Exploring Nonlocal Group Sparsity Under Transform Learning for Hyperspectral Image Denoising

Yong Chen, *Member, IEEE*, Wei He, *Senior Member, IEEE*, Xi-Le Zhao, *Member, IEEE*, Ting-Zhu Huang, Jinshan Zeng, and Hui Lin, *Senior Member, IEEE*

Abstract—Hyperspectral image (HSI) denoising has been regarded as an effective and economical preprocessing step in data subsequent applications. Recent nonlocal low-rank approximation on each full band patch group has demonstrated their superiority for HSI denoising. These methods, however, directly design low-rank regularization to the grouped patch image itself (i.e., original domain), which ignores the spatial information of the grouped patch image and cannot explore the potential structure. To address these issues, this article proposes a nonlocal group sparsifying transform learning (dubbed TLNLGS) method for HSI denoising. Motivated by the global spectral correlation in the HSI, we first impose a certain low-dimensional subspace hypothesis over the HSI to prevent the heavy computation burden with the spectral band increases, and then explore a discriminatively intrinsic nonlocal group sparse prior of the reduced image by the transform model. The learned group sparse prior can not only excavate the nonlocal self-similarity as recent nonlocal low-rank approximation methods but also preserve the local spatial smooth structure of the image. Moreover, compared with the fixed transform domain (e.g., gradient and discrete cosine transformation domains), the transform learning scheme can improve the sparse representation ability. An efficient block coordinate descent (BCD) algorithm is developed to solve the proposed model. Extensive experiments, including the simulated and real HSI datasets, indicate the superiority of the proposed TLNLGS method over the state-of-the-art HSI denoising approaches.

Yong Chen and Jinshan Zeng are with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang 330022, China (e-mail: chenyong1872008@163.com; jinshanzeng@jxnu.edu.cn).

Wei He is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: weihe1990@whu.edu.cn).

Xi-Le Zhao and Ting-Zhu Huang are with the School of Mathematical Sciences, Research Center for Image and Vision Computing, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China (e-mail: xlzhao122003@163.com; tingzhuhuang@126.com).

Hui Lin is with the Key Laboratory of Poyang Lake Wetland and Watershed Research of Ministry of Education, School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China (e-mail: huilin@cuhk.edu.hk).

Index Terms—Block coordinate descent (BCD), group sparsity, hyperspectral image (HSI) denoising, subspace representation, transform learning.

I. Introduction

HYPERSPECTRAL imaging adopts an imaging spectrometer to acquire 3-D cube data with hundreds of spectral bands covering the spectral region from ultraviolet to infrared wavelengths. Due to the abundant spectral information, hyperspectral image (HSI) has been extensively used in practical applications, such as unmixing [1], recognition [2], and classification [3]. However, influenced by the weather condition, sensor instability, calibration error, and physical mechanism, HSIs are usually polluted by various noise during the imaging process [4], [5], which destroys the image vision and constrains the advanced application tasks. Therefore, HSI denoising is regarded as a meaningful preprocessing step in remote sensing image processing.

In the past decades, numerous methods have been proposed for HSI denoising, which can be roughly divided into three categories: filtering-based [6], [7], model-based [8], [9], and learning-based approaches [10], [11], [12]. It is worth noting that the HSI denoising methods can also be classified from other perspectives. Rasti *et al.* [13] gave a review of the 2-D bandwise techniques to 3-D ones and varieties of low-rank methods. Later, Rasti *et al.* [14] again presented a study of the HSI denoising methods, which classifies the denoising methods as full-rank, low-rank, and deep learning (DL) techniques. In these categories, the model-based HSI denoising methods have received significant attention in recent years, which regard the restoration from noisy HSI as an ill-posed inverse problem and explore the prior information of HSI in the original domain or fixed domain to design a calculable regularization model. The popular and effective priors in HSI denoising can be generalized as local spatial smooth structure, global spectral correlation, and nonlocal self-similarity. The method of considering local spatial prior is mainly to expand the conventional grayscale image denoising and use the sparse model to HSI band-by-band, such as wavelet [15], sparse representation [16], spatial total variation (TV) [17], and spatio-spectral TV [18]. However, the global spectral correlation which presents a significant advantage for the denoising issue is ignored.

According to the linear mixture model [19], the global spectral correlation is hidden in the HSI. Based on this impor-
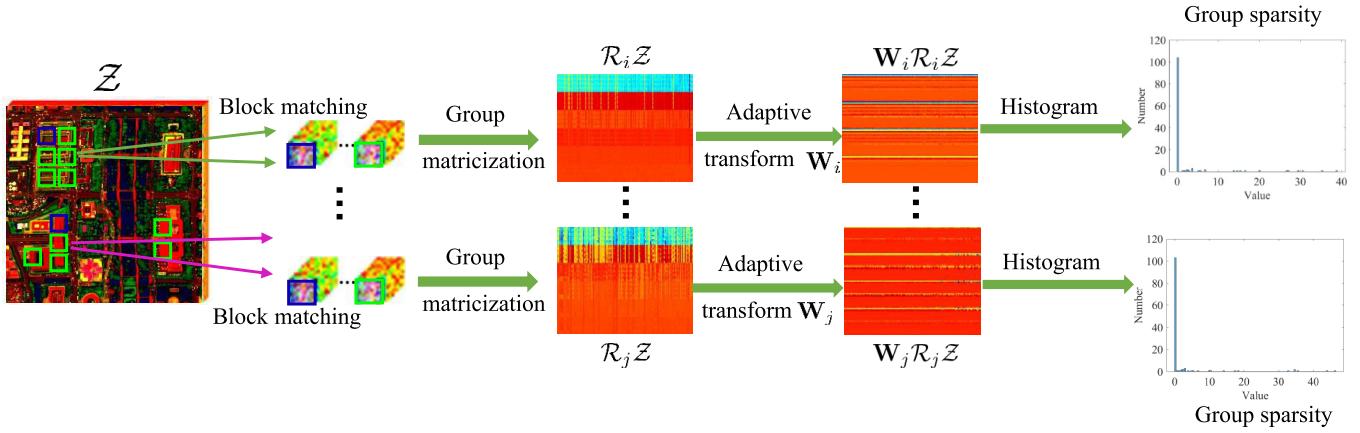
Fig. 1. Exploring nonlocal group sparsity under the transform learning scheme for the reduced image $\mathcal{Z}$.

tant prior, Zhang *et al.* [8] first proposed a classical low-rank matrix recovery (LRMR) for HSI denoising. Following the idea of the LRMR method, the nonconvex low-rank matrix approximation methods were proposed for HSI denoising [20], [21]. Furthermore, extending the low-rank matrix approximation, the low-rank tensor decompositions, including parallel factor analysis (PARAFAC) decomposition [4], Tucker decomposition [22], [23], and tensor singular value decomposition (t-SVD) [24], were introduced for HSI denoising. However, the spatial information is ignored in these methods. To capture the spatial smooth structure of HSI, He *et al.* [25] first introduced band-by-band TV regularization in the low-rank matrix recovery framework (LRTV) for HSI denoising. By extending the band-by-band TV and low-rank regularizations, the spatial–spectral TV regularization is the joint local low-rank regularization [26], low-rank Tucker decomposition [27], and t-SVD [28] for HSI denoising. The main drawback of these methods is that the spatial information of HSI is explored in the local domain, which may not achieve an optimal result.

