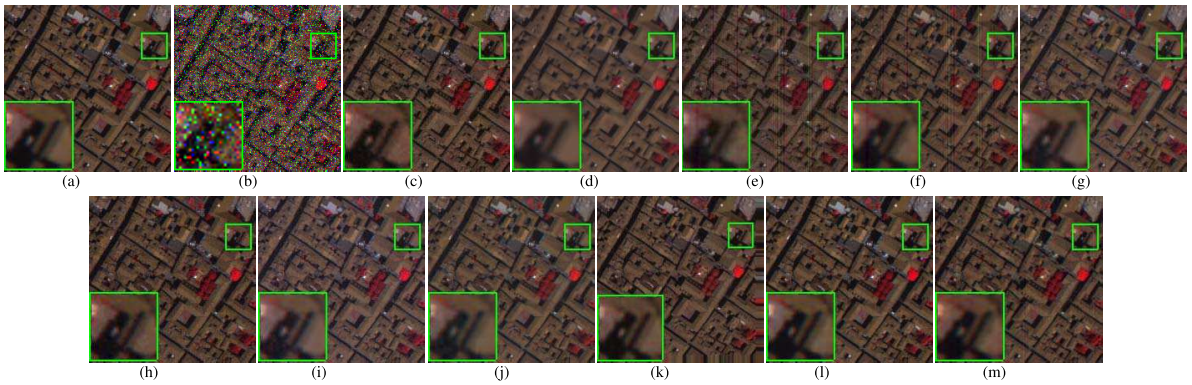

Fig. 11. Pseudo-color recovered results (R-G-B: 70-29-10 bands) of different methods on PaC dataset under Case 5. (a) Original. (b) Noisy. (c) LRMR. (d) KBR. (e) LRTR. (f) 3DTNN. (g) FastHyDe. (h) LRTFL0. (i) NGmeet. (j) E2E. (k) TDSAT. (l) UFFGNet. (m) TenMSR.

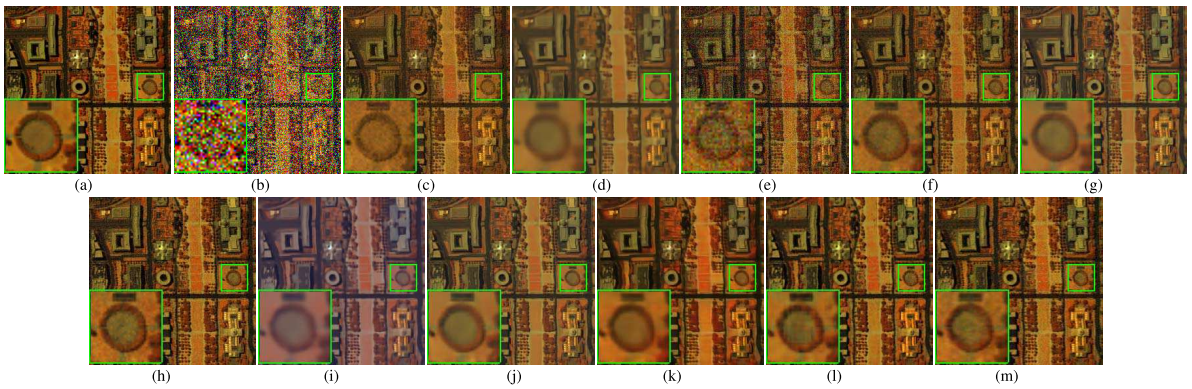

Fig. 12. Pseudo-color recovered results (R-G-B: 79-120-170 bands) of different methods on WDC dataset under Case 3. (a) Original. (b) Noisy. (c) LRMR. (d) KBR. (e) LRTR. (f) 3DTNN. (g) FastHyDe. (h) LRTFL0. (i) NGmeet. (j) E2E. (k) TDSAT. (l) UFFGNet. (m) TenMSR.

