

Double Factor-Regularized Low-Rank Tensor Factorization for Mixed Noise Removal in Hyperspectral Image

Yu-Bang Zheng, Ting-Zhu Huang, Xi-Le Zhao, Yong Chen, and Wei He

Abstract—As a preprocessing step, hyperspectral image (HSI) restoration plays a critical role in many subsequent applications. Recently, based on the framework of subspace representation and low-rank matrix/tensor factorization (LRMF/LRTF), many single factor-regularized methods add various regularization on the spatial factor to characterize its spatial prior knowledge. However, these methods neglect the common characteristics among different bands and the spectral continuity of HSIs. To tackle this issue, this article establishes a bridge between the factor-based regularization and the HSI priors and proposes a double factor-regularized LRTF model for HSI mixed noise removal. The proposed model employs LRTF to characterize the spectral global low-rankness, introduces a weighted group sparsity constraint on the spatial difference images (SpatDIs) of the spatial factor to promote the group sparsity in the SpatDIs of HSIs, and suggests a continuity constraint on the spectral factor to promote the spectral continuity of HSIs. Moreover, we develop a proximal alternating minimization-based algorithm to solve the proposed model. Extensive experiments conducted on simulated and real HSIs demonstrate that the proposed method has superior performance on mixed noise removal, as compared with the state-of-the-art methods based on subspace representation, noise modeling, and LRMF/LRTF.

Index Terms—Factor-based regularization (FR), hyperspectral image (HSI), low-rank tensor factorization (LRTF), mixed noise removal, proximal alternating minimization (PAM).

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) reflect different imaging effects of one real scene under hundreds of contiguous spectral bands. Compared with grayscale images, HSIs contain a wealth of spatial and spectral information, making them widely applied in various applications [1], such as natural disaster monitoring, terrain detection, and land use analysis. Unfortunately, real-world HSIs always suffered from all sorts of noise, including Gaussian noise, salt and pepper noise, deadlines, and stripes. The existence of above noise greatly reduces the quality of HSIs, resulting in the infeasibility of

Manuscript received January 13, 2020; revised March 17, 2020; accepted April 10, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61772003 and Grant 61876203 and in part by the Japan Society for the Promotion of Science under Grant KAKENHI 19K20308. (Corresponding authors: Ting-Zhu Huang; Xi-Le Zhao.)

Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, and Y. Chen are with the Research Center for Image and Vision Computing, School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zhengyubang@163.com; tingzhuhuang@126.com; xlzhao122003@163.com; chenyong1872008@163.com).

W. He is with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, RIKEN, Tokyo 103-0027, Japan (e-mail: wei.he@riken.jp).

subsequent applications, such as classification [2]–[4], unmixing [5]–[7], fusion [8], and target detection [9]. Hence, restoring the clean HSI from a noisy observation, i.e., HSI restoration, plays a critical role in HSI applications. Deeply exploring and accurately describing the spatial and spectral knowledge in HSIs is the core issue of HSI restoration. Among the currently HSI restoration methods, low-rankness, local continuity, and nonlocal self-similarity are the most frequently researched HSI priors [10]–[23].

Since each band of HSIs can be regarded as a grayscale image, the simplest and most direct way for HSI restoration is to employ the traditional grayscale image restoration methods, such as NLM [24], K-SVD [25], BM3D [26], and WNNM [27] to restore the target HSI band-by-band. However, the most shortcoming of this kind of method is the neglect of the spectral correlations, which is crucial prior knowledge for improving the performance of HSI restoration. Consequently, a variety of efficient methods have been introduced to simultaneously consider spatial and spectral prior, leading to a great improvement in restoring results. For example, to consider spatial information and spectral noise differences simultaneously, Yuan *et al.* [28] introduced a spectral-spatial adaptive total variation (TV) to HSI restoration and obtained a promising result. By supposing that the first several components output by principal component analysis (PCA) contain the major information of images, Chen and Qian [29] proposed an HSI restoration method, which first performs PCA on the noisy HSI, then employs the bivariate wavelet thresholding and a wavelet transform algorithm to remove the noise in the last several components, and finally obtains the restored HSI via inverse PCA. Besides, to exploit nonlocal self-similarity in both spatial and spectral domains, many restoration methods regarded the group of similar 3-D cubes as the basic unit for restoring. Examples include nonlocal group-wise spectrum-PCA [30], nonlocal tensor dictionary learning [31], and nonlocal tensor ring decomposition [32]. To utilize the spectral global low-rankness and spatial nonlocal self-similarity, under the framework of subspace representation, Zhuang and Bioucas-Dias [33] employed the BM3D to eigen images denoising and He *et al.* [34] embedded the WNNM to eigen images denoising while updating the subspace iteratively. However, the aforementioned methods only focus on removing one type of noise, especially Gaussian noise. But in most realistic scenarios, the noise in HSIs usually manifests as a mixture of several kinds of noise.

Recently, an increasing number of researchers have been

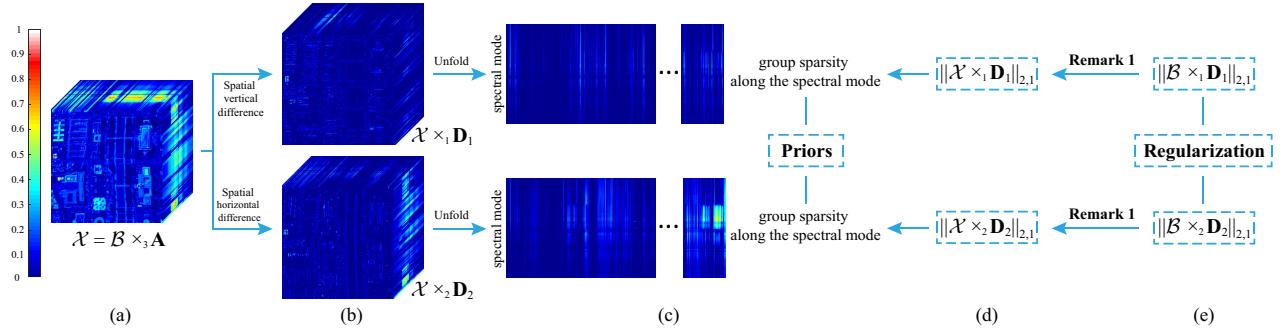


Fig. 1. The group sparsity along the spectral mode of the spatial difference images and the proposed regularization terms based on the spatial factor \mathcal{B} . (a) The HSI Washington DC Mall $\mathcal{X} \in \mathbb{R}^{256 \times 256 \times 170}$, which can be factorized as $\mathcal{B} \times_3 \mathbf{A}$, where $\mathcal{B} \in \mathbb{R}^{256 \times 256 \times r}$, $\mathbf{A} \in \mathbb{R}^{170 \times r}$, and $r \ll 170$. (b) The images obtained by performing spatial difference operators, where \mathbf{D}_1 and \mathbf{D}_2 are the first-order difference matrix. (c) The mode-3 unfolding matrices of the corresponding difference images. (d) The regularization terms on original HSI \mathcal{X} used to promote the group sparsity. (e) The regularization terms on the factor \mathcal{B} used to replace the regularization terms on original HSI \mathcal{X} .

putting great effort into mixed noise removal to adapt to real noise situations. Based on the low-rankness prior hypothesis, Zhang *et al.* [35] made the first attempt to employ low-rank matrix recovery (LRMR)-based framework to remove the mixed noise. Specifically, this method first separates the noisy HSI as overlapped full-band 3-D cubes, then unfolds 3-D cubes to matrices along the spectral mode, and finally uses the LRMR-based framework to restore unfolding matrices. Consequently, to enhance the performance of LRMR, a series of low-rank matrix approximation (LRMA)-based methods introduced nonconvex functions to approximate the rank of the unfolded HSIs [36], [37]. However, LRMA-based methods involve the singular value decomposition (SVD), leading to high computational complexity. To tackle this issue, low-rank matrix factorization (LRMF) is used to the HSI mixed noise removal problem with several flexible noise distributions hypothesis, such as a mixture of independent and identically distributed (i.i.d.) exponential power distributions [38] and a mixture of non-i.i.d. Gaussian distributions [39]. By regarding HSIs as third-order tensors, a large number of methods used the tensor decompositions/tensor rank minimization, such as Tucker decomposition [40], [41], PARAFAC decomposition [42], and tensor-SVD [43], [44], to exploit the global low-rankness of HSIs. Among them, Fan *et al.* [43] assumed that the clean HSI is a low-tubal-rank part and introduced an HSI mixed noise removal model by minimizing the convex relaxation of the tubal rank. Zheng *et al.* [44] subsequently proposed a new tensor rank, i.e., tensor fibered rank, to more flexibly and accurately explore the low-rankness of HSIs, achieving a great improvement in mixed noise removal. However, most of the aforementioned methods only considered the low-rankness and thus cannot to exhaust the intrinsic potential of HSIs, leading to a room for further enhancement.

To further boost the denoising performance, many researchers proposed a variety of HSI mixed noise removal models based on the framework of the convolutional neural network (CNN) [45]–[48], which has the powerful nonlinear fitting and feature expression ability. For example, Chang *et al.* [45] proposed a CNN-based method by modeling the spatial and the spectral features by the dilated convolution and multichannel filters, respectively. Zhang *et al.* [46] introduced

a method based on the spatial-spectral gradient network, which is capable of extracting the directional feature of sparse noise and exploring abundant spectral information of HSIs. Meanwhile, many methods considered not only the low-rankness prior but also the local continuity and nonlocal self-similarity priors. For example, He *et al.* [49] proposed a method, which employed LRMF to explore the spectral low-rankness while applying the spatial TV to each band of HSIs to exploit the spatial local continuity. To further take the spectral local continuity into consideration, the spatial-spectral TV has been proposed and integrated into various low-rank-based models, such as weighted nuclear norm [50] and low-rank Tucker decomposition (LRTD) [51], [52]. Besides, Chen *et al.* [53] conducted a group sparsity regularization on spatial difference images under the framework of LRTD, obtaining a promising result. However, the regularization terms in the aforementioned methods were added to HSIs, resulting in high computational complexity.

To reduce the computational complexity, several single factor-regularized methods have been proposed based on the framework of LRMF/low-rank tensor factorization (LRTF) [54]–[56]. Specifically, an HSI $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ can be factorized as $\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}$, where $\mathbf{A} \in \mathbb{R}^{n_3 \times r}$ and $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times r}$ ($r \ll n_3$) denote the spectral factor and the spatial factor, respectively. And the existing methods usually regarded each frontal slice of the spatial factor \mathcal{B} as a grayscale image, and subsequently used various regularization to explore its prior. Examples include the TV/framelet-regularization used to explore the spatial local continuity [54], [55] and the nonlocal low-rank regularization used to exploit the spatial nonlocal self-similarity [56]. But there are still exist two issues. On the one hand, these methods cannot fully explore the common characteristics, except the global low-rankness, of different bands. For example, most of the smooth areas in different bands of an HSI usually located at the same location since different bands present the same scene. This common characteristic causes that the spatial difference images (SpatDI) of the HSIs, shown in Fig. 1(b) and (c), are group sparse along the spectral mode¹. On the other hand, these methods neglect

¹The number of spectral vectors whose elements are approximately all zero is much more than that of the nonzero ones.

the local continuity along the spectral mode of HSIs.

In this article, we propose a double factor-regularized LRTF (LRTF-DFR) method for HSI mixed noise removal, which not only employs the LRTF framework $\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}$ to characterize the spectral global low-rankness of HSIs but also utilizes the regularization terms on the spatial factor \mathcal{B} and the spectral factor \mathbf{A} to depict the aforementioned group sparsity and spectral continuity, respectively. The contributions of this article are mainly the following three aspects.

First, a direct way to explore aforementioned group sparsity is to utilize the $\ell_{2,1}$ -norm on the SpatDIs of HSIs, i.e., $\|\mathcal{X} \times_k \mathbf{D}_k\|_{2,1}$ ($k = 1, 2$) shown in Fig. 1(d). Instead of this direct way, this article unitizes the $\ell_{2,1}$ -norm on the SpatDIs of the spatial factor \mathcal{B} , i.e., $\|\mathcal{B} \times_k \mathbf{D}_k\|_{2,1}$ shown in Fig. 1(e), and proves that $\mathcal{X} \times_k \mathbf{D}_k$ must be group sparse along the spectral mode when $\mathcal{B} \times_k \mathbf{D}_k$ is group sparse along the third mode. Meanwhile, a weighted strategy is employed to better promote the group sparsity.