The exploration of global spectral correlation can effectively preserve the spectral information of HSI, while the nonlocal self-similarity is also an important intrinsic characteristic in HSI [29], [30], [31]. Nonlocal self-similarity illustrates that there are many repeated local patches in HSI, and thus, a strong correlation exists in the nonlocal dimension by grouping the similar patches. To take full advantage of these two priors, Peng *et al.* [9] first proposed a tensor dictionary learning (TDL) to jointly model the spectral and nonlocal self-similarity of HSI. Later, different tensor models, including Kronecker-basis-representation (KBR)-based tensor sparsity measure [32], hyper-Laplacian regularized unidirectional low-rank tensor recovery (LLRT) [33], weighted low-rank tensor recovery (WLRTR) [34], and tensor ring decomposition [35], have been proposed to boost the performance of TDL. Although these methods achieve satisfactory denoising results, the computational burden is heavy as the number of spectra increases. To alleviate this problem, the subspace representation frame-

work, which was proposed for HSI subspace identification [36] and then was applied to HSI denoising [37], is used to project the HSI into a low-dimensional subspace and capture the spectral correlation [38]. The low-dimension projected factor is restored by sparse regularization [39], nonlocal self-similarity regularization [40], [41], [42], [43], and deep convolutional neural network (CNN) image prior [12], [43]. Especially, He *et al.* [41] proposed a unified subspace representation paradigm to integrate the spatial nonlocal and global spectral low-rank property simultaneously.

Although the existing nonlocal self-similarity-based HSI denoising methods that design appropriate low-rank approximation on each full band patch group have yielded good results, they still ignore some important problems. First, the low-rank structure is an external and universal characteristic, which cannot reveal the characteristic of transformation for the grouped image. Second, these methods generally ignore the local spatial smooth structure of the HSI. The reason is that they designed a regularization term separately for each prior knowledge. Thus, it needs to add more constraints to the model if simultaneously capturing these prior knowledges, which makes the model more complicated. Third, the existing methods preserve the local spatial smoothness of HSI by sparse modeling under certain transforms, such as gradient domain (i.e., TV regularization [25], [44]), wavelet domain [39], or redundant dictionaries [45]. However, the imaging area of HSI is relatively broad, which makes the spatial features and edges diverse. Thus, a certain domain cannot completely depict each local feature block in the whole image.

Mathematically, adaptive transform domains learned from the data itself can better preserve the image structure [46]. Moreover, several works have demonstrated that data-driven-based transform learning can improve the representation ability of local features and achieve promising results in various image processing [47], [48], [49], [50]. Inspired by the success of the transform learning framework, this article proposes a novel nonlocal group sparsifying transform learning (TLNLGS) method for HSI denoising, which can
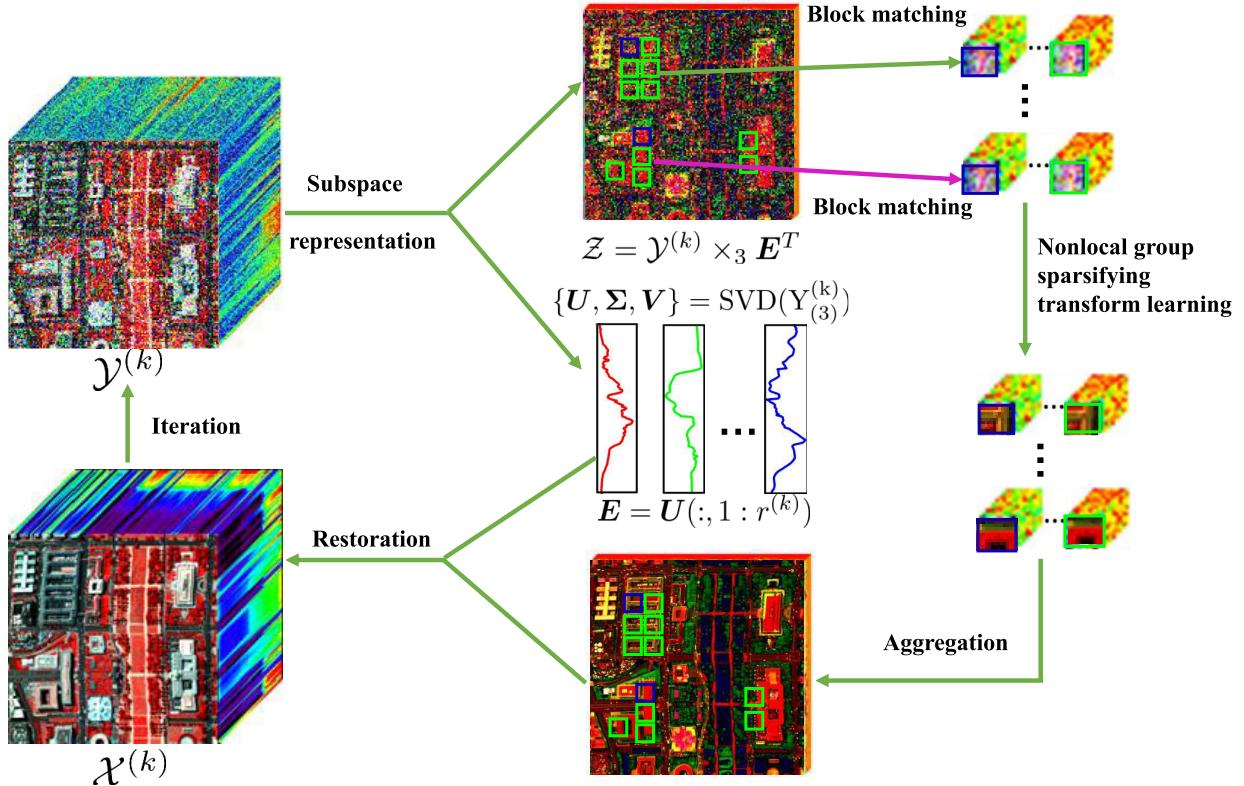
Fig. 2. Flowchart of subspace representation and TLNLGS method. It mainly includes three stages. (1) Global subspace learning. (2) Reduced image reconstruction by TLNLGS. (3) HSI restoration from the subspace and reduced image.

effectively explore the characteristic of transformation for the grouped patch image and take full advantage of the key prior knowledge of HSI. Different from the previous work of directly applying the nonlocal self-similarity on HSI itself, we assume that the HSI lies in a relatively low-dimensional subspace to preserve the global spectral correlation of HSI and prevent the heavy computation burden with the spectral band increase, which has been widely used for HSI denoising [40], [41], inpainting [38], compressive reconstruction [51], and superresolution [52], [53] issues. Moreover, instead of directly using the nonlocal low-rank approximation on the reduced image as in previous methods, we explore a discriminatively intrinsic nonlocal group sparse prior to the reduced image using the transform model (see Fig. 1), which can effectively exploit the nonlocal self-similarity and local spatial smooth structure of the HSI without introducing more constraints. Furthermore, the multiple transform domains are adaptively designed for different grouped images, which is more effective than the shared and fixed transform domains on diverse grouped images (see Tables V and VI). The flowchart of the proposed TLNLGS method is shown in Fig. 2. The main novelties and contributions are summarized as follows.