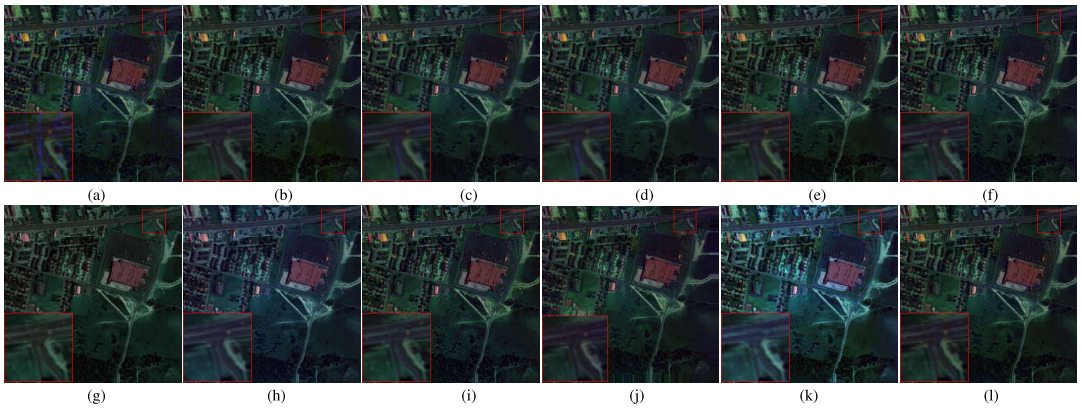

Fig. 13. Pseudo-color recovered results (R-G-B: 30-110-207 bands) of different methods on real Urban dataset. (a) Observed. (b) LRMR. (c) KBR. (d) LRTR. (e) 3DTNN. (f) FastHyDe. (g) LRTFL0. (h) NGmeet. (i) E2E. (j) TDSAT. (k) UFFGNet. (l) TenMSR.

still exist stripe noise. The denoising results of FastHyDe, LRTFL0, NGmeet, E2E, and TDSAT have noticeable image edge blurring. UFFGNet can better preserve image details, but the restored image appears to have a color cast. As a comparison, the proposed TenMSR successfully maintains the structure and details of the image while removing stripes.

The GaoFen-5 dataset is collected by the SAST-5000B satellite platform. GaoFen-5 data have 256×256 spatial pixels and 155 spectral bands and mainly contain the Gaussian noise and the stripe noise. Fig. 14 shows the pseudo-color images of the original data and denoising results of different methods. Compared with the baselines, our method shows superiority in removing stripe noise and preserving the overall structure of the image. Moreover, in order to verify the ability of our method in capturing the multiple tensor subspace structure of real data, Fig. 15 presents the multi-subspace structures of the affinity matrix on two real datasets. From the visual effects, one can see that the Urban dataset lies in four subspaces, and the GaoFen-5 dataset can be roughly divided into two groups, i.e., lies in two subspaces. This figure shows that the proposed TenMSR can effectively and superiorly identify the tensor multi-subspace structure of real spectral image—that is ignored in the single matrix/tensor subspace representation methods.

VI. Discussion

In this section, we will discuss the effects of number of subspaces on denoising performance for the proposed TenMSR,