Second, we note that each spectral vector (tube) of an HSI \mathcal{X} can be mathematically expressed as the linear combination of all column of the spectral factor \mathbf{A} , i.e., columns of \mathbf{A} can be regarded as a basis of the spectral space of \mathcal{X} . Since continuous bases tend to generate continuous data, this article promotes the continuity of spectral vectors of \mathcal{X} by boosting the continuity of columns of \mathbf{A} .

Third, an efficient proximal alternating minimization [57] (PAM)-based algorithm is developed to solve the proposed LRTF-DFR method. Extensive experiments conducted on simulated and real HSIs exhibit the superior performance of the proposed LRTF-DFR on mixed noise removal, spatial image recovery, and spectral signatures preserving, as compared with several excellent methods.

The rest of this article is organized as follows. Section II gives some notations and preliminary knowledge. Section III presents the proposed LRTF-DFR-based HSI restoration model and develops a PAM-based solving algorithm with a detailed computational complexity analysis. Section IV carries out extensive experiments to illustrate the superiority of the proposed model. Section V summarizes this article.

II. NOTATIONS AND PRELIMINARIES

Following the nomenclatures of [58], we summarize notations used in this article in Table I. Next, we introduce the definitions of mode- k unfolding, mode- k tensor-matrix product, and Tucker rank.

Definition 1 (Mode- k Unfolding): For an N th-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$, its mode- k unfolding $\mathbf{X}_{(k)}$ is an $n_k \times \prod_{i \neq k} n_i$ matrix, which satisfies that $\mathbf{X}_{(k)}(i_k, j)$ is mapped by $\mathcal{X}(i_1, i_2, \dots, i_N)$, where $j = 1 + \sum_{s=1, s \neq k}^N (i_s - 1) J_s$ with $J_s = \sum_{m=1, m \neq k}^{s-1} n_m$. The corresponding operator and inverse operator are denoted as $\mathbf{X}_{(k)} = \text{Unfold}_k(\mathcal{X})$ and $\mathcal{X} = \text{Fold}_k(\mathbf{X}_{(k)})$, respectively.

Definition 2 (Mode- k Tensor-Matrix Product): The mode- k tensor-matrix product of an $n_1 \times n_2 \times \cdots \times n_N$ tensor \mathcal{X} and a $J \times n_k$ matrix \mathbf{A} is an $n_1 \times \cdots \times n_{k-1} \times J \times n_{k+1} \times \cdots \times n_N$

TABLE I
NOTATION DECLARATIONS.

Notations	Interpretations
$x, \mathbf{x}, \mathbf{X}, \mathcal{X}$	scalar, vector, matrix, tensor
x_{i_1, i_2, \dots, i_N} or $\mathcal{X}(i_1, i_2, \dots, i_N)$	the (i_1, i_2, \dots, i_N) th element of \mathcal{X}
$\mathbf{x}_{:, i_2, i_3}$ or $\mathcal{X}(:, i_2, i_3)$	the (i_2, i_3) th column of a third-order tensor \mathcal{X}
$\mathbf{x}_{i_1, :, i_3}$ or $\mathcal{X}(i_1, :, i_3)$	the (i_1, i_3) th row of a third-order tensor \mathcal{X}
$\mathbf{x}_{i_1, i_2, :}$ or $\mathcal{X}(i_1, i_2, :)$	the (i_1, i_2) th tubal of a third-order tensor \mathcal{X}
$\ \mathcal{X}\ _F$	Frobenius norm $\ \mathcal{X}\ _F = \sqrt{\sum_{i_1, i_2, \dots, i_N} x_{i_1, i_2, \dots, i_N} ^2}$
$\ \mathcal{X}\ _1$	ℓ_1 -norm $\ \mathcal{X}\ _1 = \sum_{i_1, i_2, \dots, i_N} x_{i_1, i_2, \dots, i_N} $
$\ \mathcal{X}\ _{2,1}$	$\ell_{2,1}$ -norm of third-order tensors $\ \mathcal{X}\ _{2,1} = \sum_{i_1, i_2} x_{i_1, i_2, :} _2$
$\mathbf{X}_{(k)}$	the mode- k unfolding of \mathcal{X}
$\langle \cdot, \cdot \rangle$	inner product
\otimes	Kronecker product
\odot	component-wise multiplication
\oslash	component-wise division
\times_k	mode- k tensor-matrix product

tensor denoted by $\mathcal{X} \times_k \mathbf{A}$ and satisfied

$$(\mathcal{X} \times_k \mathbf{A})_{i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_N} = \sum_{i_k=1}^{n_k} x_{i_1, i_2, \dots, i_N} \cdot a_{j, i_k}.$$

According to above two definitions, we have

$$\mathcal{Y} = \mathcal{X} \times_k \mathbf{A} \Leftrightarrow \mathbf{Y}_{(k)} = \mathbf{A} \mathbf{X}_{(k)},$$

and

$$\begin{aligned} \mathcal{Y} &= \mathcal{X} \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \cdots \times_N \mathbf{A}_N \Leftrightarrow \\ \mathbf{Y}_{(k)} &= \mathbf{A}_k \mathbf{X}_{(k)} (\mathbf{A}_N \otimes \cdots \otimes \mathbf{A}_{k+1} \otimes \mathbf{A}_{k-1} \otimes \cdots \otimes \mathbf{A}_1)^T. \end{aligned}$$

Definition 3 (Tucker Rank): For an N th-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$, its Tucker rank is defined as

$$\text{Rank}_T(\mathcal{X}) = (\text{Rank}(\mathbf{X}_{(1)}), \text{Rank}(\mathbf{X}_{(2)}), \dots, \text{Rank}(\mathbf{X}_{(N)})).$$

Interested readers can obtain a more detailed introduction for tensors in [58].

III. PROPOSED HSI RESTORATION MODEL AND SOLVING ALGORITHM

A. Problem Formulation and Motivation

By simply assuming the noisy HSI $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is corrupted by the additive Gaussian noise \mathcal{N} and the additive sparse noise \mathcal{S} (salt and pepper noise, deadlines, and stripes), the HSI degradation formulation can be expressed as

$$\mathcal{Y} = \mathcal{X} + \mathcal{N} + \mathcal{S}.$$

The key issue of recovering the clean HSI \mathcal{X} from \mathcal{Y} is to exactly characterize the HSI prior by establishing appropriate regularization terms. Mathematically, a generalized HSI restoration model can be formulated as

$$\min_{\mathcal{X}, \mathcal{S}} \frac{1}{2} \|\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2 + \tau \mathcal{R}(\mathcal{X}) + \mu \|\mathcal{S}\|_1, \quad (1)$$

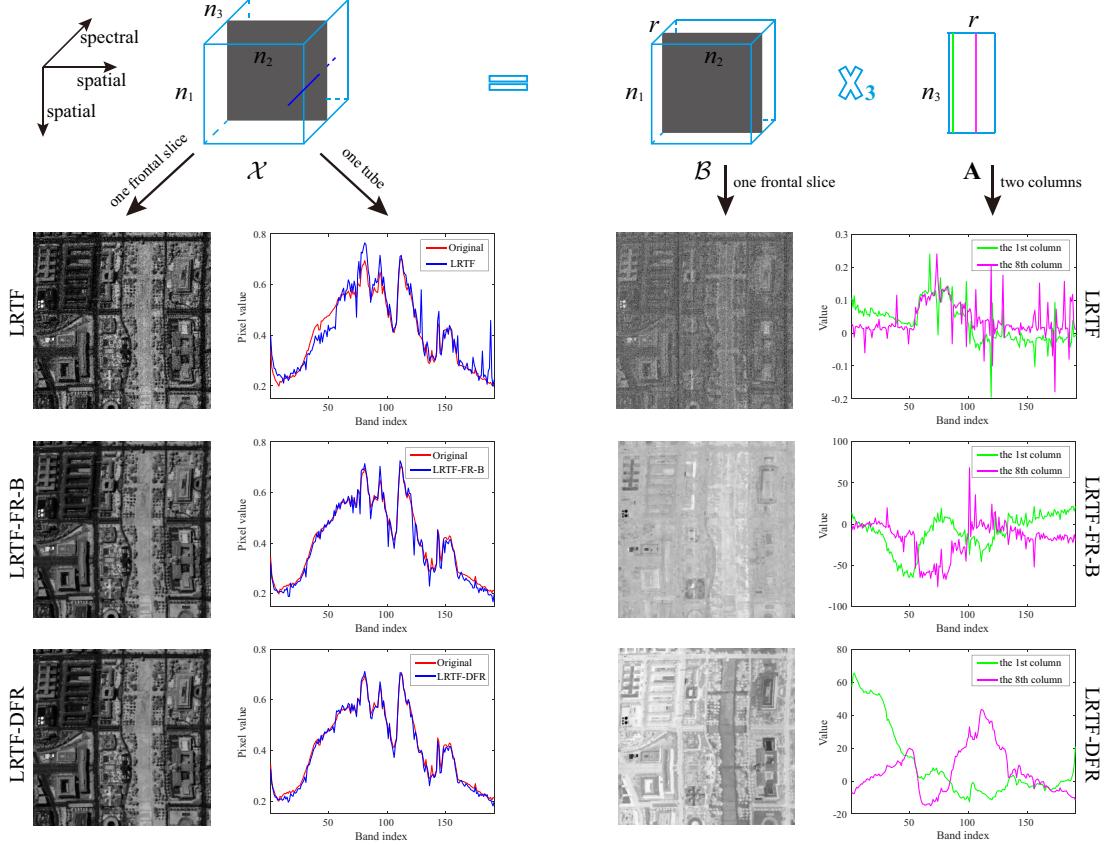


Fig. 2. The impact of each term in the proposed LRTF-DFR on the performance of HSI restoration. From the second to fourth row: the results of the method without factor-based regularization (LRTF), the method only with factor \mathcal{B} -based regularization (LRTF-FR-B), and the proposed LRTF-DFR, respectively. From left to right: restoration results at band 72, restoration results of the spectral vector at spatial location (150,150), one frontal slice of the spatial factor \mathcal{B} , and two columns of the spectral factor \mathbf{A} .

where τ and μ are regularization parameters; $\mathcal{R}(\mathcal{X})$ is the regularization term designed to exploit the HSI prior. For example, $\mathcal{R}(\mathcal{X})$ was designed as tensor nuclear norm [43], [44] and Schatten p -norm [36] to explore low-rankness prior, spatial-spectral TV [52] and enhanced 3-D TV [59] to characterize spatial-spectral continuity prior, and nonlocal low-rank regularization [51] to exploit nonlocal self-similarity prior. However, adding regularization on HSIs will inevitably lead to high computational complexity, since HSIs usually have a large scale.

Fortunately, the spectral low-rankness of HSIs can be represented as the mode-3 tensor-matrix product of a low-dimensional tensor and a low-dimensional matrix, i.e., the clean HSI \mathcal{X} can be approximately factorized as

$$\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}, \quad (2)$$

where $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times r}$ ($r \ll n_3$) denotes the spatial factor and $\mathbf{A} \in \mathbb{R}^{n_3 \times r}$ represents the spectral factor. Since the spatial factor \mathcal{B} usually reflects the spatial characteristics of HSIs [33], [34], many subspace-based and LRMF/LRTF-based methods employed various regularization on \mathcal{B} and their models can be generally formulated as

$$\min_{\mathbf{A}, \mathcal{B}, \mathcal{S}} \frac{1}{2} \|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A} - \mathcal{S}\|_F^2 + \tau \mathcal{R}(\mathcal{B}) + \mu \|\mathcal{S}\|_1, \quad (3)$$

where \mathbf{A} satisfies $\mathbf{A}^T \mathbf{A} = \mathbf{I}$ in subspace-based methods and $\mathcal{R}(\mathcal{B})$ is the regularization term designed to exploit its prior. For example, $\mathcal{R}(\mathcal{B})$ was designed as the framelet-regularization [55] and nonlocal low-rank regularization [56] to explore the spatial local continuity and the spatial nonlocal self-similarity, respectively. However, the existing regularization on \mathcal{B} cannot fully explore the common characteristics among different bands of HSIs, such as the spectral group sparsity in SpatDIs of HSIs. Meanwhile, the spectral continuity of HSIs was also neglected.