1) We explore a discriminatively intrinsic nonlocal group sparsity of the reduced image based on the transform learning scheme for HSI denoising, which can simultaneously capture the nonlocal self-similarity and local smooth structure, and get rid of more constraints caused by more prior knowledge of HSI.

2) We design multiple transform domains to model the different grouped images, which can improve the sparse representation ability of diverse spatial information compared with the fixed transform or shared transform domains.

3) We design an effective algorithm based on block coordinate descent (BCD) for solving the proposed model with convergence guarantee. The experiments on the simulated and real HSIs show that our method achieves better performances compared with the state-of-the-art approaches.

The rest of this article is organized as follows. The notation and problem formulation are introduced in Section II. Section III presents the proposed TLNLGS model. The optimization procedure is developed in Section IV. Section V illustrates the experimental results and discussions of our method. Finally, Section VI concludes this article.

II. NOTATION AND PROBLEM FORMULATION

A. Notation

Tensors, matrices, vectors, and scalars are represented by calligraphic letter (e.g., $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$), boldface capital letters (e.g., $\mathbf{X} \in \mathbb{R}^{I \times J}$), boldface lowercase letter (e.g., $\mathbf{x} \in \mathbb{R}^I$), and lowercase or uppercase letters (e.g., $i, I \in \mathbb{R}$), respectively. $\mathcal{X}(i_1, i_2, \ldots, i_n)$ or $x_{i_1, i_2, \ldots, i_n}$ denotes an element value of $\mathcal{X}$ in position $(i_1, i_2, \ldots, i_n)$. The mode-k matricization of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$ is represented as $\mathbf{X}_{(k)} \in \mathbb{R}^{I_k \times I_1, \ldots, I_{k-1} I_{k+1}, \ldots, I_n}$. In contrast, a tensor can be folded by the unfolding matrices $\mathbf{X}_{(k)}$ along the k-mode $\mathcal{X} = \text{fold}_k(\mathbf{X}_{(k)})$. The tube, row, and column fibers of a 3-D tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ are vector defined as $\mathbf{x}_{i_1, i_2, :}$, $\mathbf{x}_{i_1, :, i_3}$, and $\mathbf{x}_{:, i_2, i_3}$, respectively. Moreover,
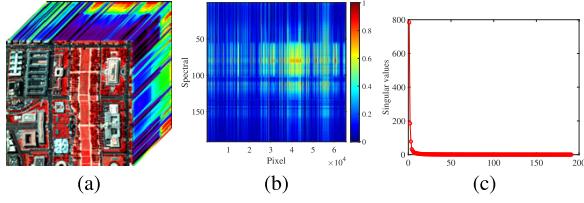
Fig. 3. (a) Washington dc Mall. (b) Unfolding matrix of (a) along spectral dimension. (c) Distribution of the singular values of unfolding matrix (b).

Fig. 4. Global spectral correlation of HSI via subspace representation.

the Frobenius norm of a tensor $\mathcal{X}$ is calculated as $\|\mathcal{X}\|_F = (\sum_{i_1,i_2,...,i_n} (x_{i_1,i_2,...,i_n})^2)^{(1/2)}$. The mode-$k$ multiplication of $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$ with a matrix $\mathbf{U} \in \mathbb{R}^{J \times I_k}$ is denoted by $\mathcal{X} \times_k \mathbf{U}$, and $(\mathcal{X} \times_k \mathbf{U})_{i_1,...,i_{k-1},j,i_{k+1},...,i_n} = \sum_{i_k} x_{i_1,i_2,...,i_n} u_{j,i_k}$.

B. Problem Formulation

Let $\mathcal{X} \in \mathbb{R}^{M \times N \times B}$ be a clean $M \times N$ HSI with B spectral bands, and $\mathcal{Y} \in \mathbb{R}^{M \times N \times B}$ be the observed HSI degraded by the Gaussian noise, then the degradation model between these two HSIs can be formulated as

$$\mathcal{Y} = \mathcal{X} + \mathcal{N} \tag{1}$$

where $\mathcal{N} \in \mathbb{R}^{M \times N \times B}$ denotes the Gaussian noise with zero mean and variance σ^2.

The problem of HSI denoising is to restore the clean HSI $\mathcal{X}$ from the noisy HSI $\mathcal{Y}$. However, directly estimating $\mathcal{X}$ from the degradation model (1) is an ill-posed problem, i.e., there exist multiple unexpected solutions. To solve this ill-posed problem, regularization is an effective method to explore the prior knowledge about the desired solution.

Based on the regularization method, the restoration of clean HSI $\mathcal{X}$ can be solved by the following regularization model:

$$\arg\min_{\mathcal{X}} \frac{1}{2}\|\mathcal{Y} - \mathcal{X}\|_F^2 + \lambda R(\mathcal{X}) \tag{2}$$

where the first term is the data fidelity term, $R(\mathcal{X})$ is the regularization term that explores the prior knowledge about the desired HSI, and λ is a regularization parameter used to balance the data fidelity term and regularization term.

III. PROPOSED TLNLGS METHOD

A. Global Spectral Correlation Via Subspace Representation

From the perspective of unmixing, each spectral signature can be approximately represented by a linear combination of a small number of pure spectral endmembers, and the number of endmembers is much less than the spectral band [19], which means that the global spectral correlation exists in the HSI. To further illustrate the correlation of pixel values in the spectral dimension, we present the distribution of singular values of $\mathbf{X}_{(3)}$ (unfolding matrix along spectral dimension) of the Washington dc Mall HSI in Fig. 3. From the curve, we can observe that the distributions of singular values of unfolding matrices decay rapidly, which again indicates the high correlation of HSI in the spectral dimension.

To capture the global spectral correlation of HSI, the convex matrix nuclear norm [8], [25], nonconvex matrix rank
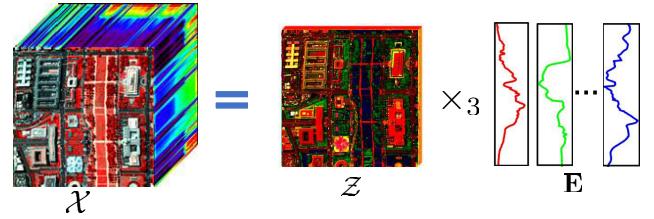
approximation [20], [21], and low-rank tensor decomposition [27], [35] are extensively applied to design the regularization $R(\mathcal{X})$ for HSI denoising. Recently, extending the representation of linear mixture model, i.e., HSI can be decomposed as the product of endmember and abundance matrices, subspace representation is widely presented to describe the global spectral correlation. Based on subspace representation, the clean HSI can be represented as

$$\mathcal{X} = \mathcal{Z} \times_3 \mathbf{E} \tag{3}$$

where $\mathbf{E} \in \mathbb{R}^{B \times r}$ ($r \ll B$) is the basis matrix with orthogonal columns, and $\mathcal{Z} \in \mathbb{R}^{M \times N \times r}$ is the reduced image maintained the similar characteristics with $\mathcal{X}$. Fig. 4 presents the global spectral correlation of HSI via subspace representation.