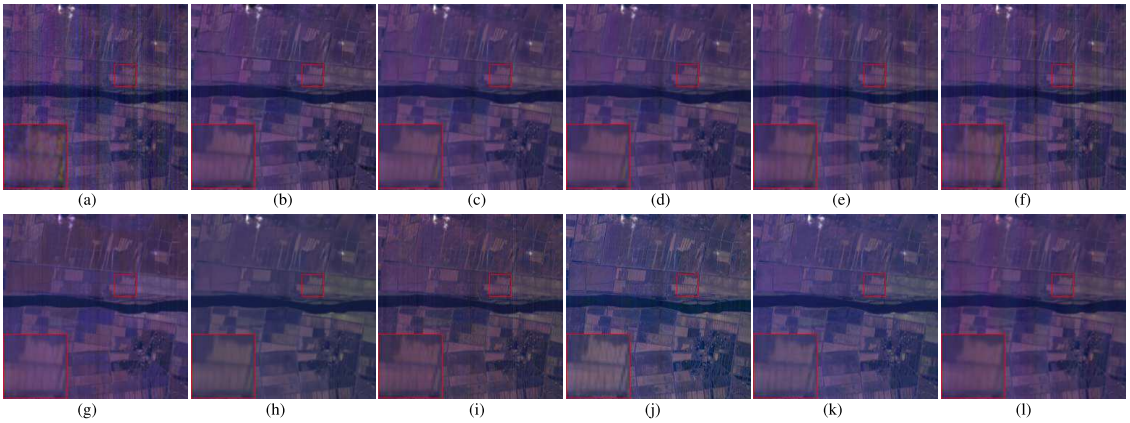

Fig. 14. Pseudo-color recovered results (R-G-B: 96-151-54 bands) of different methods on real GaoFen-5 dataset. (a) Observed. (b) LRMR. (c) KBR. (d) LRTR. (e) 3DTNN. (f) FastHyDe. (g) LRTFL0. (h) NGmeet. (i) E2E. (j) TDSAT. (k) UFFGNet. (l) TenMSR.

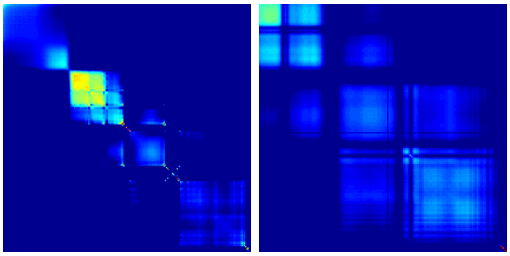

Fig. 15. Multi-subspace structure of the proposed method on (a) Urban and (b) GaoFen-5 datasets.

TABLE V

DENOISING PERFORMANCE BY DIFFERENT SETTINGS ON SENTINEL2A AND PAVIA DATASETS UNDER $\sigma = 0.10$ AND $\nu = 0.10$

Different functions	Index	LeakyReLU	ReLU	Softmax	Softsign	Tanh
Sentinel2A	MPSNR	30.73	30.72	30.70	30.71	30.73
	MSSIM	0.7580	0.7576	0.7580	0.7578	0.7582
PaC	MPSNR	33.97	33.97	33.98	33.99	33.98
	MSSIM	0.9434	0.9436	0.9435	0.9435	0.9437

the effects of the nonlinear transform, the effects of regularization of the TenMSR coefficient, and the numerical convergence behavior of the proposed algorithm.

A. Effects of Number of Subspaces on TenMSR

For the acquired (multitemporal) RSIs, due to the wavelength difference or collected period variability, the spectral bands of RSI usually lie in multiple tensor subspaces. The previous numerical experiments demonstrate that the proposed TenMSR can finely excavate the intrinsic tensor multi-subspace structure of RSI. Meanwhile, the learned representation tensor admits the block-diagonal structure, and then, we can identify the number of subspaces in which the spectral bands of RSI are distributed.

Here, we discuss the effects of numbers of subspaces on the restoration performance. We take the PaC dataset corrupted by Gaussian noise and salt and pepper noise with $\sigma = 0.10$ and $\nu = 0.10$ as the example. From the block structure shown in Fig. 10, one can get that the spectral bands of PaC data lie in three tensor subspaces. Fig. 16 shows the recovery images and lists the recovered MPSNR and MSSIM values. One can see that the proposed method with different numbers of subspaces can remove the mixed noise. However, when the number equals three, the proposed TenMSR performs best

TABLE VI

RECOVERY PERFORMANCE OF THE PROPOSED METHOD WITH DIFFERENT NONLINEAR FUNCTIONS

Dataset	Index	Noisy	TenMSR/TNN	TenMSR/NTNN	TenMSR/3DTNN	TenMSR
Sentinel2A	MPSNR	13.59	28.60	28.75	30.62	**30.73**
	MSSIM	0.09	0.7253	0.6642	**0.7627**	0.7582
Pavia	MPSNR	13.64	30.38	30.60	33.89	**33.98**
	MSSIM	0.19	0.9079	0.9065	0.9428	**0.9437**

in keeping sharp edges and details. The values listed on the images show that the proposed TenMSR achieves the highest MPSNR and MSSIM values when the number of subspaces is three. Therefore, the proposed TenMSR can finely excavate the intrinsic tensor multi-subspace structure of RSI, achieving satisfactory performance.