B. LRTF-DFR-Based HSI Restoration Model

To consider the aforementioned spectral group sparsity and spectral continuity in the HSI \mathcal{X} , we introduce regularization on factors in (2) and propose an LRTF-DFR-based HSI restoration method as follows:

$$\begin{aligned} \min_{\mathbf{A}, \mathcal{B}, \mathcal{S}} & \frac{1}{2} \|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A} - \mathcal{S}\|_F^2 + \tau \sum_{k=1}^2 \|\mathcal{W}_k \odot (\mathcal{B} \times_k \mathbf{D}_k)\|_{2,1} \\ & + \lambda \|\mathbf{D}_3 \mathbf{A}\|_F^2 + \mu \|\mathcal{W}_s \odot \mathcal{S}\|_1, \end{aligned} \quad (4)$$

where τ , λ , and μ are regularization parameters; \mathbf{D}_k ($k = 1, 2, 3$) are first-order difference matrices; \mathcal{W}_k ($k = 1, 2$) and \mathcal{W}_s are weight tensors used to better promote the group sparsity [53] and the sparsity [60], respectively. Especially, all

frontal slices of \mathcal{W}_k are the same matrices, denoted as \mathbf{W}_k . Most importantly, the Remarks 1 and 2 give a detail illustration of why the group sparsity constraint on $\mathcal{B} \times_k \mathbf{D}_k$ ($k = 1, 2$) can promote the group sparsity in $\mathcal{X} \times_k \mathbf{D}_k$ and the continuity constraint on \mathbf{A} can promote the spectral continuity of \mathcal{X} .

Remark 1 (Factor \mathcal{B} -Based Regularization): With the Definition 2, we have

$$\begin{aligned}\mathcal{X} \times_k \mathbf{D}_k &= (\mathcal{B} \times_3 \mathbf{A}) \times_k \mathbf{D}_k = (\mathcal{B} \times_k \mathbf{D}_k) \times_3 \mathbf{A} \\ \Leftrightarrow [\mathcal{X} \times_k \mathbf{D}_k](i, j, :) &= \sum_{t=1}^r [\mathcal{B} \times_k \mathbf{D}_k](i, j, t) \mathbf{A}(:, t).\end{aligned}$$

It can be easily to obtain that $[\mathcal{X} \times_k \mathbf{D}_k](i, j, :) = 0$ when $[\mathcal{B} \times_k \mathbf{D}_k](i, j, :) = 0$. Therefore, when $\mathcal{B} \times_k \mathbf{D}_k$ is group sparse along the third mode, $\mathcal{X} \times_k \mathbf{D}_k$ must be group sparse along the third (spectral) mode. Moreover, the above conclusion is also true in reverse. This is because $[\mathcal{B} \times_k \mathbf{D}_k](i, j, :) = 0$ when $[\mathcal{X} \times_k \mathbf{D}_k](i, j, :) = 0$, since $\mathbf{A}(:, 1)$, $\mathbf{A}(:, 2)$, \dots , and $\mathbf{A}(:, r)$ are linearly independent.

Remark 2 (Factor \mathbf{A} -Based Regularization): By rewriting $\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}$ as $\mathbf{X}_{(3)} = \mathbf{AB}_{(3)}$, it is not hard to find that each spectral vector of \mathcal{X} (column of $\mathbf{X}_{(3)}$) can be mathematically expressed as a linear combination of all column of the spectral factor \mathbf{A} . This implies that columns of \mathbf{A} can be regarded as a basis of the spectral space of \mathcal{X} . Since continuous bases usually tend to generate continuous data, boosting the continuity of the columns of \mathbf{A} can promote the continuity of the spectral vector of \mathcal{X} , as compared with the case without the continuity constraint.

We take an example shown in Fig. 2 to illustrate the impact of the factor-based regularization terms in the proposed LRTF-DFR on the performance of mixed noise removal. By comparing the second and third rows, we find that owing to the group sparsity constraint on $\mathcal{B} \times_k \mathbf{D}_k$ ($k = 1, 2$), the frontal slice of \mathcal{B} obtained by LRTF-FR-B contain more geometrical features than that obtained by LRTF, leading to a significant improvement in HSI restoration. By comparing the third and fourth rows, we find that owing to the column continuity constraint on \mathbf{A} , the spectral vector obtained by the proposed LRTF-DFR is much smoother and closer to the original ones than that by LRTF-FR-B. The above observations empirically illustrate the significance of the factor-based regularization terms and consistent with the previous theoretical discussions.

In summary, the proposed LRTF-DFR has the following two advantages.

- *Double factor-based regularization for HSI prior:* Compared with the single factor-regularized methods [54]–[56], the proposed LRTF-DFR explores more HSI prior knowledge in both spatial and spectral domains since both spatial and spectral factors are fully utilized.
- *Lower computational complexity:* Compared with the method using regularization on the original HSI [52], [53], the proposed LRTF-DFR has lower computational complexity since the scales of factors are much smaller than that of the original HSI.

C. PAM-Based Solving algorithm

Within the framework of PAM-based algorithm, the problem (4) can be solved by alternately updating

$$\left\{ \begin{array}{l} \text{Step 1: } \mathbf{A}^{l+1} = \underset{\mathbf{A}}{\operatorname{argmin}} f(\mathbf{A}, \mathcal{B}^l, \mathcal{S}^l) + \frac{\rho}{2} \|\mathbf{A} - \mathbf{A}^l\|_F^2, \\ \text{Step 2: } \mathcal{B}^{l+1} = \underset{\mathcal{B}}{\operatorname{argmin}} f(\mathbf{A}^{l+1}, \mathcal{B}, \mathcal{S}^l) + \frac{\rho}{2} \|\mathcal{B} - \mathcal{B}^l\|_F^2, \\ \text{Step 3: } \mathcal{S}^{l+1} = \underset{\mathcal{S}}{\operatorname{argmin}} f(\mathbf{A}^{l+1}, \mathcal{B}^{l+1}, \mathcal{S}) + \frac{\rho}{2} \|\mathcal{S} - \mathcal{S}^l\|_F^2, \end{array} \right. \quad (5)$$

where $f(\mathbf{A}, \mathcal{B}, \mathcal{S})$ is the objective function of (4) and $\rho > 0$ is the proximal parameter.

1) *Update \mathbf{A} :* In Step 1 of (5), the \mathbf{A} -subproblem is as follows:

$$\begin{aligned}\underset{\mathbf{A}}{\operatorname{argmin}} \quad & \frac{1}{2} \|\mathcal{Y} - \mathcal{B}^l \times_3 \mathbf{A} - \mathcal{S}^l\|_F^2 \\ & + \lambda \|\mathbf{D}_3 \mathbf{A}\|_F^2 + \frac{\rho}{2} \|\mathbf{A} - \mathbf{A}^l\|_F^2,\end{aligned}\quad (6)$$

whose solver can be directly obtained by solving the following Sylvester matrix equation:

$$\begin{aligned}\mathbf{A} \mathbf{B}_{(3)}^l (\mathbf{B}_{(3)}^l)^T + 2\lambda \mathbf{D}_3^T \mathbf{D}_3 \mathbf{A} + \rho \mathbf{A} \\ = (\mathbf{Y}_{(3)} - \mathbf{S}_{(3)}^l) (\mathbf{B}_{(3)}^l)^T + \rho \mathbf{A}^l.\end{aligned}\quad (7)$$

To efficiently solve (7), we introduce the following theorem.

Theorem 1 (A Fast Solution of Sylvester Matrix Equation): Supposing that $\mathbf{A} \in \mathbb{R}^{m \times m}$, $\mathbf{B} \in \mathbb{R}^{n \times n}$, and $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m \times n}$. The classical Sylvester matrix equation

$$\mathbf{AX} + \mathbf{XB} = \mathbf{Y}$$

has a unique solution if only if $\mathbf{I}_n \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I}_m$ is a invertible matrix. Especially, if

$$\mathbf{A} = \mathbf{U}_1 \Lambda_1 \mathbf{U}_1^T \quad \text{and} \quad \mathbf{B} = \mathbf{U}_2 \Lambda_2 \mathbf{U}_2^T,$$

this unique solution can be expressed as

$$\mathbf{X} = \mathbf{U}_1 ((1 \otimes \mathbf{T}) \odot (\mathbf{U}_1^T \mathbf{Y} \mathbf{U}_2)) \mathbf{U}_2^T,$$

where Λ_1 and Λ_2 are diagonal matrices, \mathbf{U}_1 and \mathbf{U}_2 are unitary matrices, and

$$\begin{aligned}\mathbf{T} = & (\operatorname{diag}^2(\Lambda_1), \operatorname{diag}(\Lambda_1), \dots, \operatorname{diag}(\Lambda_1)) \\ & + (\operatorname{diag}(\Lambda_2), \operatorname{diag}(\Lambda_2), \dots, \operatorname{diag}(\Lambda_2))^T.\end{aligned}$$

In (7), it is not hard to find that the matrix $\mathbf{D}_3^T \mathbf{D}_3$ is a circulant matrix and the matrix $\mathbf{B}_{(3)}^l (\mathbf{B}_{(3)}^l)^T$ is a symmetric matrix. Therefore, we utilize one-dimensional fast Fourier transformation (FFT) and singular value decomposition (SVD) to diagonalize $\mathbf{D}_3^T \mathbf{D}_3$ and $\mathbf{B}_{(3)}^l (\mathbf{B}_{(3)}^l)^T$, respectively. That is,

$$\mathbf{D}_3^T \mathbf{D}_3 = \mathbf{F}_1^T \Psi_1 \mathbf{F}_1 \quad \text{and} \quad \mathbf{B}_{(3)}^l (\mathbf{B}_{(3)}^l)^T = \mathbf{U}_1 \Sigma_1 \mathbf{U}_1^T,$$

where \mathbf{F}_1 is one-dimensional discrete Fourier transformation (DFT) matrix. By using Theorem 1, we can efficiently solve (7) as

$$\mathbf{A}^{l+1} = \mathbf{F}_1^T ((1 \otimes \mathbf{T}_1) \odot (\mathbf{F}_1 \mathbf{G} \mathbf{U}_1)) \mathbf{U}_1^T, \quad (8)$$

² $\operatorname{diag}(\Lambda)$ is a column vector whose elements are diagonal elements of Λ .

where $\mathbf{G} = (\mathbf{Y}_{(3)} - \mathbf{S}_{(3)}^l)(\mathbf{B}_{(3)}^l)^T + \rho \mathbf{A}^l$ and

$$\begin{aligned}\mathbf{T}_1 = & 2\lambda(\text{diag}(\Psi_1), \text{diag}(\Psi_1), \dots, \text{diag}(\Psi_1)) + \rho \text{ones}^3(n_3, r) \\ & + (\text{diag}(\Sigma_1), \text{diag}(\Sigma_1), \dots, \text{diag}(\Sigma_1))^T.\end{aligned}$$

2) *Update \mathcal{B}* : In Step 2 of (5), the \mathcal{B} -subproblem is as follows:

$$\begin{aligned}\underset{\mathcal{B}}{\operatorname{argmin}} \quad & \frac{1}{2} \|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A}^{l+1} - \mathcal{S}^l\|_F^2 \\ & + \tau \sum_{k=1}^2 \|\mathcal{W}_k \odot (\mathcal{B} \times_k \mathbf{D}_k)\|_{2,1} + \frac{\rho}{2} \|\mathcal{B} - \mathcal{B}^l\|_F^2.\end{aligned}\tag{9}$$

We use the alternating direction method of multipliers (ADMM) [61] to solve (9). By introducing two auxiliary variables \mathcal{Z}_1 and \mathcal{Z}_2 , we rewrite (9) as

$$\begin{aligned}\underset{\mathcal{B}, \mathcal{Z}_1, \mathcal{Z}_2}{\operatorname{argmin}} \quad & \frac{1}{2} \|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A}^{l+1} - \mathcal{S}^l\|_F^2 \\ & + \tau \sum_{k=1}^2 \|\mathcal{W}_k \odot \mathcal{Z}_k\|_{2,1} + \frac{\rho}{2} \|\mathcal{B} - \mathcal{B}^l\|_F^2\end{aligned}\tag{10}$$

$$\text{s.t. } \mathcal{B} \times_k \mathbf{D}_k = \mathcal{Z}_k, \quad k = 1, 2.$$