Using subspace decomposition, the clean HSI can be transformed into a low-dimensional subspace to estimate, which reduces the computation complexity. Moreover, subspace representation has been proven to be an effective method for handling HSI restoration [38], [41]. Based on the exploration of subspace representation of HSI, the estimation of HSI in the regularization model (2) can be formulated as

$$\arg\min_{\mathbf{E}, \mathcal{Z}} \frac{1}{2}\|\mathcal{Y} - \mathcal{Z} \times_3 \mathbf{E}\|_F^2 \tag{4}$$

which transforms the estimation of the original HSI into two representation factors.

B. Nonlocal Group Sparsity Prior of Reduced Image Under Transform Learning

Although the restoration model (4) effectively explores the spectral prior, it ignores the spatial prior information of HSI. Previous works have been demonstrated that simultaneously considering the spectral and spatial priors can achieve better results [38], [41]. Since the reduced image $\mathcal{Z}$ inherits the spatial characteristic of the original HSI as shown in Fig. 4, the regularization of spatial prior of the original HSI $\mathcal{X}$ can be transformed to regularize the reduced image $\mathcal{Z}$. Therefore, the denoising model (4) is extended to the following formulation:

$$\arg\min_{\mathbf{E}, \mathcal{Z}} \frac{1}{2}\|\mathcal{Y} - \mathcal{Z} \times_3 \mathbf{E}\|_F^2 + \lambda R(\mathcal{Z}) \tag{5}$$

where $R(\mathcal{Z})$ is the regularization term related to the spatial prior depiction of reduced image $\mathcal{Z}$.

TV regularization is usually designed to constrain the reduced image [54]. However, TV regularization simply explores the local spatial feature and ignores the nonlocal image block information. The spatial nonlocal self-similarity

prior of the reduced image is a powerful tool to maintain the spatial information, and the state-of-the-art denoising results have been achieved in [40], [41] through integrating such prior information. The principle of nonlocal similarity is that there are many repeated local patterns across the image. The step of grouping similar 3-D patches of reduced image $\mathcal{Z}$ is shown in the block matching part of Fig. 1. By grouping the similar patches together, the existing works mainly explored the correlation characteristic and used the nonlocal low-rank approximation to describe grouped image $\mathcal{R}_i \mathcal{Z}$ [40], [41].

The low-rank approximation regularization of the grouped image can achieve satisfactory denoising results under the subspace representation framework. However, the use of low-rank regularization alone ignores the local piecewise smooth structure of the grouped image. It is well-known that HSI contains the local smooth structure, such that the reduced image has a similar feature [25], [38]. Since the grouped image is extracted from the reduced image, certain smooth structure is also inherited in the grouped image. To compensate the local smooth structure of grouped image $\mathcal{R}_i \mathcal{Z}$, the low-rank approximation and sparse regularization in certain transforms or over dictionaries can be used to constrain $\mathcal{R}_i \mathcal{Z}$. However, the intuition behind this idea is to add two regularization constraints into the model, which makes the model more complex and difficult to optimize. Moreover, early works used the sparse prior in a fixed transform domain, e.g., gradient domain, wavelets' domain, and discrete cosine transform (DCT), to exploit the local smooth structure. Due to the fact that the spatial structures of HSI are diverse, a certain domain cannot completely depict the local feature in the whole image. Although the data-driven-based synthesis dictionary learning is the effective adaptive sparse excavation tool, the sparse coding step is an NP-hard problem [55]. Furthermore, this combination failed to reveal the characteristic of transformation of the grouped image since they directly designed regularizations on the original domain or fixed domain.

To overcome the aforementioned drawbacks, we explore the discriminatively intrinsic nonlocal group sparse regularization of the reduced image under the transform learning scheme. Fig. 1 illustrates in detail the process of exploring nonlocal group sparsity under the transform learning model for the reduced image $\mathcal{Z}$ extracted from the real Washington dc Mall dataset. Instead of directly designing the regularization to the image patch $\mathcal{V}_i \mathcal{Z} \in \mathbb{R}^{p \times p \times r}$, where $\mathcal{V}_i$ is a linear operator extracting the ith pixel overlapping cube patch of $\mathcal{Z}$, we construct the grouped image as $\mathcal{R}_i \mathcal{Z} \in \mathbb{R}^{p^2 r \times d}$. The operator $\mathcal{R}_i$ first divides the reduced image $\mathcal{Z} \in \mathbb{R}^{M \times N \times r}$ as an overlapped patch tensor $\mathcal{V}_i \mathcal{Z}$ with the size of $p \times p \times r$, then finds d similar patches for each ith key patch by block matching, and finally reshapes the nonlocal grouped patch as $\mathcal{R}_i \mathcal{Z}$, where the column dimension is the nonlocal self-similarity mode. Therefore, for a given transform $\mathbf{W}_i$ for each grouped image, the group sparse prior of the reduced image is explored in the transform domain. The horizontal axis is the ℓ_2-norm value of each row for $\mathbf{W}_i \mathcal{R}_i \mathcal{Z}$, and the vertical axis is the number of corresponding values in the statistical histogram. It is clear that most ℓ_2-norm values are zero or near zero, which indicates that there are many zero rows in

$\mathbf{W}_i \mathcal{R}_i \mathcal{Z}$. Therefore, the reduced image has the nonlocal group sparse prior under the transform model.

In reality, the grouped image may be polluted by some noise and even have some errors, leading to the deviation from desirable group sparsity constraint. Therefore, the solution of the data can be modeled as $\mathbf{W}_i \mathcal{R}_i \mathcal{Z} = \mathbf{G}_i + \mathbf{N}_i$, where $\mathbf{G}_i$ is the group sparse component, and $\mathbf{N}_i$ is the noise or error. Based on this observation, the regularization term $R(\mathcal{Z})$ can be designed as

$$R(\mathcal{Z}) = \sum_i \left(\frac{1}{2} \|\mathbf{W}_i \mathcal{R}_i \mathcal{Z} - \mathbf{G}_i\|_F^2 + \tau \|\mathbf{G}_i\|_{2,1} \right) \quad (6)$$

where τ is a regularization parameter, and $\ell_{2,1}$ is used to describe the group sparse prior of $\mathbf{G}_i$, which is defined as $\|\mathbf{R}\|_{2,1} = \sum_k \|\mathbf{R}(:, k)\|_2$.

It is worth noting that instead of using fixed transforms or synthesis dictionaries, we use a learning-based transform $\mathbf{W}_i$ to explore the group sparsity of $\mathcal{R}_i \mathcal{Z}$. Moreover, since the image scenes are diverse in whole HSI, different transforms $\mathbf{W}_i$ are adaptively obtained from each grouped image by group sparsifying transform learning. For the dimension of the group sparsifying transform $\mathbf{W}_i$, it usually can be overcomplete [48] or square [47]. In our work, following the strategy in [47], we set $\mathbf{W}_i$ as a square matrix and impose the orthogonal properties (i.e., $\mathbf{W}_i \in \mathbb{R}^{p^2 r \times p^2 r}$ and $\mathbf{W}_i^T \mathbf{W}_i = \mathbf{I}$). Recently, Xue *et al.* [56], [57] have applied the sparsity transform learning framework for multidimensional data recovery. However, they mainly learn a basis transform matrix to capture the low-rank property, i.e., the second dimension of the transform matrix is much smaller than the first dimension. In contrast, the proposed method learns a redundancy transform matrix to explore the potential structure of the grouped image.