B. Effects of the Nonlinear Transform

To show the effect of the nonlinear transform, Fig. 17 presents the comparisons of the singular value curves and the accumulation energy ratio (AccEgy) of singular values under the linear transform (FFT) and the nonlinear transform (Tanh) of the multi-subspace representation coefficient. We observe that the coefficient tensor under the nonlinear transform exhibits more significantly low-rank property than those obtained by the linear transform, i.e., N-3DTNN under nonlinear transform can more accurately characterize the low rankness of the data. Besides, the results shown in Table V demonstrate the superior performance of the nonlinear transform over the linear transform. This shows that the nonlinear transform is a key factor in mining the low rankness of the representation tensor.

To test the effect of different nonlinear transforms on denoising performance, taking Sentinel2A and PaC datasets corrupted by Gaussian noise and salt and pepper noise with $\sigma = 0.10$ and $\nu = 0.10$ as examples, we estimate different nonlinear functions, including LeakyReLU, ReLU, Softmax, Softsign, and Tanh. Table VI showsthe recovery performance of the proposed method with different nonlinear functions. One can see that the proposed method can achieve the promising performance for different nonlinear transforms. Without loss of generality, here we choose Tanh in our method.

C. Effects of Regularization of TenMSR Coefficient

In this work, we propose the N-3DTNN to fully explore the tensor low rankness of the multiple tensor subspace representation coefficient. To test the superiority of the

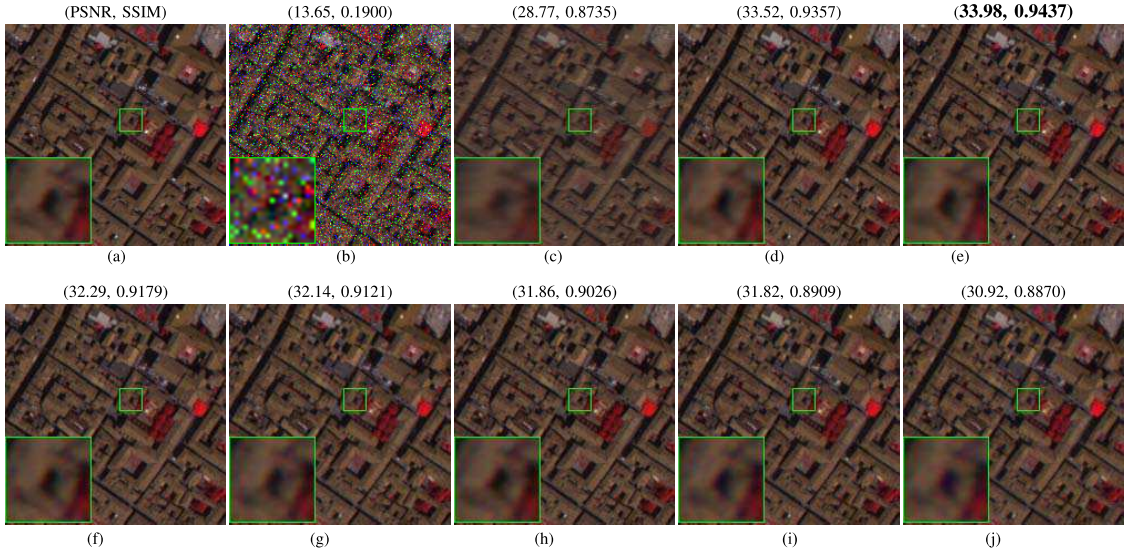

Fig. 16. Pseudo-color recovered results (R-G-B: 70-29-10 bands) of the proposed method on PaC dataset with different numbers of subspaces. (a) Original. (b) Noisy. (c) One-subspace. (d) Two-subspaces. (e) TenMSR (three-subspaces). (f) Four-subspaces. (g) Five-subspaces. (h) Six-subspaces. (i) Seven-subspaces. (j) Eight-subspaces.