The augmented Lagrangian function of (10) is

$$\begin{aligned}L_\beta(\mathcal{B}, \mathcal{Z}_k, \mathcal{P}_k) = & \frac{1}{2} \|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A}^{l+1} - \mathcal{S}^l\|_F^2 \\ & + \sum_{k=1}^2 \left\{ \tau \|\mathcal{W}_k \odot \mathcal{Z}_k\|_{2,1} + \langle \mathcal{B} \times_k \mathbf{D}_k - \mathcal{Z}_k, \mathcal{P}_k \rangle \right. \\ & \left. + \frac{\beta}{2} \|\mathcal{B} \times_k \mathbf{D}_k - \mathcal{Z}_k\|_F^2 \right\} + \frac{\rho}{2} \|\mathcal{B} - \mathcal{B}^l\|_F^2,\end{aligned}\tag{11}$$

where \mathcal{P}_k ($k = 1, 2$) are Lagrange multipliers and $\beta > 0$ is the penalty parameter. To solve (10), we can alternately update \mathcal{B} , \mathcal{Z}_k , and \mathcal{P}_k as

$$\begin{cases} \mathcal{B}^{l+1,p+1} = \underset{\mathcal{B}}{\operatorname{argmin}} L_\beta(\mathcal{B}, \mathcal{Z}_k^p, \mathcal{P}_k^p), \\ \mathcal{Z}_k^{p+1} = \underset{\mathcal{Z}_k}{\operatorname{argmin}} L_\beta(\mathcal{B}^{l+1,p+1}, \mathcal{Z}_k, \mathcal{P}_k^p), \\ \mathcal{P}_k^{p+1} = \mathcal{P}_k^p + \beta(\mathcal{B}^{l+1,p+1} \times_k \mathbf{D}_k - \mathcal{Z}_k^{p+1}). \end{cases}\tag{12}$$

Next, we solve $\mathcal{B}^{l+1,p+1}$ and \mathcal{Z}_k^{p+1} in (12), respectively. For $\mathcal{B}^{l+1,p+1}$, we solve the following problem:

$$\begin{aligned}\underset{\mathcal{B}}{\operatorname{argmin}} \quad & \frac{1}{2} \|\mathcal{Y} - \mathcal{B} \times_3 \mathbf{A}^{l+1} - \mathcal{S}^l\|_F^2 + \frac{\rho}{2} \|\mathcal{B} - \mathcal{B}^l\|_F^2 \\ & + \sum_{k=1}^2 \frac{\beta}{2} \left\| \mathcal{B} \times_k \mathbf{D}_k - \mathcal{Z}_k^p + \frac{\mathcal{P}_k^p}{\beta} \right\|_F^2,\end{aligned}\tag{13}$$

which can be directly solved by the following equation:

$$\mathcal{B} \times_3 ((\mathbf{A}^{l+1})^T \mathbf{A}^{l+1}) + \sum_{k=1}^2 \beta \mathcal{B} \times_k (\mathbf{D}_k^T \mathbf{D}_k) + \rho \mathcal{B} = \mathcal{K},\tag{14}$$

where

$$\begin{aligned}\mathcal{K} = & (\mathcal{Y} - \mathcal{S}^l) \times_3 (\mathbf{A}^{l+1})^T \\ & + \sum_{k=1}^2 \beta \left(\mathcal{Z}_k^p - \frac{\mathcal{P}_k^p}{\beta} \right) \times_k \mathbf{D}_k^T + \rho \mathcal{B}^l.\end{aligned}$$

³ones(m, n) is an $m \times n$ matrix whose elements are all 1.

And (14) can be equivalently rewritten as

$$\mathbf{B}_{(3)}^T ((\mathbf{A}^{l+1})^T \mathbf{A}^{l+1}) + \rho \mathbf{B}_{(3)}^T + \mathbf{C} \mathbf{B}_{(3)}^T = \mathbf{K}_{(3)}^T,\tag{15}$$

where $\mathbf{C} = \beta[(\mathbf{I}_{n_2} \otimes \mathbf{D}_1^T \mathbf{D}_1) + (\mathbf{D}_2^T \mathbf{D}_2 \otimes \mathbf{I}_{n_1})]$. It is easy to find that the matrix \mathbf{C} has a structure of block circulant with circulant blocks and the matrix $(\mathbf{A}^{l+1})^T \mathbf{A}^{l+1}$ is a symmetric matrix. Thus, we employ two-dimensional FFT and SVD to diagonalize \mathbf{C} and $(\mathbf{A}^{l+1})^T \mathbf{A}^{l+1}$, respectively. That is,

$$\mathbf{C} = \mathbf{F}_2^T \Psi_2 \mathbf{F}_2 \quad \text{and} \quad (\mathbf{A}^{l+1})^T \mathbf{A}^{l+1} = \mathbf{U}_2 \Sigma_2 \mathbf{U}_2^T,$$

where \mathbf{F}_2 is two-dimensional DFT matrix. By using Theorem 1, we can efficiently solve (14) as

$$\mathcal{B}^{l+1,p+1} = \text{Fold}_3 \left([\mathbf{F}_2^T ((1 \otimes \mathbf{T}_2) \odot (\mathbf{F}_2 \mathbf{K}_{(3)}^T \mathbf{U}_2)) \mathbf{U}_2^T]^T \right),\tag{16}$$

where

$$\begin{aligned}\mathbf{T}_2 = & (\text{diag}(\Psi_2), \text{diag}(\Psi_2), \dots, \text{diag}(\Psi_2)) + \rho \text{ones}(n_1 n_2, r) \\ & + (\text{diag}(\Sigma_2), \text{diag}(\Sigma_2), \dots, \text{diag}(\Sigma_2))^T.\end{aligned}$$

For \mathcal{Z}_k^{p+1} ($k = 1, 2$), we solve the following problem:

$$\begin{aligned}\underset{\mathcal{Z}_k}{\operatorname{argmin}} \quad & \tau \|\mathcal{W}_k \odot \mathcal{Z}_k\|_{2,1} \\ & + \frac{\beta}{2} \left\| \mathcal{B}^{k+1,p+1} \times_k \mathbf{D}_k - \mathcal{Z}_k + \frac{\mathcal{P}_k^p}{\beta} \right\|_F^2,\end{aligned}\tag{17}$$

which can be directly solved by

$$\mathcal{Z}_k^{p+1}(i, j, :) = \text{shrink}_{2,1} \left(\hat{\mathcal{Z}}_k(i, j, :), |\mathbf{W}_k(i, j)| \cdot \frac{\tau}{\beta} \right),\tag{18}$$

where

$$\begin{aligned}\hat{\mathcal{Z}}_k &= \mathcal{B}^{l+1,p+1} \times_k \mathbf{D}_k + \frac{\mathcal{P}_k^p}{\beta}, \\ \mathbf{W}_k(i, j) &= \frac{1}{\|\hat{\mathcal{Z}}_k(i, j, :)\|_2 + \varepsilon}, \\ \text{shrink}_{2,1}(\mathbf{x}, \xi) &= \begin{cases} \frac{\|\mathbf{x}\|_2 - \xi}{\|\mathbf{x}\|_2} \mathbf{x}, & \text{if } \xi < \|\mathbf{x}\|_2, \\ 0 & \text{otherwise,} \end{cases}\end{aligned}$$

and ε is a small constant for avoiding the appearance of singularities.

3) *Update \mathcal{S}* : In Step 3 of (5), the \mathcal{S} -subproblem is as follows:

$$\begin{aligned}\underset{\mathcal{S}}{\operatorname{argmin}} \quad & \frac{1}{2} \|\mathcal{Y} - \mathcal{B}^{l+1} \times_3 \mathbf{A}^{l+1} - \mathcal{S}^l\|_F^2 \\ & + \mu \|\mathcal{W}_s \odot \mathcal{S}\|_1 + \frac{\rho}{2} \|\mathcal{S} - \mathcal{S}^l\|_F^2,\end{aligned}\tag{19}$$

which has the following solution:

$$\mathcal{S}^{l+1} = \text{shrink}_1 \left(\hat{\mathcal{S}}, \mathcal{W}_s \odot \frac{\mu}{1 + \rho} \right),\tag{20}$$

where

$$\hat{\mathcal{S}} = \frac{\mathcal{Y} - \mathcal{B}^{l+1} \times_3 \mathbf{A}^{l+1} + \rho \mathcal{S}^l}{1 + \rho},$$

$$\mathcal{W}_s(i, j, m) = \frac{1}{|\hat{\mathcal{S}}(i, j, m)| + \varepsilon},$$

Algorithm 1 PAM-Based Solver for the LRTF-DFR-Based HSI Restoration Model.

Input: The degraded observation HSI $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, rank r , parameters $\tau, \lambda, \mu, \beta$, and $\rho = 0.1$.

Initialization: $l = 0$, $l_{\max} = 50$, $p_{\max} = 10$, $\mathbf{A}^0 = \text{rand}(n_3, r)$, $\mathcal{B}^0 = \text{rand}(n_1, n_2, r)$, and $\mathcal{S}^0 = 0$.

- 1: **while** not converged and $l < l_{\max}$ **do**
- 2: Update \mathbf{A}^{l+1} via (8).
- 3: **Initialization:** $p = 0$, $\mathcal{Z}_k^0 = 0$ ($k = 1, 2$), and $\mathcal{P}_k^0 = 0$.
- 4: **while** $p < p_{\max}$ **do**
- 5: Update $\mathcal{B}^{l+1,p+1}$ via (16).
- 6: Update \mathcal{Z}_k^{p+1} via (18), $k = 1, 2$.
- 7: Update \mathcal{P}_k^{p+1} via (12), $k = 1, 2$.
- 8: Let $p = p + 1$.
- 9: Check the convergence condition:

$$\|\mathcal{B}^{l+1,p} - \mathcal{B}^{l+1,p-1}\|_F / \|\mathcal{B}^{l+1,p-1}\|_F < 10^{-4}.$$
- 10: **end while**
- 11: Let $\mathcal{B}^{l+1} = \mathcal{B}^{l+1,p}$.
- 12: Update \mathcal{S}^{l+1} via (20).
- 13: Let $l = l + 1$.
- 14: Check the convergence condition:

$$\|\mathcal{B}^l \times_3 \mathbf{A}^l - \mathcal{B}^{l-1} \times_3 \mathbf{A}^{l-1}\|_F / \|\mathcal{B}^{l-1} \times_3 \mathbf{A}^{l-1}\|_F < 10^{-4}.$$
- 15: **end while**

Output: The restored HSI $\mathcal{X} = \mathcal{B} \times_3 \mathbf{A}$.

and

$$[\text{shrink}_1(\mathcal{X}, \xi)]_{i,j,m} = \text{sign}(x_{i,j,m}) \max(|x_{i,j,m}| - \xi, 0).$$

By summarizing the aforementioned solving process, we describe the pseudocode of the developed PAM-based algorithm for solving the LRTF-DFR-based HSI restoration model in Algorithm 1.

D. Computational Complexity Analysis

We analyze the computational complexity of the developed PAM-based solving algorithm on a degraded HSI $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. As shown in Algorithm 1, the computational cost at each outer iteration mainly lies in the updating of \mathbf{A} , \mathcal{B} , \mathcal{Z}_k , \mathcal{P}_k , \mathcal{S} , and the number of inner iterations p , where $k = 1, 2$. First, updating \mathbf{A} via (8) involves SVD, one-dimensional FFT, and several matrix multiplications, which leads to $\mathcal{O}(rn_1n_2n_3 + r^2n_3 + rn_3\log(n_3))$ cost. Second, updating \mathcal{B} via (16) needs $\mathcal{O}((r + n_1 + n_2)n_1n_2n_3 + r^2n_3 + rn_1n_2\log(n_1n_2))$ cost, since it involves SVD, two-dimensional FFT, and several tensor-matrix product operations. Third, the computational cost of updating \mathcal{Z}_k ($k = 1, 2$) via (18) is $\mathcal{O}(n_1n_2n_3)$. Fourth, the computational cost of updating \mathcal{P}_k via (12) is $\mathcal{O}(n_1^2n_2r)$. Finally, the computational cost of updating \mathcal{S} via (20) is $\mathcal{O}(n_1n_2n_3)$. In summary, the computational cost at each outer iteration of the developed PAM-based algorithm is $\mathcal{O}(rn_3\log(n_3) + p((r + n_1 + n_2)n_1n_2n_3 + rn_1n_2\log(n_1n_2)))$.

IV. NUMERICAL EXPERIMENTS

We test the performance of the proposed LRTF-DFR-based HSI restoration method⁴ by conducting extensive experiments

⁴The code of the proposed LRTF-DFR-based HSI restoration method is available at https://github.com/YuBangZheng/code_LRTFDFR.

on both simulated and real HSIs. To comprehensively evaluate the proposed method, we compare five excellent HSI restoration methods, i.e., LRMR [35], NMoG [39], SNLRSF [56], LRTDTV [52], and LRTDGS [53]. We select LRMR and NMoG since they represent the classic methods based on matrix rank minimization and noise modeling, respectively. We select SNLRSF, LRTDTV, and LRTDGS since they are the most relevant to the proposed method among the newer methods for mixed noise removal. The detailed relations and distinctions between the proposed method and the compared methods are summarized in Table II.