C. Proposed Model

Based on the above-mentioned analysis, it is natural for us to incorporate the nonlocal group sparse term of reduced image (6) into the denoising model (5), and then the proposed model can be formulated as

$$\min_{\mathbf{E}, \mathcal{Z}, \mathbf{W}_i, \mathbf{G}_i} \frac{1}{2} \|\mathcal{Y} - \mathcal{Z} \times_3 \mathbf{E}\|_F^2$$
$$+ \lambda \sum_i \left\{ \frac{1}{2} \|\mathbf{W}_i \mathcal{R}_i \mathcal{Z} - \mathbf{G}_i\|_F^2 + \tau \|\mathbf{G}_i\|_{2,1} \right\}$$
$$\text{s.t. } \mathbf{E}^T \mathbf{E} = \mathbf{I}, \quad \mathbf{W}_i^T \mathbf{W}_i = \mathbf{I}. \quad (7)$$

The basic idea of the proposed model is that the nonlocal self-similarity and local smooth structure of the reduced image can be well-depicted by the group sparsity under the transform learning framework; meanwhile, the subspace representation captures the global spectral correlation of HSI. Under the unified framework of subspace representation, synthesis-based group sparse prior thoroughly excavates the potential characteristics of the reduced image, yielding better HSI denoising results.

Remark (The Group Sparsity of $\mathcal{R}_i \mathcal{Z}$ Under Transform Matrix $\mathbf{W}_i$): It is well-known that HSI contains the local smooth structure such that the reduced image $\mathcal{Z}$ has a similar feature. Since all the patches of the grouped image are

extracted from the reduced image, and the column of $\mathcal{R}_i \mathcal{Z}$ is stacked by vectoring the image patches, the grouped image $\mathcal{R}_i \mathcal{Z}$ also has a smooth structure along the column direction. Under the sparsifying transform matrix $\mathbf{W}_i$, the matrix $\mathcal{R}_i \mathcal{Z}$ is sparse. Moreover, benefiting from the nonlocal self-similarity property, the grouped image $\mathcal{R}_i \mathcal{Z}$ has a strong correlation, indicating that each column of $\mathcal{R}_i \mathcal{Z}$ can be linearly represented by a small number of identical columns in $\mathbf{W}_i^T$. Therefore, the representation coefficient $\mathbf{G}_i$ should have the property of group sparsity.

D. Difference Between TLNLGS and Related Methods

In this section, we discuss the difference between TLNLGS and related methods in detail.

1) Difference With Nonlocal Low-Rank-Based Methods: The representative global local factorization (GLF) [40] and nonlocal meets global (NGmeet) methods [41] used the subspace representation to capture the global spectral of HSI and then designed the nonlocal prior to reconstruct the reduced image. However, they are very different from our work. These two methods ignored the important spatial piecewise smoothness of HSI. Moreover, they designed the low-rank approximation on the grouped patch image in the original domain, failing to further reveal the characteristic of transformation. In contrast, our method explores a discriminatively intrinsic nonlocal group sparse prior to the reduced image by the transform learning scheme, which can effectively exploit the low-rank and local spatial smooth structure of HSI.

2) Difference With Group-Sparsity-Based Regularization Methods: The group sparse regularization was proposed in [58] for HSI denoising and then was improved in [59]. Recently, combining the low-rank prior, the group sparsity regularized low-rank approximation methods were proposed for HSI denoising [54]. Again, these methods directly used group sparse regularization to constrain the reduced image in the original domain, wavelet domain, discrete cosine domain, or fixed gradient domain. The proposed method regularizes the nonlocal grouped image of the reduced image in the multiple transform domains by group sparse regularization, which reveals the discriminatively intrinsic prior to the reduced image. The experimental results show that the adaptive multiple transform is indeed better than a fixed domain (see Table VI).

IV. ALGORITHM

It is generally difficult to directly solve the proposed model due to the coupling of these optimization variables. To address this, we are motivated to use the simple BCD algorithm framework for optimization. There are mainly four steps in each iteration: 1) global subspace $\mathbf{E}$ learning; 2) group sparsity $\mathbf{G}_i$ coding; 3) transform $\mathbf{W}_i$ learning; and 4) reduced image $\mathcal{Z}$ reconstruction. In the following, we present the optimization details of each step.

A. Global Subspace $\mathbf{E}$ Learning

For a fixed coupled variable $\mathcal{Z}$, the subproblem of global subspace $\mathbf{E}$ learning in model (7) is presented as

$$\min_{\mathbf{E}} \frac{1}{2}\|\mathcal{Y} - \mathcal{Z} \times_3 \mathbf{E}\|_F^2, \quad \text{s.t. } \mathbf{E}^T \mathbf{E} = \mathbf{I}. \tag{8}$$

According to [32], this subproblem has the closed-form solution of $\mathbf{E} = \mathbf{U}\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular vectors of $\mathbf{Y}_{(3)}\mathbf{Z}_{(3)}^T$, respectively. However, optimization is only suitable for the first iteration. The reason is that the iterative refinement strategy is used in our algorithm, and then the size of $\mathbf{E}$ is changed in the next iteration. Therefore, following [41], the orthogonal subspace $\mathbf{E}$ can be learned from singular value decomposition (SVD) of $\mathbf{Y}_{(3)}$, which is a simple operation.

B. Group Sparsity $\mathbf{G}_i$ Coding

Given the transform $\mathbf{W}_i$ and reduced image $\mathcal{Z}$, we solve the model (7) for group sparsity coding as

$$\min_{\mathbf{G}_i} \frac{1}{2}\|\mathbf{W}_i \mathcal{R}_i \mathcal{Z} - \mathbf{G}_i\|_F^2 + \tau \|\mathbf{G}_i\|_{2,1} \tag{9}$$

which is a standard group sparsity coding problem. Let $\mathbf{W}_i \mathcal{R}_i \mathcal{Z} = \mathbf{R}$, the closed-form solution of each row of $\mathbf{G}_i$ is computed by the following thresholding operator:

$$\mathbf{G}_i(j,:) = \begin{cases} \dfrac{\|\mathbf{R}(j,:)\|_2 - \tau}{\|\mathbf{R}(j,:)\|_2}\mathbf{R}(j,:), & \text{if } \|\mathbf{R}(j,:)\|_2 > \tau \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{10}$$

C. Transform $\mathbf{W}_i$ Learning

The subproblem of model (7) with respect to transform $\mathbf{W}_i$ can be learned by the following:

$$\min_{\mathbf{W}_i} \frac{1}{2}\|\mathbf{W}_i \mathcal{R}_i \mathcal{Z} - \mathbf{G}_i\|_F^2, \quad \text{s.t. } \mathbf{W}_i^T \mathbf{W}_i = \mathbf{I}. \tag{11}$$

With the orthogonal constraint, the closed-form solution of transform $\mathbf{W}_i$ can be achieved by $\mathbf{W}_i = \mathbf{U}\mathbf{V}^T$ [32], where $\mathcal{R}_i \mathcal{Z} \mathbf{G}_i^T = \mathbf{U}\mathbf{S}\mathbf{V}^T$ denotes the condensed SVD of $\mathcal{R}_i \mathcal{Z} \mathbf{G}_i^T$.