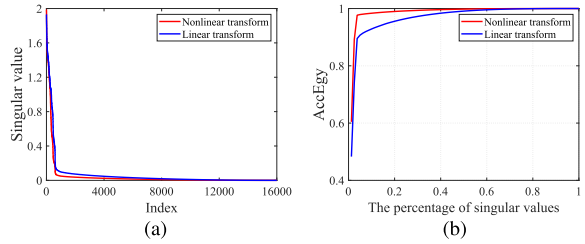

Fig. 17. Comparison of singular value curves of the representation tensor after the linear and nonlinear transform on the PaC dataset. (a) Singular values. (b) AccEgy.

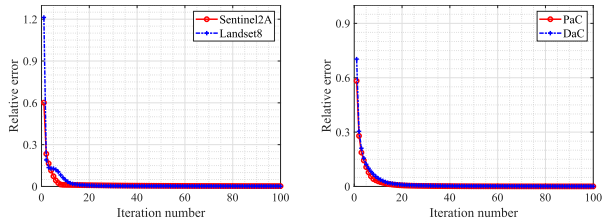

Fig. 18. Relative error curves of the proposed method on different datasets.

proposed N-3DTNN, we impose different low-rank regularizations on the tensor representation coefficient, including TNN, nonlinear TNN (NTNN), and 3DTNN, and the corresponding methods are named as TenMSR/TNN, TenMSR/NTNN, and TenMSR/3DTNN, respectively. Note that NTNN combines the nonlinear transform and TNN. Taking Sentinel2A and PaC datasets as examples, Table V presents the denoising performance of different settings and TenMSR. One can observe that the proposed TenMSR achieves the best results and TenMSR/3DTNN method obtains the second-best results. This demonstrates that considering the low rankness of the data along three dimensions is better than the traditional tensor nuclear norm that only considers the low rankness of the data in the third dimension. TenMSR/NTNN obtains higher MPSNR and MSSIM values than TenMSR/TNN, which indicates that the low rankness of the data is more obvious under the nonlinear transformation. As a comparison, the proposed TenMSR achieves the promising results in most cases, which reflects the effectiveness of N-3DTNN and the accuracy of our TenMSR in characterizing the multiple tensor subspace structure of RSIs.

D. Convergence Analysis

Theorem 1 theoretically proves the convergence of the proposed algorithm. Here, we show the numerical convergence performance of Algorithm 1. Taking Sentinel2A, Landset8, PaC, and DaC datasets under Case 2 as examples, Fig. 18 plots the relative error curves of the proposed TenMSR on different datasets. We can observe that as the number of iterations increases, the relative error has been decreasing and tends to 0—that shows the numerical convergence of the proposed algorithm.

VII. Conclusion

In this work, we propose a TenMSR method for RSI mixed noise removal. Different from the single matrix/tensor subspace methods, TenMSR can accurately describe the tensor multi-subspace structure embedded in RSI. Moreover, an N-3DTNN is used to fully characterize the tensor low rankness of the representation coefficient. We develop a PAM-based algorithm to solve the proposed model and prove that the sequences produced by the proposed algorithm can converge to a critical point. Numerical experiments on various datasets and mixed noise scenarios show the outstanding denoising performance over the existing original space and single subspace methods. In the future, we will attempt to apply the proposed method to solve other tasks, such as RSI cloud removal task and the completion task.

References

[1] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.

[2] H. Zhang, L. Liu, W. He, and L. Zhang, "Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3071–3084, May 2020.

[3] Y. Chen, M. Chen, W. He, J. Zeng, M. Huang, and Y.-B. Zheng, "Thick cloud removal in multitemporal remote sensing images via low-rank regularized self-supervised network," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5506613.

[4] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, Feb. 2019.

[5] F. Yuan, Y. Chen, W. He, and J. Zeng, "Feature fusion-guided network with sparse prior constraints for unsupervised hyperspectral image quality improvement," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5511912.

[6] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019.

[7] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "HSI-DeNet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, Feb. 2019.

[8] L. Zhuang, M. K. Ng, L. Gao, J. Michalski, and Z. Wang, "Eigenimage2Eigenimage (E2E): A self-supervised deep learning network for hyperspectral image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 11, pp. 16262–16276, Nov. 2024.