All parameters involved in the compared methods are carefully adjusted based on the authors' suggestions in their articles to obtain the optimal performance. The gray values of HSIs are normalized into the interval [0,1] band-by-band. All experiments are implemented by using MATLAB (R2019a) on Windows 10 with an Intel Core i9-9900K 3.60 GHz processor and 32 GB of RAM.

A. Simulated Data Experiments

We employ two public HSIs in this section, one is the simulated Indian Pines dataset⁵ with the size of (145,145,224), which is also used in the compared methods LRTDTV and LRTDGS, the other is the Washington DC Mall dataset⁶ with the size of (256,256,191), which is also used in the compared methods LRMR, NMoG, SNLRSF, and LRTDGS. Three quantitative evaluation indexes, including the mean of peak signal-to-noise rate (MPSNR) over all bands, the mean of structural similarity (MSSIM) over all bands, and the mean of spectral angle mapping (MSAM) over all spectral vectors, are selected to evaluate the overall quality of the restored results quantitatively. Since the noise in HSIs usually manifests as a mixture of several kinds of noise in real noise scenarios, we consider the following five cases.

Case 1: Gaussian Noise. Gaussian noise with zero mean is added to all bands, and the noise standard deviation in each band is randomly sampled from the interval [0.1,0.2].

Case 2: Gaussian Noise + Salt and Pepper Noise. Gaussian noise is added in the same way as Case 1. Furthermore, salt and pepper noise is added to all bands, and the noise proportion in each band is randomly sampled from the interval [0.1,0.2].

Case 3: Gaussian Noise + Salt and Pepper Noise + Stripe. Gaussian noise and salt and pepper noise are added in the same way as Case 2. Furthermore, 40% of all bands are randomly selected to add stripes, and the number of stripes in each selected band is randomly sampled from the set [6,7,· · ·,15]. Especially, all elements of the whole column will become to a certain value randomly sampled from the interval [0.6,0.8] if this column is selected to add the stripe.

Case 4: Gaussian Noise + Salt and Pepper Noise + Deadline. Gaussian noise and salt and pepper noise are added in the same way as Case 2. Furthermore, 20% of all bands are randomly selected to add deadlines with the width of randomly sampled from the set [1,2,3], and the number of

⁵This dataset is generated in the same way as [49].

⁶<http://lesun.weebly.com/hyperspectral-data-set.html>.

TABLE II
THE DETAILED RELATIONS AND DISTINCTIONS BETWEEN THE PROPOSED METHOD AND THE COMPARED METHODS.

Method	Low-rankness			Local continuity			Nonlocal self-similarity			Group sparsity of SpatDI	Noise assumption
	spatial	spectral	both	spatial	spectral	both	spatial	spectral	both		
LRMR	<input checked="" type="checkbox"/>			—			—			—	Gaussian and Laplace distributions
	minimizing the matrix rank			—			—				
NMoG	<input checked="" type="checkbox"/>			—			—			—	non-i.i.d. mixture of Gaussian distributions
	LRMF			—			—				
SNLRSF	<input checked="" type="checkbox"/>			—			<input checked="" type="checkbox"/> nonlocal low-rank regularization on factor			—	Gaussian and Laplace distributions
	subspace representation			—			—				
LRTDTV	<input checked="" type="checkbox"/>			SSTV			—			—	Gaussian and Laplace distributions
	LRTD			—			—				
LRTDGS	<input checked="" type="checkbox"/>			weighted $\ell_{2,1}$ -norm on SpatDI of HSIs			—			weighted $\ell_{2,1}$ -norm on SpatDI of HSIs	Gaussian and Laplace distributions
	LRTD			—			—				
LRTF-DFR	<input checked="" type="checkbox"/>			double factor-based regularization			—			weighted $\ell_{2,1}$ -norm on SpatDI of spatial factor	Gaussian and Laplace distributions
	LRTF			—			—				

TABLE III
THE QUANTITATIVE COMPARISON OF ALL COMPARED METHODS FOR DIFFERENT DATASETS AND CASES.

Dataset	Case	Indexes	Noise	LRMR	NMoG	SNLRSF	LRTDTV	LRTDGS	LRTF-DFR	
Indian Pines	Case 1	MPSNR	16.642	32.718	31.992	40.709	37.988	40.477	43.935	
		MSSIM	0.2697	0.8682	0.8289	0.9875	0.9822	0.9918	0.9971	
		MSAM	16.984	2.3167	2.6231	0.7881	1.1982	0.8877	0.6132	
	Case 2	MPSNR	11.752	31.103	30.290	31.026	35.112	38.569	41.403	
		MSSIM	0.1442	0.8292	0.7987	0.8644	0.9674	0.9865	0.9965	
		MSAM	27.294	2.7518	3.1394	3.0024	1.6669	1.1046	0.8490	
	Case 3	MPSNR	11.711	30.006	29.674	29.813	33.440	35.534	39.165	
		MSSIM	0.1411	0.8225	0.7972	0.8577	0.9608	0.9758	0.9951	
		MSAM	27.204	3.4053	3.4743	3.5800	2.1837	2.0246	1.3131	
	Case 4	MPSNR	11.518	29.401	29.566	29.188	33.102	33.487	38.739	
		MSSIM	0.1404	0.8808	0.7955	0.8416	0.9583	0.9490	0.9938	
		MSAM	28.402	4.2670	3.7107	4.1584	2.4177	3.4266	1.7831	
Washington DC Mall	Case 5	MPSNR	11.409	28.216	29.019	27.827	32.307	32.263	37.948	
		MSSIM	0.1362	0.7913	0.7916	0.8286	0.9548	0.9343	0.9942	
		MSAM	28.566	5.0425	4.1730	5.1917	2.8256	4.0121	1.7715	
	Mean time (s)			—	25.971	99.320	20.014	102.18	73.057	41.512
	Case 1	MPSNR	16.658	31.771	33.744	36.885	32.574	34.167	35.686	
		MSSIM	0.2915	0.9056	0.9376	0.9698	0.9135	0.9398	0.9585	
		MSAM	42.481	10.340	9.4375	5.1332	6.2833	7.0713	5.0720	
	Case 2	MPSNR	11.055	30.113	32.125	28.448	30.225	32.804	34.626	
		MSSIM	0.1257	0.8760	0.9174	0.8788	0.8669	0.9220	0.9493	
		MSAM	51.274	10.481	13.636	8.2613	7.4229	7.8837	5.4400	
	Case 3	MPSNR	10.874	29.245	31.893	27.225	29.579	31.981	34.464	
		MSSIM	0.1216	0.8642	0.9150	0.8549	0.8548	0.9158	0.9477	
		MSAM	50.975	11.073	11.082	9.0752	7.7969	10.105	5.7019	
	Case 4	MPSNR	11.071	29.772	32.083	28.654	30.019	32.565	34.061	
		MSSIM	0.1245	0.8723	0.9210	0.8786	0.8654	0.9200	0.9462	
		MSAM	51.653	11.004	11.632	8.6462	7.7726	8.4999	5.9657	
	Case 5	MPSNR	10.863	28.858	31.918	27.515	29.545	31.859	33.933	
		MSSIM	0.1201	0.8593	0.9221	0.8600	0.8508	0.9164	0.9450	
		MSAM	51.462	11.752	8.6390	9.3490	8.0001	10.036	6.3431	
	Mean time (s)			—	64.054	208.11	58.121	290.47	240.88	119.08

deadlines in each selected band is randomly sampled from the set $[6, 7, \dots, 10]$. Especially, all elements of the whole column will become to zero if this column is selected to add the deadline.

Case 5: Gaussian Noise + Salt and Pepper Noise + Stripe + Deadline. Gaussian noise and salt and pepper noise are added in the same way as Case 2. Furthermore, stripes and deadlines

are added in the same way as Case 3 and Case 4, respectively.

1) *Quantitative Comparison:* We report the MPSNR, MSSIM, MSAM, and the mean time values obtained by all compared methods in Table III. Especially, we highlight the best results by bold. As observed, the proposed LRTF-DFR achieves overall superior results over the compared ones under nearly all cases. Although under case 1, SNLRSF outputs

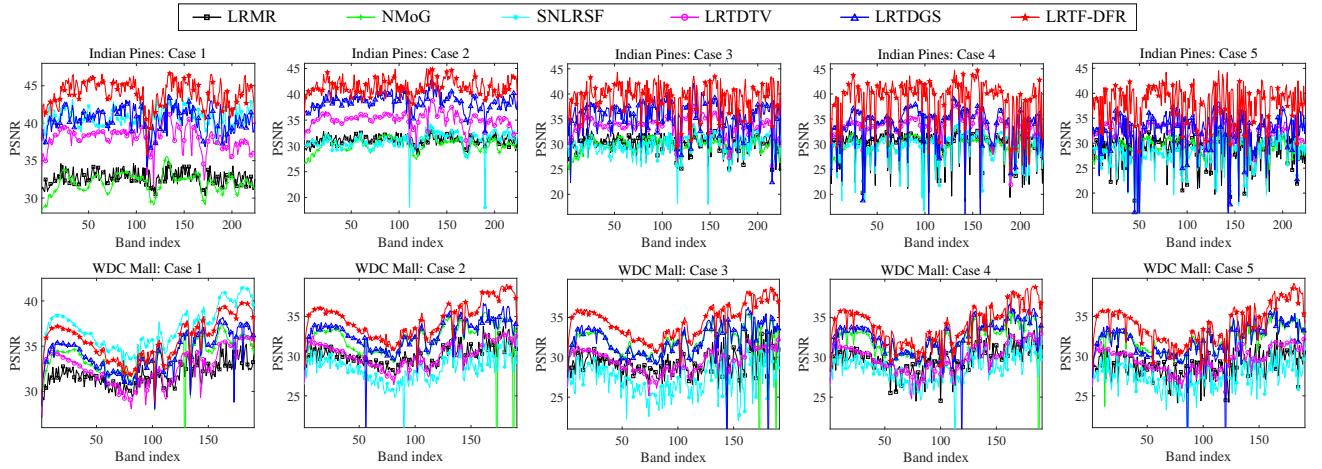


Fig. 3. PSNR values of all bands obtained by different compared methods under different cases. From left to right: the results of case 1, case 2, case 3, case 4, and case 5, respectively. The first and the second rows are the results on HSIs Indian Pines and Washington DC Mall, respectively.

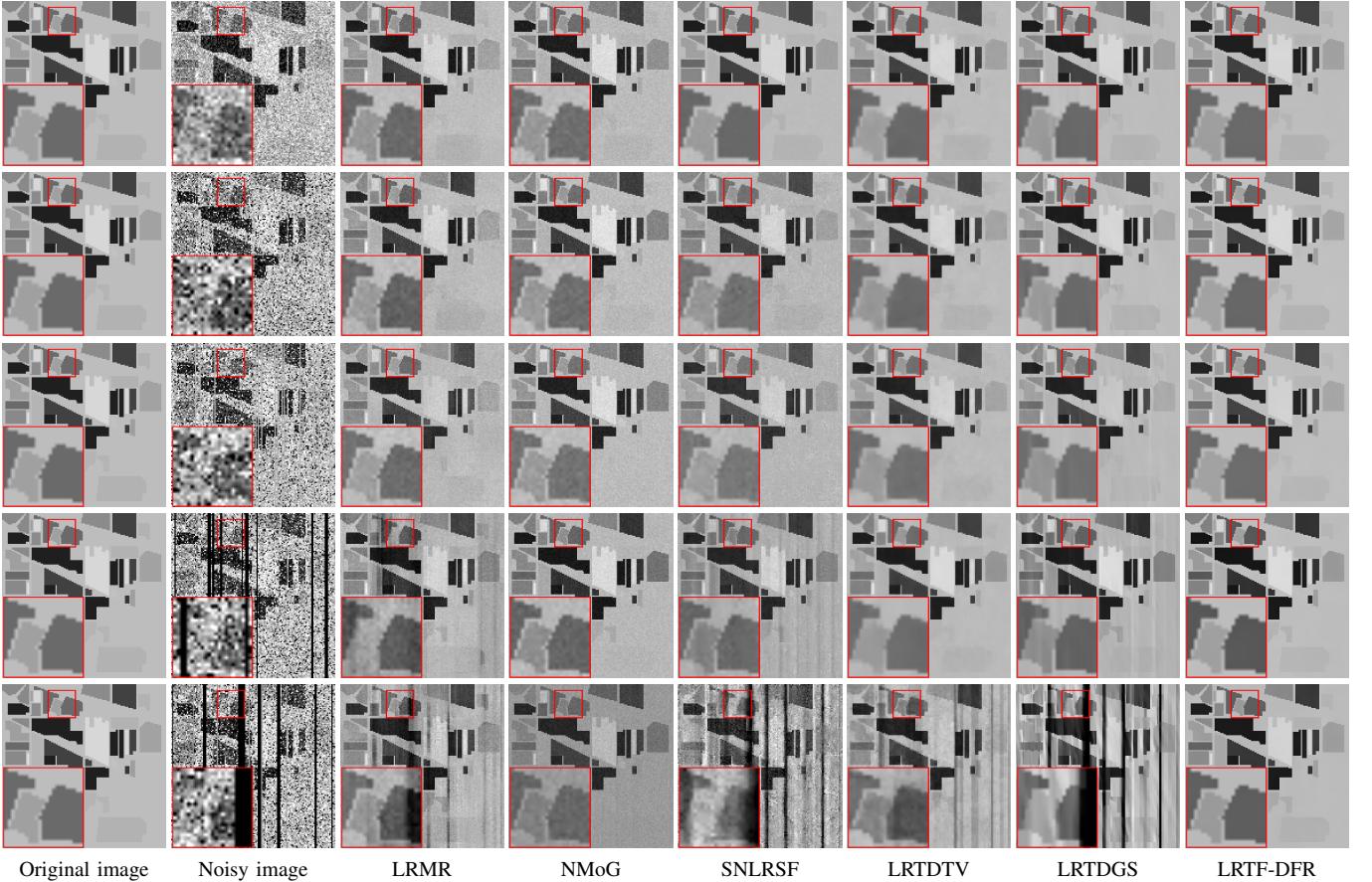


Fig. 4. Restoration results at band 48 of the Indian Pines dataset. From top to bottom: the results under case 1, case 2, case 3, case 4, and case 5, respectively. From left to right: the original image, the noisy image, the results of LRMR, NMoG, SNLRSF, LRTDTV, LRTDGS, and LRTF-DFR, respectively.