D. Reduced Image $\mathcal{Z}$ Reconstruction

With obtained $\mathbf{E}$, $\mathbf{G}_i$, and $\mathbf{W}_i$, the reduced image $\mathcal{Z}$ reconstruction problem reduces to

$$\min_{\mathcal{Z}} \frac{1}{2}\|\mathcal{Y} - \mathcal{Z} \times_3 \mathbf{E}\|_F^2 + \lambda \sum_i \frac{1}{2}\|\mathbf{W}_i \mathcal{R}_i \mathcal{Z} - \mathbf{G}_i\|_F^2. \tag{12}$$

Since the orthogonal matrix preserves the Frobenius norm, the following equation holds:

$$\|\mathbf{W}_i \mathcal{R}_i \mathcal{Z} - \mathbf{G}_i\|_F^2 = \left\|\mathcal{R}_i \mathcal{Z} - \mathbf{W}_i^T \mathbf{G}_i\right\|_F^2 \tag{13}$$

where $\mathbf{W}_i^T \mathbf{G}_i$ represents the restored patch via group sparsifying approximation.

Based on (13), problem (12) is equivalent to

$$\min_{\mathcal{Z}} \frac{1}{2}\|\mathcal{Y} - \mathcal{Z} \times_3 \mathbf{E}\|_F^2 + \lambda \sum_i \frac{1}{2}\left\|\mathcal{R}_i \mathcal{Z} - \mathbf{W}_i^T \mathbf{G}_i\right\|_F^2 \tag{14}$$

which is a least-square problem with the closed-form solution achieved by solving the normal linear equations' system

$$\left(\mathcal{I} + \lambda \sum_i \mathcal{R}_i^T \mathcal{R}_i\right)\mathcal{Z} = \left(\mathcal{Y} \times_3 \mathbf{E}^T\right) + \lambda \sum_i \mathcal{R}_i^T \left(\mathbf{W}_i^T \mathbf{G}_i\right). \tag{15}$$

Algorithm 1 TLNLGS-Based HSI Denoising

Require: Noisy HSI $\mathcal{Y}$, parameters λ, τ, $\gamma = 2$, and $\delta = 0.1$.
 1: Initialize: $\mathcal{X}^{(0)} = \mathcal{Y}^{(0)} = \mathcal{Y}$, initializing $\mathbf{W}_i$ using 2-D DCT, and estimating $r^{(0)}$ using HySime [36].
 2: **for** $k = 1 : K$ **do**
 3:   Update subspace $r^{(k)} = r^{(k-1)} + \gamma * (k - 1)$.
 4:   Calculate $\mathcal{Y}^{(k)} = \mathcal{X}^{(k-1)} + \delta(\mathcal{Y} - \mathcal{X}^{(k-1)})$.
 5:   Learn orthogonal subspace $\mathbf{E}$ and initialize reduce image $\mathcal{Z}$ via SVD on $\mathbf{Y}^{(k)}_{(3)}$ and $\text{fold}_3(\mathbf{E}^T \mathbf{Y}^{(k)}_{(3)})$, respectively.
 6:   Group sparsity $\mathbf{G}_i$ coding for all i via (10).
 7:   Transform $\mathbf{W}_i$ learning for all i via (11).
 8:   Reduced image $\mathcal{Z}$ reconstruction via (15).
 9:   Denoised HSI reconstruction $\mathcal{X}^{(k)} = \mathcal{Z} \times_3 \mathbf{E}$.
10: **end for**
Ensure: Denoised HSI $\mathcal{X}^{(k)}$.

Here, $\mathcal{R}_i^T : \mathbb{R}^{p^2 r \times d} \rightarrow \mathbb{R}^{M \times N \times r}$ represents the inverse operator of $\mathcal{R}_i$, which aggregates all the restored patches into the image with the size of $\mathbb{R}^{M \times N \times r}$. $\sum_i \mathcal{R}_i^T \mathcal{R}_i$ counts the number of the patches in all the $\mathcal{R}_i \mathcal{Z}$ that contain the pixel.

Based on the above iterations, we can obtain the entire algorithm for solving the model (7), as summarized in Algorithm 1. Since the rank (the number of column of $\mathbf{E}$) of globe subspace $\mathbf{E}$ results in a balance between the oversmooth (lower rank) and detail preservation (higher rank), we refine the rank of $\mathbf{E}$ on Step 3 in Algorithm 1 to improve the denoising results. Moreover, the iteration regularization on Step 4 in Algorithm 1 has also been demonstrated to boost the final results.

E. Convergence Analysis

In the following, we establish the convergence theorem of the proposed algorithm.

Theorem 1: Given any finite initialization, the sequence generated by Algorithm 1 converges to a critical point of (7).

Proof: Let $f(\mathbf{E}, \mathcal{Z}, \mathbf{W}_i, \mathbf{G}_i) = (1/2)\|\mathcal{Y} - \mathcal{Z} \times_3 \mathbf{E}\|_F^2 + \lambda \sum_i \{(1/2)\|\mathbf{W}_i \mathcal{R}_i \mathcal{Z} - \mathbf{G}_i\|_F^2 + \tau \|\mathbf{G}_i\|_{2,1}\}$ be the objective function in (7). We can easily verify that the objective f is block multiconvex, i.e., f is convex with respect to any block of variables while fixing the other blocks [60], and Lipschitz differentiable on any bounded set, i.e., ∇f is Lipschitz continuous. Note that the iterate for each block in Algorithm 1 follows the exact minimization update scheme, i.e., [60, (1.3a) in page 1759]. Moreover, we can show that f is a Kurdyka–Łojasiewicz function (see [61, Definition 3]) via following the similar analysis in the proof of [61, Proposition 2] and [62, Corollary 1(c)]. Thus, the global convergence of the suggested Algorithm 1 (i.e., convergence regardless of initial point) can be established by [60, Theorem 2.8].

F. Analysis of Computational Complexity

The computational complexity of the proposed TLNLGS is given in the following. We assume that the size of input HSI $\mathcal{Y}$ is $M \times N \times B$, the subspace is r, the patch size is $d \times d$, and the number of similar image patches is d. The orthogonal subspace $\mathbf{E}$ is learned via SVD on $\mathbf{Y}^{(k)}_{(3)}$, and the

complexity is $O(MNB^2)$. The computational complexity of group sparsity coding is $O(p^4 r^2 d)$, which is mainly from the cost of forming matrix $\mathbf{R}$. The transform update is first to form matrix $\mathcal{R}_i \mathcal{Z} \mathbf{G}_i^T$, and then use the SVD; thus, the computational complexity is $O(p^4 r^2 d + p^6 r^3)$. The reconstruction of reduced image $\mathcal{Z}$ is a simple division, whose computational complexity is $O(MNBb + Ip^4 r^2 d)$, where I is the number of grouped images. Therefore, the total computation complexity of Algorithm 1 is $O(MNB^2 + (p^4 r^2 d + p^6 r^3)I + MNBb + Ip^4 r^2 d)$. Although the number of grouped image is large, the group sparsity coding and transform update can be performed using parallel computing.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, several experiments are implemented to evaluate the effectiveness of the proposed TLNLGS on the simulated and real datasets. Eleven state-of-the-art HSI denoising methods have also been conducted for comparison, including low-rank tensor approximation (LRTA) method [63], TDL [9], low-rank tensor decomposition with TV regularization method (LRTDTV) [27], parameter-free method (HyRes) [39], fast HSI denoising method (FastHyDe) [38], GLF [40], KBR [32], LLRT [33], WLRTR [34], disentangled spatiospectral deep prior (DS2DP) method [12], and NGmeet [41]. Note that these methods involve low-rank matrix approximation-based, low-rank tensor decomposition-based, nonlocal self-similarity-based, and DL-based state-of-the-art methods for a comprehensive comparison. The codes of the compared methods can be downloaded from the authors' homepage, and the hyperparameters in different experiments are set based on either authors' codes or suggestions in the reference articles. The results are performed in MATLAB R2020b using a desktop of 32-GB RAM, with an Intel Core i9-10850K CPU at 3.60 GHz.