[9] H. Zhu, L. Jiao, W. Ma, F. Liu, and W. Zhao, "A novel neural network for remote sensing image matching," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2853–2865, Sep. 2019.

[10] H. Zhang, M. Lin, G. Yang, and L. Zhang, "ESCNet: An end-to-end superpixel-enhanced change detection network for very-high-resolution remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 1, pp. 28–42, Jan. 2023.

[11] L. Zhuang, X. Fu, M. K. Ng, and J. M. Bioucas-Dias, "Hyperspectral image denoising based on global and nonlocal low-rank factorizations," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10438–10454, Dec. 2021.

[12] Y. Bu, Y. Zhao, J. Xue, J. Yao, and J. C.-W. Chan, "Transferable multiple subspace learning for hyperspectral image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.

[13] M. Ding, X. Fu, T.-Z. Huang, J. Wang, and X.-L. Zhao, "Hyperspectral super-resolution via interpretable block-term tensor modeling," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 641–656, Apr. 2021.

[14] L. Zhuang and M. K. Ng, "FastHyMix: Fast and parameter-free hyperspectral image mixed noise removal," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4702–4716, Aug. 2023.

[15] H. Ji, C. Liu, Z. Shen, and Y. Xu, "Robust video denoising using low rank matrix completion," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1791–1798.

[16] H. Ji, S. Huang, Z. Shen, and Y. Xu, "Robust video restoration by joint sparse and low rank matrix approximation," *SIAM J. Imag. Sci.*, vol. 4, no. 4, pp. 1122–1142, Jan. 2011.

[17] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.

[18] Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, and W. Zhang, "Hyperspectral image restoration via iteratively regularized weighted Schatten p-norm minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4642–4659, Aug. 2016.

[19] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 700–711, Feb. 2013.

[20] X. Cao, Q. Zhao, D. Meng, Y. Chen, and Z. Xu, "Robust low-rank matrix factorization under general mixture noise distributions," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4677–4690, Oct. 2016.

[21] H. Fan, J. Li, Q. Yuan, X. Liu, and M. Ng, "Hyperspectral image denoising with bilinear low rank matrix factorization," *Signal Process.*, vol. 163, pp. 132–152, Oct. 2019.

[22] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

[23] Y. Chen, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, "Hyperspectral image restoration using framelet-regularized low-rank nonnegative matrix factorization," *Appl. Math. Model.*, vol. 63, pp. 128–147, Nov. 2018.

[24] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[25] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.

[26] X. Guo, X. Huang, L. Zhang, and L. Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, pp. 50–63, Sep. 2013.

[27] W. He, N. Yokoya, L. Yuan, and Q. Zhao, "Remote sensing image reconstruction using tensor ring completion and total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8998–9009, Nov. 2019.

[28] Y. Luo, X. Zhao, Z. Li, M. K. Ng, and D. Meng, "Low-rank tensor function representation for multi-dimensional data recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 5, pp. 3351–3369, May 2024.

[29] M. Ding, X.-L. Zhao, J.-H. Yang, Z. Zhou, and M. K. Ng, "Bilateral tensor low-rank representation for insufficient observed samples in multidimensional image clustering and recovery," *SIAM J. Imag. Sci.*, vol. 18, no. 1, pp. 20–59, Mar. 2025.

[30] B.-Z. Li, X.-L. Zhao, X. Zhang, T.-Y. Ji, X. Chen, and M. K. Ng, "A learnable group-tube transform induced tensor nuclear norm and its application for tensor completion," *SIAM J. Imag. Sci.*, vol. 16, no. 3, pp. 1370–1397, Sep. 2023.

[31] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.

[32] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[33] M. Ding, J.-H. Yang, X.-L. Zhao, J. Zhang, and M. K. Ng, "Nonconvex low-rank tensor representation for multi-view subspace clustering with insufficient observed samples," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 6, pp. 3583–3597, Jun. 2025.