slightly higher MPSNR and MSSIM values than the proposed method on the Washington DC Mall dataset, it obviously underperforms the proposed method under other cases. For the running time, the proposed method costs a shorter time than LRTDTV and LRTDGS. The main reason is that the proposed LRTF-DFR characterizes the HSI priors by introducing constraints on factors, while LRTDTV and LRTDGS by directly adding constraints on the original HSI. Meanwhile,

the proposed method also costs a shorter time than NMoG with the reason that NMoG spends much time in noise learning. To compare the performance of each band, Fig. 3 shows the PSNR values of each band obtained by six compared methods. As observed, the proposed LRTF-DFR achieves the highest PSNR values among all compared methods in most bands and cases.

2) *Qualitative Comparison:* To visually compare the restoration results, Fig. 4 shows band 48 of the Indian Pines

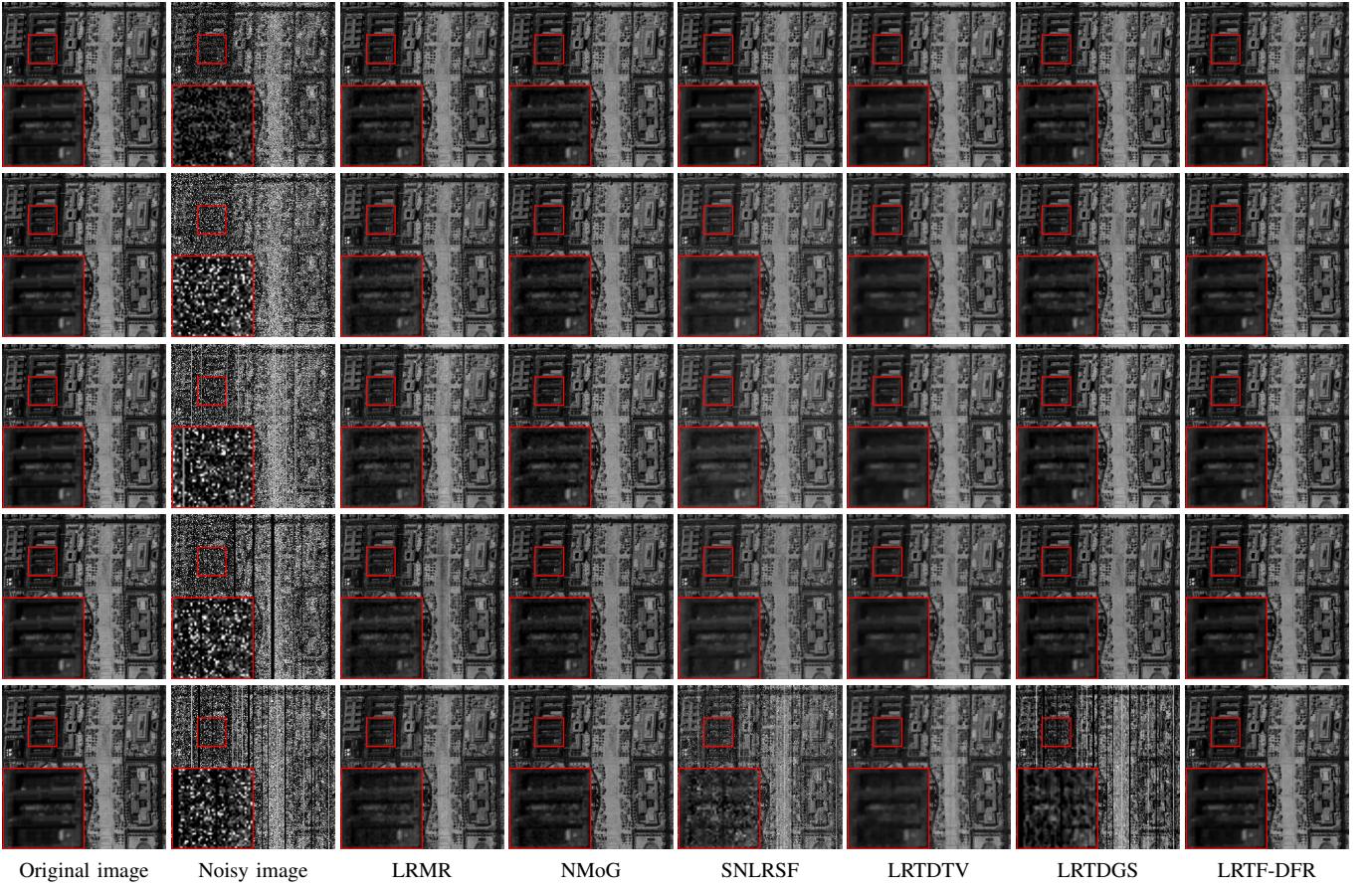


Fig. 5. Restoration results at band 86 of the Washington DC Mall dataset. From top to bottom: the results under case 1, case 2, case 3, case 4, and case 5, respectively. From left to right: the original image, the noisy image, the results of LRMR, NMoG, SNLRSF, LRTDTV, LRTDGS, and LRTF-DFR, respectively.

dataset restored by six compared methods under all cases. We observe from Fig. 4 that the proposed LRTF-DFR obtains the best visual results among all compared methods, both in noise removal and the recovery of global structure and local details. Here LRMR, NMoG, and SNLRSF are not able to fully remove the noises, and LRTDTV and LRTDGS are not capable of well recovering the local details, especially the edges. Especially for case 5, the results of LRMR, SNLRSF, LRTDTV, and LRTDGS contain larger amounts of the trace of the thick deadlines. Comparatively, the proposed LRTF-DFR completely removes all noises while finely preserving the global structure and local details of the target HSI. Fig. 5 shows band 86 of the Washington DC Mall dataset restored by six compared methods under all cases. It can be seen that the results of the proposed LRTF-DFR have an obvious improvement in noise removal and details preserving, as compared with the other methods.

To further compare the performance of the spectral curve recovery, cases 2 and 5 are selected as two representative cases. Fig. 6 shows the spectral curves at one spatial location of the restoration results by different compared methods under cases 2 and 5. It is observed that the spectral curves obtained by the proposed LRTF-DFR can better approximate the original ones than those produced by the compared methods. In summary, the above observations illustrate that the proposed LRTF-DFR has the best performance on mixed noise removal, spatial

image recovery, and spectral signatures preserving, among all compared methods.

The reason why the proposed LRTF-DFR is superior to the compared ones is that it can fully utilize the relevant information among different bands. More specifically, the proposed LRTF-DFR not only utilizes the global low-rankness and the local continuity but also assumes that most of the smooth areas in different bands located at the same location, i.e., the SpatDI_s of HSIs are group sparse along the spectral mode. This assumption makes proposed LRTF-DFR to fully exploit the common characteristics among different bands. By contrast, LRMR, NMoG, and SNLRSF fail to consider the local continuity, LRTDTV does not consider the spectral group sparsity in the SpatDI_s, and LRTDGS does not consider the local continuity in the spectral domain.

B. Real Data Experiments

This section employs two real-world HSIs, one is the Urban dataset⁶ with size of (307,307,210), which is also used in all compared methods, the other is the real Indian Pines dataset⁶ with size of (145,145,220), which is also used in the compared methods SNLRSF, LRTDTV, and LRTDGS.

1) *Urban Dataset*: Fig. 7 shows the restoration results at band 109 of the real HSI Urban, including the visual effects and the vertical mean profiles. Here the vertical mean

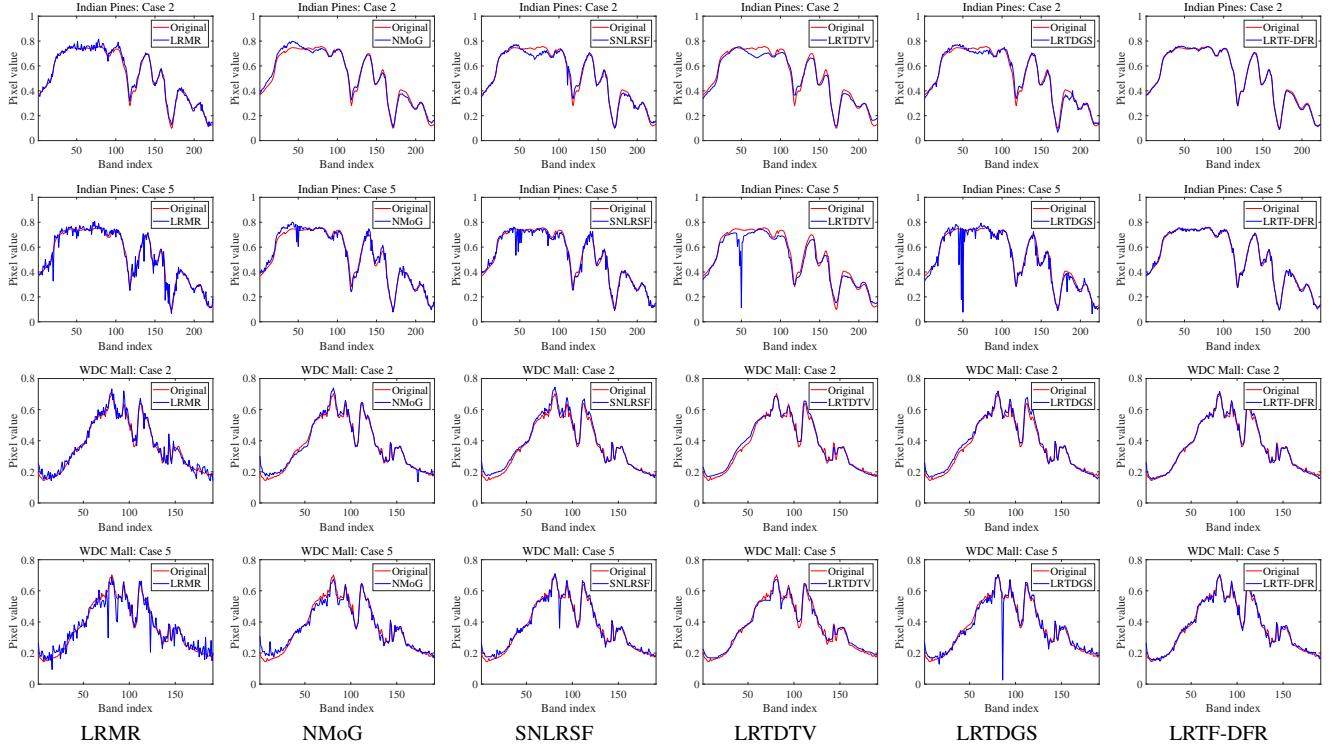


Fig. 6. Spectral curves of the restoration results by different compared methods. From left to right: the results of LRMR, NMoG, SNLRSF, LRTDTV, LRTDGS, and LRTF-DFR, respectively. The first two rows are the results at spatial location (30,30) of the Indian Pines dataset under case 2 and case 5, respectively. The last two rows are the results at spatial location (151,151) of the Washington DC Mall dataset under case 2 and case 5, respectively.