A. Experiments on HSI With Synthetic Noise

1) Experimental Setting: We choose two synthetic HSI datasets. One is the Washington dc Mall dataset (WDC,[1]) whose size is $256 \times 256 \times 191$, with complex texture and boundary information. The other one is the Pavia City Center dataset (PaC,[2]), and the size of PaC is $200 \times 200 \times 80$. To simulate the noisy HSI data, seven different types of Gaussian noises are added to the clean HSI data. The noise settings are additive Gaussian noise $N(\mathbf{0}, \sigma^2 \mathbf{I})$ with $\sigma = 10, 30, 80, 100$, and Gaussian noise $N(\mathbf{0}, \Sigma^2)$ where Σ is a diagonal matrix with diagonal elements sampled from a uniform distribution $U[10, 100]$ and $U[30, 80]$. The signal-to-noise ratios (SNRs) associated with different Gaussian noise cases on two datasets are listed in Table I. Moreover, to demonstrate that the proposed method can be extended to remove mixed noise, we add the mixtures of Gaussian noise, impulse noise, dead line noise, and stripe noise to simulate the noisy image. The standard deviation of the Gaussian noise and percentages of impulse noise are uniformly sampled within the

---

[1]https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html
[2]http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

TABLE I
SNR of Simulated Noisy HSI for the Gaussian Noise Cases

σ	2	10	30	80	100	[10, 100]	[30, 80]
WDC	29.307	15.327	5.785	-2.735	-4.673	[-4.673, 15.327]	[-2.735, 5.785]
PaC	29.225	15.246	5.704	-2.816	-4.754	[-4.754, 15.246]	[-2.816, 5.704]
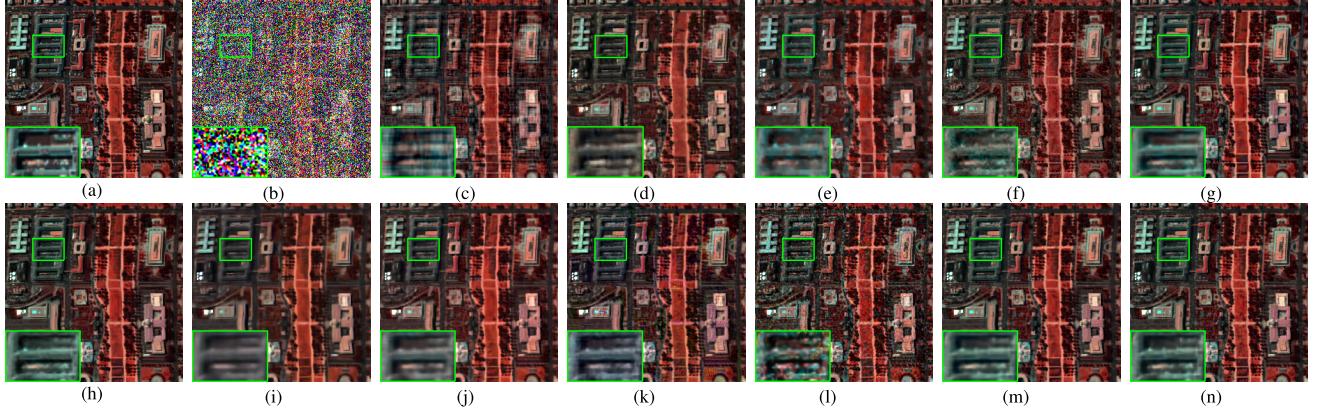
Fig. 5. Denoised results of the WDC dataset in the case $\sigma = 100$. The false color image is composed by bands (R: 108, G: 25, and B: 11). (a) Original. (b) Noisy. (c) LRTA. (d) TDL. (e) LRTDTV. (f) HyRes. (g) FastHyDe. (h) GLF. (i) KBR. (j) LLRT. (k) WLRTR. (l) DS2DP. (m) NGmeet. (n) TLNLGS.
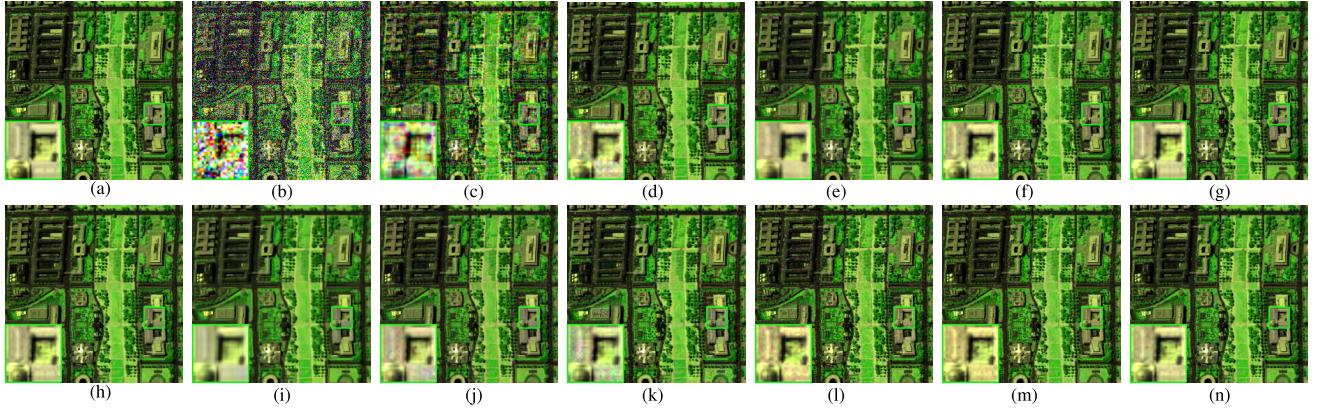
Fig. 6. Denoised results of the WDC dataset in the case $\sigma = [30, 80]$. The false color image is composed by bands (R: 109, G: 95, and B: 160). (a) Original. (b) Noisy. (c) LRTA. (d) TDL. (e) LRTDTV. (f) HyRes. (g) FastHyDe. (h) GLF. (i) KBR. (j) LLRT. (k) WLRTR. (l) DS2DP. (m) NGmeet. (n) TLNLGS.

range of $U[10, 100]$ and $U[0, 0.1]$, respectively. In addition, we severally select 10% of all the bands to add dead line and stripe noises. For the parameters' setting of our method, we will give in the discussion part.