[34] H. Zhang et al., "Learnable transform-assisted tensor decomposition for spatio-irregular multidimensional data recovery," *ACM Trans. Knowl. Discovery Data*, vol. 19, no. 1, pp. 1–23, Jan. 2025.

[35] Q. Zhang, Y. Zheng, Y. Dong, C. Yu, and Q. Yuan, "Hyperspectral image mixed noised removal via jointly spatial and spectral difference constraint with low-rank tensor factorization," *Eng. Appl. Artif. Intell.*, vol. 149, Jun. 2025, Art. no. 110508.

[36] T.-X. Jiang, M. K. Ng, J. Pan, and G.-J. Song, "Nonnegative low rank tensor approximations with multidimensional image applications," *Numerische Math.*, vol. 153, no. 1, pp. 141–170, Jan. 2023.

[37] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4589–4604, Oct. 2017.

[38] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-H. Ma, and T.-Y. Ji, "Mixed noise removal in hyperspectral image via low-fibered-rank regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 734–749, Jan. 2020.

[39] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, Apr. 2020.

[40] J.-M. Wu, S.-B. Yin, T.-X. Jiang, G.-S. Liu, and X.-L. Zhao, "PALADIN: A novel plug-and-play 3D CS-MRI reconstruction method," *Inverse Problems*, vol. 41, no. 3, Mar. 2025, Art. no. 035014.

[41] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.

[42] C. He, Y. Wei, K. Guo, and H. Han, "Removal of mixed noise in hyperspectral images based on subspace representation and nonlocal low-rank tensor decomposition," *Sensors*, vol. 24, no. 2, p. 327, Jan. 2024.

[43] J. Lin, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, and L. Zhuang, "A tensor subspace representation-based method for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7739–7757, Sep. 2021.

[44] L. Sun and C. He, "Hyperspectral image mixed denoising using difference continuity-regularized nonlocal tensor subspace low-rank learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[45] X. Fu, Y. Guo, M. Xu, and S. Jia, "Hyperspectral image denoising via robust subspace estimation and group sparsity constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5512716.

[46] L. Sun and B. Jeon, "A novel subspace spatial–spectral low rank learning method for hyperspectral denoising," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