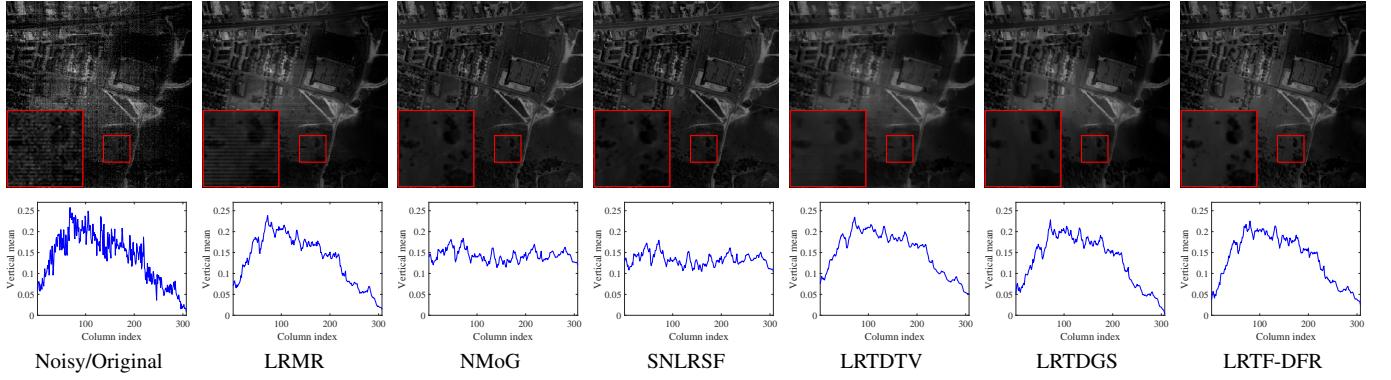


Fig. 7. Restoration results at band 109 of the real HSI Urban. From left to right: the noisy/original result, the results of LRMR, NMoG, SNLRSF, LRTDTV, LRTDGS, and LRTF-DFR, respectively. The first and the second rows are the visual results and the vertical mean profiles, respectively.

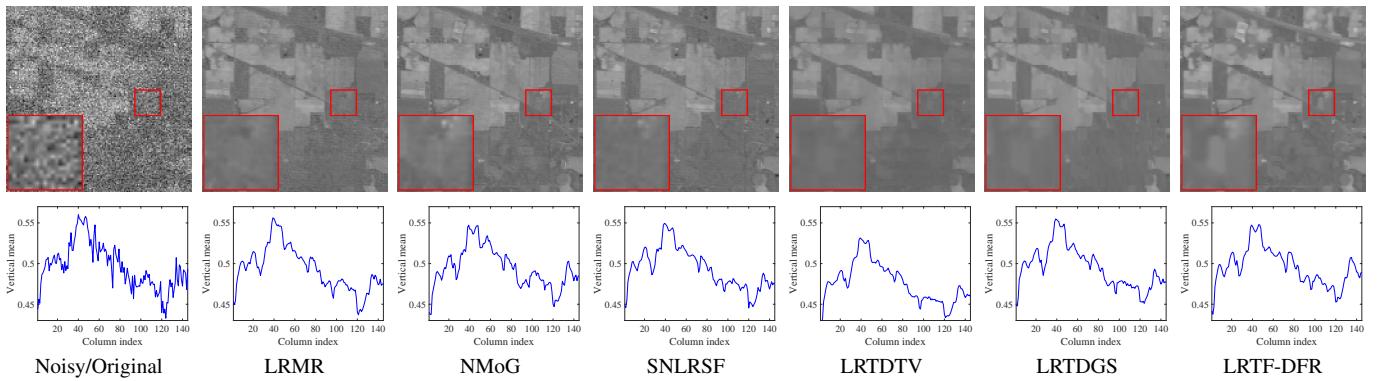


Fig. 8. Restoration results at band 220 of the real HSI Indian Pines. From left to right: the noisy/original result, the results of LRMR, NMoG, SNLRSF, LRTDTV, LRTDGS, and LRTF-DFR, respectively. The first and the second rows are the visual results and the vertical mean profiles, respectively.

TABLE IV
THE INFLUENCE OF THE NUMBER OF INNER ITERATION UNDER CASE 5 IN SIMULATED DATA EXPERIMENTS.

Dataset	p	2	4	6	8	10	12	14	16	18	20
Indian Pines	MPSNR	33.284	36.354	37.146	37.541	37.986	38.112	38.148	38.152	38.178	38.184
	CPU time (s)	14.941	20.745	27.441	35.012	41.424	49.014	57.454	65.143	73.451	81.354
Washington DC Mall	MPSNR	33.210	33.547	33.697	33.821	33.921	33.924	33.931	33.942	33.956	33.961
	CPU time (s)	43.853	62.963	82.847	100.01	120.83	139.98	157.92	178.78	198.14	217.22

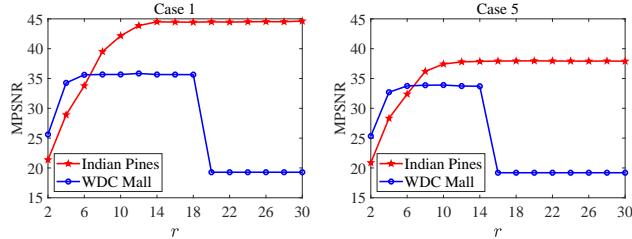


Fig. 9. Sensitivity analysis of the rank r under cases 1 and 5.

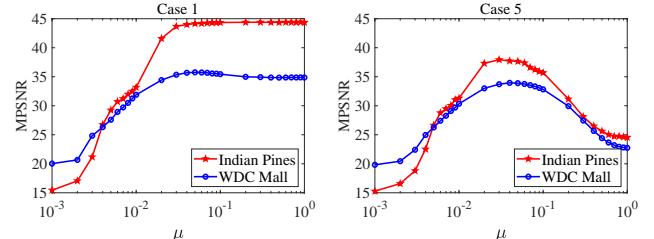


Fig. 11. Sensitivity analysis of the parameter μ under cases 1 and 5.

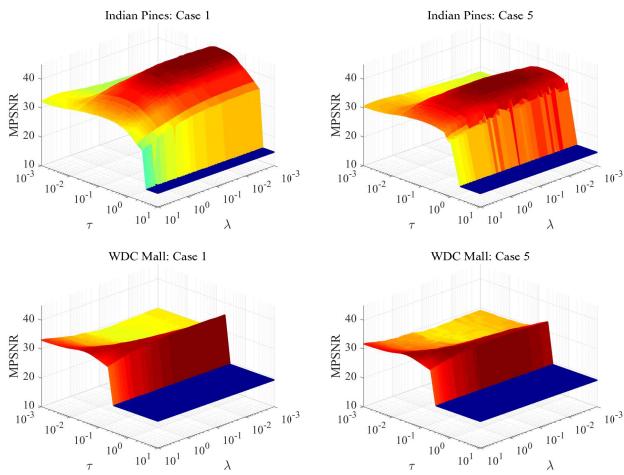


Fig. 10. Sensitivity analysis of the parameters τ and λ under cases 1 and 5.

profile is a curve obtained by calculating the mean value of pixels in each column. As observed, all compared methods can effectively remove the mixed noise except LRMR, which cannot completely remove the stripes. But by observing the enlarged box in Fig. 7, we can see that the results of LRTDTV and LRTDGS lose some local details. NMoG and SNLRSF have excellent ability to preserve the image details, whereas their vertical mean profiles deviate from the basic trend of the observed ones, implying the change of image contrast. Comparatively, the proposed LRTF-DFR completely removes all noises, precisely preserves the global structure and local details, and finely retains the image contrast.

2) *Indian Pines Dataset*: Fig. 8 shows the restoration results at band 220 of the real HSI Indian Pines, including the visual effects and the vertical mean profiles. It is obvious that the result of LRMR contains a little noise, the reason is that LRMR only considers the low-rankness prior. NMoG, SNLRSF, LRTDTV, LRTDGS, and the proposed LRTF-DFR all have excellent ability to eliminate the noise and can produce relatively smooth vertical mean profiles. But the proposed

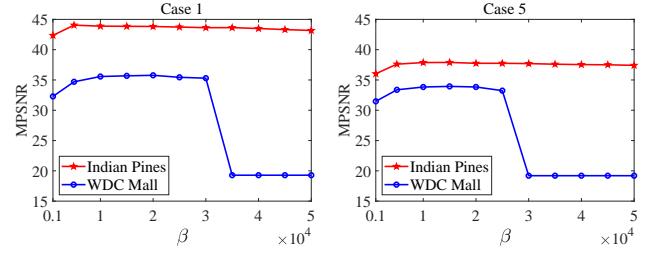


Fig. 12. Sensitivity analysis of the parameter β under cases 1 and 5.

LRTF-DFR offers major advantages over the other ones for recovering local details, especially the edges.

C. Discussions

1) *Inner Iteration Analysis*: We test the influence of the number of inner iteration p on the simulated datasets under cases 5. Table IV reports the MPSNR and CPU time values with respect to different p . It is seen that the proposed LRTF-DFR obtains a stable and superior performance when $p \geq 10$. Considering that the CPU time is increased with increasing of p , we set p to be 10 in all experiments.

2) *Parameter Analysis*: We test the sensitivity of the parameters on the simulated datasets under cases 1 and 5 since they represent the Gaussian noise case and mixed noise case, respectively. As shown in Algorithm 1, the proposed LRTF-DFR method involves five important parameters: the rank r , regularization parameters τ , λ , and μ , and the penalty parameter β .

The rank r mainly characterizes the spectral global correlation of HSIs. The sensitivity analysis of the rank r is presented in Fig. 9. It is seen that the proposed LRTF-DFR exhibits stable and superior performance within a certain range of the rank r . Considering that larger r leads to higher computational complexity, we set r to be 12 and 8 for the simulated Indian Pines dataset and Washington DC Mall dataset, respectively. Especially, the rank r is estimated by the well-known HySime algorithm [62] in real data experiments.

TABLE V
PARAMETERS SETTING OF THE PROPOSED LRTF-DFR IN SIMULATED DATA EXPERIMENTS.

Dataset	r	τ	λ	μ	β
Indian Pines	12	0.2	0.01	0.04	15000
Washington DC Mall	8				

TABLE VI
PARAMETERS SETTING OF THE PROPOSED LRTF-DFR IN REAL DATA EXPERIMENTS.

Dataset	r	τ	λ	μ	β
Urban Indian Pines	Estimated by the HySime algorithm [62]	0.2	5	0.04	15000

The parameters τ and λ determine the weights of the factor \mathcal{B} -based and the factor \mathbf{A} -based regularization terms, respectively. Fig. 10 shows their sensitivity analysis. As observed, the performance of the proposed LRTF-DFR is robust to the parameter λ , while it is sensitive to the parameter τ , especially on the Washington DC Mall dataset. Moreover, as the parameters τ and λ changes, the MPSNR values have the same changing tendency for different noisy cases. This implies that these two parameters are robust to different noisy cases.

The parameter μ determines the weight of the sparse noise term. Fig. 11 shows the sensitivity analysis of μ . It is seen that the proposed LRTF-DFR has stable and superior performance under case 1, when μ is larger than 0.04. This is because case 1 only involves Gaussian noise. For case 5, the performance of the proposed method is slightly sensitive to μ and the higher MPSNR values are achieved at $\mu = 0.03, 0.04$, and 0.05 .

The sensitivity analysis of the penalty parameter β is shown in Fig. 12. It is obvious that the performance of the proposed LRTF-DFR is robust within a certain range of β .

Under the guidance of the above analysis, Tables V and VI list the parameters setting of the proposed method in simulated data experiments and real data experiments, respectively.

3) *Convergence Analysis*: Since the ADMM and weighted strategies are embedded into the PAM framework, it is still an open problem for theoretically proving the convergence of the developed algorithm [57], [63]. Instead, we numerically demonstrate the convergence. Fig. 13 shows the relative change of the restored HSIs in each iteration and its previous iteration under cases 1 and 5. It is observed that for all testing datasets, the values of the relative change achieved by the developed algorithm monotonically decrease and gradually tend to zeros, as the number of iterations increases. This justifies the strong convergence of the developed PAM-based solving algorithm numerically.