[47] W. He et al., "Non-local meets global: An iterative paradigm for hyperspectral image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 2089–2107, Apr. 2022.
[48] C. He, L. Sun, W. Huang, J. Zhang, Y. Zheng, and B. Jeon, "TSLRLN: Tensor subspace low-rank learning with non-local prior for hyperspectral image mixed denoising," *Signal Process.*, vol. 184, Jul. 2021, Art. no. 108060.
[49] Y. Chen, W. Lai, W. He, X.-L. Zhao, and J. Zeng, "Hyperspectral compressive snapshot reconstruction via coupled low-rank subspace representation and self-supervised deep network," *IEEE Trans. Image Process.*, vol. 33, pp. 926–941, 2024.
[50] Q. Zhang, Y. Zheng, Q. Yuan, M. Song, H. Yu, and Y. Xiao, "Hyperspectral image denoising: From model-driven, data-driven, to model-data-driven," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 10, pp. 13143–13163, Oct. 2024.
[51] G. Fu, F. Xiong, J. Lu, and J. Zhou, "SSUMamba: Spatial–spectral selective state space model for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5527714.
[52] L. Zhuang, M. K. Ng, L. Gao, and Z. Wang, "Eigen-CNN: Eigenimages plus eigennoise level maps guided network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5512018.
[53] M. Li, Y. Fu, and Y. Zhang, "Spatial–spectral transformer for hyperspectral image denoising," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, 2023, pp. 1368–1376.
[54] K. Wei, Y. Fu, and H. Huang, "3-D quasi-recurrent neural network for hyperspectral image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 363–375, Jan. 2021.
[55] M. Li, Y. Fu, T. Zhang, and G. Wen, "Supervise-assisted self-supervised deep-learning method for hyperspectral image restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 4, pp. 7331–7344, Apr. 2025.
[56] E. Pan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Progressive hyperspectral image destriping with an adaptive frequencial focus," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5517312.
[57] X. Li, M. Ding, Y. Gu, and A. Pižurica, "An end-to-end framework for joint denoising and classification of hyperspectral images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 7, pp. 3269–3283, Jul. 2023.
[58] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6861–6870.
[59] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.
[60] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, Jan. 2013.
[61] B.-Z. Li, X.-L. Zhao, T.-Y. Ji, X.-J. Zhang, and T.-Z. Huang, "Nonlinear transform induced tensor nuclear norm for tensor completion," *J. Sci. Comput.*, vol. 92, no. 3, pp. 1–30, Sep. 2022.
[62] M. Nikolova and M. K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Sci. Comput.*, vol. 27, no. 3, pp. 937–966, Jan. 2005.
[63] X.-L. Zhao, J.-H. Yang, T.-H. Ma, T.-X. Jiang, M. K. Ng, and T.-Z. Huang, "Tensor completion via complementary global, local, and nonlocal priors," *IEEE Trans. Image Process.*, vol. 31, pp. 984–999, 2022.
[64] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods," *Math. Program.*, vol. 137, nos. 1–2, pp. 91–129, Feb. 2013.
[65] J.-H. Yang, C. Chen, H.-N. Dai, M. Ding, Z.-B. Wu, and Z. Zheng, "Robust corrupted data recovery and clustering via generalized transformed tensor low-rank representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 7, pp. 8839–8853, Jul. 2024.
[66] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.
[67] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1888–1902, Aug. 2018.
[68] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising based on low rank and sparse representations," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 1847–1850.
[69] F. Xiong, J. Zhou, and Y. Qian, "Hyperspectral restoration via $L\_\{0\}$ gradient regularized low-rank tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10410–10425, Dec. 2019.
[70] Q. Zhang et al., "Three-dimension spatial–spectral attention transformer for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5531213.
[71] S. Liu, X. L. Zhao, J. S. Leng, B. Z. Li, J. H. Yang, and X. Y. Chen, "Revisiting high-order tensor singular value decomposition from basic element perspective," *IEEE Trans. Signal Process.*, vol. 72, pp. 4589–4603, 2024.

Jing-Hua Yang received the Ph.D. degree from Macau University of Science and Technology in 2023.

She is currently an Assistant Professor with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China. Her current research interests include data mining, image processing, tensor analysis, and federated learning. More information can be found on her homepage https://jinghua-yang.github.io/

Heng-Chao Li (Senior Member, IEEE) received the B.S. and M.S. degrees from Southwest Jiaotong University, Chengdu, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008.

From 2013 to 2014, he was a Visiting Scholar working with Prof. William J. Emery at the University of Colorado at Boulder, Boulder, CO, USA. He is currently a Professor with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu. His research interests include tensor analysis, remote sensing image processing, and pattern recognition.

Meng Ding received the Ph.D. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2021.

From 2019 to 2020, he was an exchange Ph.D. Student with Oregon State University, Corvallis, OR, USA, supported by China Scholarship Council. He is currently an Associate Professor with the School of Mathematics, Southwest Jiaotong University, Chengdu. His current research interests include image processing, tensor analysis, and machine learning.

Xi-Le Zhao (Senior Member, IEEE) received the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2009 and 2012, respectively.

He worked as a Post-Doctoral Researcher working with Prof. Michael K. Ng at Hong Kong Baptist University, Hong Kong, from 2013 to 2014. He worked as a Visiting Scholar working with Prof. Jose Bioucas Dias at the University of Lisbon, Lisbon, Portugal, from 2016 to 2017. He is currently a Professor with the School of Mathematical Sciences, UESTC. His research interests include scientific computing, image processing, and machine learning. More information can be found on his homepage https://zhaoxile.github.io/

Wen-Yu Hu received the Ph.D. degree in computational mathematics from Dalian University of Technology, Dalian, China, in 2012.

From March 2016 to March 2017, he was a Visiting Scholar with the National Centre for Computer Animation (NCCA), Bournemouth University, U.K. He is currently a Professor with the School of Mathematics and Computer Science, Gannan Normal University, Ganzhou, China. His research interests include tensor learning and subspace clustering.