V. CONCLUSION

In this article, we proposed an LRTF-DFR model for HSI mixed noise removal, which employed the LRTF framework to characterize the spectral global low-rankness of HSIs, introduced a group sparsity constraint on the SpatDI of the spatial factor to promote the group sparsity in the SpatDI of HSIs, and suggested a continuity constraint on the spectral

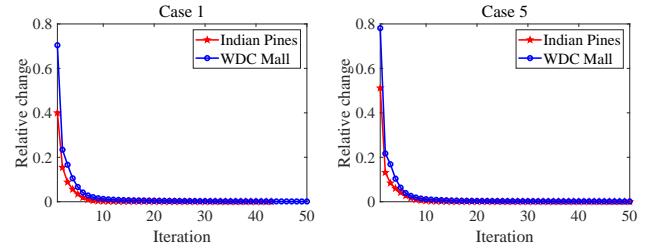


Fig. 13. Relative change curves with respect to the iteration number under cases 1 and 5.

factor to promote the spectral continuity of HSIs. Especially, two weighted strategies are employed to better promote the group sparsity and sparsity, respectively. To solve the proposed LRTF-DFR model, a PAM-based algorithm was developed with a numerically guaranteed convergence. By comparing with several excellent methods, including LRMR [35], NMoG [39], SNLRSF [56], LRTDTV [52], and LRTDGS [53], in extensive numerical experiments, the proposed method exhibited its superior performance on mixed noise removal, spatial image recovery, and spectral signatures preserving.

In the future, we will attempt to combine the proposed LRTF-DFR with the convolutional neural network [45], [64]–[66] to learn a more appropriate regularization on factors and further improve the ability of mixed noise removal.

ACKNOWLEDGMENT

The authors would like to express their sincere thanks to the editors and reviewers for their constructive comments, which are all valuable and helpful for improving this article. The authors would like to express their sincere thanks to the medical staff on the front line of fighting the COVID-19. It is their dedication and sacrifice that provide people a secure and stable research environment in this special time.

REFERENCES

- [1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, “Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.
- [2] H. Zhang, J. Li, Y. Huang, and L. Zhang, “A nonlocal weighted joint sparse representation classification method for hyperspectral imagery,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2056–2065, 2014.
- [3] S. Prasad, D. Labate, M. Cui, and Y. Zhang, “Morphologically decoupled structured sparsity for rotation-invariant hyperspectral image analysis,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4355–4366, 2017.
- [4] 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.
- [5] X. Fu, W. Ma, T. Chan, and J. M. Bioucas-Dias, “Self-dictionary sparse regression for hyperspectral unmixing: Greedy pursuit and pure pixel search are related,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1128–1141, 2015.
- [6] X. Fu, W. Ma, J. M. Bioucas-Dias, and T. Chan, “Semiblind hyperspectral unmixing in the presence of spectral library mismatches,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5171–5184, 2016.
- [7] W. He, H. Zhang, and L. Zhang, “Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3909–3921, 2017.

- [8] J. Li, X. Liu, Q. Yuan, H. Shen, and L. Zhang, "Antinoise hyperspectral image fusion by mining tensor low-multilinear-rank and variational properties," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 10, pp. 7832–7848, 2019.
- [9] Y. Liu, G. Gao, and Y. Gu, "Tensor matched subspace detector for hyperspectral target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 4, pp. 1967–1974, 2017.
- [10] J. Li, Q. Yuan, H. Shen, and L. Zhang, "Noise removal from hyperspectral image with joint spectral-spatial distributed sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5425–5439, 2016.
- [11] X.-L. Zhao, W. Wang, T.-Y. Zeng, T.-Z. Huang, and M. K. Ng, "Total variation structured total least squares method for image restoration," *SIAM Journal on Scientific Computing*, vol. 35, no. 6, pp. B1304–B1320, 2013.
- [12] X.-L. Zhao, F. Wang, and M. K. Ng, "A new convex optimization model for multiplicative noise and blur removal," *SIAM Journal on Imaging Sciences*, vol. 7, no. 1, pp. 456–475, 2014.
- [13] Y. Chang, L. Yan, H. Fang, and C. Luo, "Anisotropic spectral-spatial total variation model for multispectral remote sensing image destriping," *IEEE Transactions on Image Processing*, vol. 24, no. 6, pp. 1852–1866, 2015.
- [14] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, "Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm," *Journal of Computational and Applied Mathematics*, p. 112680, to be published. doi: 10.1016/j.cam.2019.112680.
- [15] M. Ding, T.-Z. Huang, T.-Y. Ji, T.-Y. Ji, X.-L. Zhao, and J.-H. Yang, "Low-rank tensor completion using matrix factorization based on tensor train rank and total variation," *Journal of Scientific Computing*, vol. 81, no. 2, pp. 941–964, 2019.
- [16] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 8, pp. 1888–1902, 2018.
- [17] J.-H. Yang, X.-L. Zhao, T.-Y. Ji, T.-H. Ma, and T.-Z. Huang, "Low-rank tensor train for tensor robust principal component analysis," *Applied Mathematics and Computation*, vol. 367, p. 124783, 2020.
- [18] W. He, N. Yokoya, L. Yuan, and Q. Zhao, "Remote sensing image reconstruction using tensor ring completion and total variation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 8998–9009, 2019.
- [19] Q. Yao, J. T.-Y. Kwok, T. Wang, and T. Liu, "Large-scale low-rank matrix learning with nonconvex regularizers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2628–2643, 2019.
- [20] W. He, H. Zhang, H. Shen, and L. Zhang, "Hyperspectral image denoising using local low-rank matrix recovery and global spatial-spectral total variation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 713–729, 2018.
- [21] Y.-B. Zheng, T.-Z. Huang, T.-Y. Ji, X.-L. Zhao, T.-X. Jiang, and T.-H. Ma, "Low-rank tensor completion via smooth matrix factorization," *Applied Mathematical Modelling*, vol. 70, pp. 677–695, 2019.
- [22] J.-H. Yang, X.-L. Zhao, T.-H. Ma, Y. Chen, T.-Z. Huang, and M. Ding, "Remote sensing images destriping using unidirectional hybrid total variation and nonconvex low-rank regularization," *Journal of Computational and Applied Mathematics*, vol. 363, pp. 124–144, 2020.
- [23] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4274–4288, 2018.
- [24] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 60–65.
- [25] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [26] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
- [27] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2862–2869.
- [28] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3660–3677, 2012.
- [29] G. Chen and S. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 3, pp. 973–980, 2011.
- [30] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian, "Denoising of multispectral images via nonlocal groupwise spectrum-PCA," in *Conference on Colour in Graphics, Imaging, and Vision*, 2010, pp. 261–266.
- [31] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2949–2956.
- [32] Y. Chen, W. He, N. Yokoya, T.-Z. Huang, and X.-L. Zhao, "Nonlocal tensor-ring decomposition for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1348–1362, 2020.
- [33] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 730–742, 2018.
- [34] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 6868–6877.
- [35] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, 2014.
- [36] 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 Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4642–4659, 2016.
- [37] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, and G. He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5366–5380, 2017.
- [38] X. Cao, Q. Zhao, D. Meng, Y. Chen, and Z. Xu, "Robust low-rank matrix factorization under general mixture noise distributions," *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4677–4690, 2016.
- [39] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-i.i.d. noise structure," *IEEE Transactions on Cybernetics*, vol. 48, no. 3, pp. 1054–1066, 2018.
- [40] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 138–142, 2008.
- [41] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2061–2069, 2008.
- [42] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3717–3724, 2012.
- [43] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 10, pp. 4589–4604, 2017.
- [44] 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 Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 734–749, 2020.
- [45] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "HSI-DeNet: Hyperspectral image restoration via convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 667–682, 2019.
- [46] Q. Zhang, Q. Yuan, J. Li, X. Liu, H. Shen, and L. Zhang, "Hybrid noise removal in hyperspectral imagery with a spatial-spectral gradient network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 10, pp. 7317–7329, 2019.
- [47] W. Dong, H. Wang, F. Wu, G. Shi, and X. Li, "Deep spatial-spectral representation learning for hyperspectral image denoising," *IEEE Transactions on Computational Imaging*, vol. 5, no. 4, pp. 635–648, 2019.
- [48] B. Lin, X. Tao, and J. Lu, "Hyperspectral image denoising via matrix factorization and deep prior regularization," *IEEE Transactions on Image Processing*, vol. 29, pp. 565–578, 2020.
- [49] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE*

- Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 178–188, 2016.
- [50] Q. Wang, Z. Wu, J. Jin, T. Wang, and Y. Shen, “Low rank constraint and spatial spectral total variation for hyperspectral image mixed denoising,” *Signal Processing*, vol. 142, pp. 11–26, 2018.
- [51] H. Zhang, L. Liu, W. He, and L. Zhang, “Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition,” *IEEE Transactions on Geoscience and Remote Sensing*, to be published. doi: 10.1109/TGRS.2019.2947333.
- [52] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X. Zhao, and D. Meng, “Hyperspectral image restoration via total variation regularized low-rank tensor decomposition,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1227–1243, 2018.
- [53] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, “Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition.” *IEEE Transactions on Cybernetics*, to be published. doi: 10.1109/TCYB.2019.2936042.
- [54] H. K. Aggarwal and A. Majumdar, “Hyperspectral unmixing in the presence of mixed noise using joint-sparsity and total variation,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4257–4266, 2016.
- [55] Y. Chen, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, “Hyperspectral image restoration using framelet-regularized low-rank nonnegative matrix factorization,” *Applied Mathematical Modelling*, vol. 63, pp. 128–147, 2018.
- [56] C. Cao, J. Yu, C. Zhou, K. Hu, F. Xiao, and X. Gao, “Hyperspectral image denoising via subspace-based nonlocal low-rank and sparse factorization,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 3, pp. 973–988, 2019.
- [57] 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,” *Mathematical Programming*, 2013.
- [58] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [59] J. Peng, Q. Xie, Q. Zhao, Y. Wang, D. Meng, and Y. Leung, “Enhanced 3DTV regularization and its applications on hyper-spectral image denoising and compressed sensing,” *arXiv preprint arXiv:1809.06591*, 2018.
- [60] E. J. Candès, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *Journal of Fourier analysis and applications*, vol. 14, no. 5, pp. 877–905, 2008.
- [61] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [62] J. M. Bioucas-Dias and J. M. P. Nascimento, “Hyperspectral subspace identification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.
- [63] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota, “Clarke subgradients of stratifiable functions,” *SIAM Journal on Optimization*, vol. 18, no. 2, pp. 556–572, 2007.
- [64] K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep CNN denoiser prior for image restoration,” in the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2808–2817.
- [65] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, “Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1205–1218, 2019.
- [66] X.-L. Zhao, W.-H. Xu, T.-X. Jiang, Y. Wang, and M. K. Ng, “Deep plug-and-play prior for low-rank tensor completion,” *Neurocomputing*, to be published. doi: <https://doi.org/10.1016/j.neucom.2020.03.018>.



Ting-Zhu Huang received the B.S., M.S., and Ph.D. degrees in computational mathematics from the Department of Mathematics, Xi'an Jiaotong University, Xi'an, China. He is currently a Professor with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China.

His current research interests include scientific computation and applications, numerical algorithms for image processing, numerical linear algebra, preconditioning technologies, and matrix analysis with applications.

Dr. Huang is an Editor of the Scientific World Journal, Advances in Numerical Analysis, the Journal of Applied Mathematics, the Journal of Pure and Applied Mathematics: Advances in Applied Mathematics, and the Journal of Electronic Science and Technology, China.



Xi-Le Zhao 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 is currently a Professor with the School of Mathematical Sciences, UESTC.

His current research interests include image processing, computer vision, and machine learning.



Yong Chen received the B.S. degree in School of Science from East China University of Technology, Nanchang, China, in 2015. He is currently working toward the Ph.D. degree with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China. From 2018 to 2019, he was a research intern with the Geoinformatics unit, RIKEN Center for Advanced Intelligence Project, Japan.

His research interests include remote sensing image processing, sparse optimization, and low rank representation.



Wei He (S'14-M'17) received the B.S. degree in School of Mathematics and statistics and the Ph.D. degree in Surveying, Mapping and Remote Sensing (LIESMARS) from Wuhan University, Wuhan, China, in 2012 and 2017, respectively. He is currently a researcher with the Geoinformatics unit, RIKEN Center for Advanced Intelligence Project, Japan.

His research interests include image quality improvement, remote sensing image processing, matrix/tensor analysis and deep learning.



Yu-Bang Zheng received the B.S. degree in information and computing science from Anhui University of Finance and Economics, Bengbu, China, in 2017. He is currently working toward the Ph.D. degree in the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China. More information can be found in his homepage <https://yubangzheng.github.io>.

His current research interests include sparse and low-rank modeling for high-dimensional image processing problems.