2) Quantitative Indices: Five quantitative indices, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM) [64], feature similarity (FSIM) [65], erreur relative globale adimensionnelle de synthese (ERGAS) [66], and spectral angle mapper (SAM), are used to evaluate the denoising performance of different methods. PSNR, SSIM, and FSIM are mainly used to measure the spatial information, while ERGAS and SAM measure the ability of spectral information preservation. The better restoration results are reflected by larger PSNR, SSIM and, FSIM values, and smaller ERGAS and SAM values. Moreover, time cost in seconds is given.

3) Experimental Results on WDC Dataset: In terms of visual comparison, two representative denoised cases obtained

by different methods are presented. Figs. 5 and 6 present the restoration results of Gaussian noise with $\sigma = 100$ and Gaussian noise with $\sigma = [30, 80]$, respectively. Since LRTA and LRTDTV directly use LRTA for the original image, the performance is limited. Although the noise removal effects by the nonlocal-based methods TDL, KBR, LLRT, and WLRTR are improved, the restored spectral signatures are distorted. The HyRes method can eliminate obvious noise, but the detail is destroyed as presented in the enlarged box. Subspace representation with the nonlocal self-similarity-based methods FastHyDe, GLF, and NGmeet obtain satisfactory results compared with other methods, but the restoration results are slightly smoothed in $\sigma = 100$. The DL-based method DS2DP can eliminate most of the noises, but the noise removal is incomplete as shown in the enlarged box. On the contrary, by exploring the group sparsity on the nonlocal image, TLNLGS removes heavy Gaussian noises while preserving the

TABLE II

QUANTITATIVE RESULTS OF COMPARISON METHODS ON THE WDC DATASET UNDER DIFFERENT NOISE CASES

σ	Index	Noisy	LRTA	TDL	LRTDTV	HyRes	FastHyDe	GLF	KBR	LLRT	WLRTR	DS2DP	NGmeet	TLNLGS
2	PSNR	42.11	46.50	50.44	46.38	50.46	50.41	49.93	45.44	46.94	47.69	44.09	50.09	50.52
	SSIM	0.987	0.994	0.998	0.996	0.998	0.998	0.998	0.996	0.996	0.996	0.995	0.998	0.998
	FSIM	0.994	0.997	0.999	0.998	0.999	0.999	0.999	0.998	0.998	0.998	0.997	0.999	0.999
	ERGAS	33.13	21.39	13.53	20.34	14.24	13.87	15.41	22.21	18.96	18.98	26.90	13.91	13.21
	SAM	3.868	2.559	1.428	1.944	1.565	1.455	1.565	1.693	1.996	1.857	2.300	1.506	1.363
	Time		1.8	24.5	181.4	3.6	19.7	180.2	2727.4	1864.8	3170.3	6813.9	104.8	331.7
10	PSNR	28.13	38.61	41.90	38.40	42.28	41.40	41.61	40.60	41.99	42.61	40.28	42.44	42.90
	SSIM	0.781	0.973	0.989	0.977	0.989	0.989	0.990	0.986	0.990	0.990	0.984	0.991	0.992
	FSIM	0.903	0.986	0.993	0.988	0.994	0.993	0.993	0.992	0.994	0.994	0.991	0.994	0.995
	ERGAS	165.67	49.30	34.32	50.64	33.02	38.10	37.44	38.47	35.94	31.76	40.94	32.83	30.13
	SAM	16.408	5.559	3.396	4.482	3.408	3.878	3.800	2.944	3.438	2.918	3.953	3.208	2.709
	Time		4.8	24.5	163.2	2.8	4.6	88.8	2878.8	4237.7	3169.8	6834.6	104.6	168.7
30	PSNR	18.59	32.75	35.07	34.44	36.13	37.19	37.47	34.83	36.33	36.99	34.51	37.76	37.84
	SSIM	0.368	0.916	0.954	0.943	0.961	0.971	0.973	0.953	0.967	0.970	0.945	0.974	0.975
	FSIM	0.693	0.951	0.974	0.968	0.978	0.982	0.984	0.972	0.981	0.982	0.968	0.984	0.985
	ERGAS	497.00	94.37	72.92	77.75	65.15	57.86	55.99	73.84	65.89	58.26	77.66	53.42	52.91
	SAM	36.442	8.092	6.119	6.149	6.205	5.229	4.991	4.968	5.169	4.713	6.424	4.400	4.297
	Time		9.0	23.6	163.4	2.7	4.8	75.0	3000.2	4173.4	3154.3	6855.6	103.6	179.6
80	PSNR	10.07	27.81	29.38	29.94	30.45	32.29	32.52	29.43	31.04	31.30	28.73	32.32	32.71
	SSIM	0.093	0.782	0.850	0.861	0.880	0.920	0.923	0.845	0.894	0.899	0.833	0.918	0.926
	FSIM	0.445	0.883	0.918	0.919	0.934	0.952	0.954	0.908	0.940	0.942	0.911	0.953	0.957
	ERGAS	1324.58	166.36	138.93	130.20	123.18	99.54	96.76	137.45	116.10	110.80	149.42	98.85	94.50
	SAM	59.512	10.596	8.842	8.692	9.602	7.413	7.109	7.963	7.104	7.899	9.488	6.732	6.577
	Time		6.5	22.4	191.7	2.6	5.0	66.7	5328.2	4146.3	3161.9	6855.4	99.7	108.1
100	PSNR	8.13	26.71	27.73	28.86	29.34	31.25	31.32	28.45	29.94	29.77	27.79	31.16	31.56
	SSIM	0.062	0.733	0.799	0.828	0.848	0.901	0.901	0.810	0.864	0.860	0.801	0.895	0.905
	FSIM	0.391	0.861	0.893	0.902	0.918	0.941	0.942	0.886	0.923	0.923	0.894	0.941	0.945
	ERGAS	1656.79	188.76	167.63	147.15	139.68	112.02	111.00	153.98	131.28	131.77	167.12	112.80	108.04
	SAM	64.289	11.249	9.852	9.317	10.086	8.055	7.758	8.695	7.632	9.058	10.226	7.274	7.089
	Time		3.4	20.7	189.9	2.7	5.1	71.8	4492.5	4116.9	3217.9	6833.5	117.7	106.6
[10, 100]	PSNR	14.06	21.91	22.14	30.57	35.45	37.17	36.82	30.77	29.37	29.96	31.17	30.57	36.62
	SSIM	0.229	0.492	0.511	0.866	0.955	0.972	0.970	0.884	0.820	0.835	0.896	0.883	0.972
	FSIM	0.558	0.747	0.776	0.914	0.974	0.983	0.981	0.931	0.917	0.915	0.940	0.937	0.982
	ERGAS	1063.11	387.18	420.51	121.15	69.99	57.84	60.01	119.64	146.85	184.36	116.07	171.50	61.05
	SAM	54.376	29.758	31.572	7.106	6.556	5.253	5.297	7.982	12.445	15.944	8.701	13.405	5.009
	Time		9.8	28.0	222.8	2.7	4.9	72.4	5522.0	4320.2	3518.4	6834.5	118.6	108.9
[30, 80]	PSNR	13.73	23.13	29.53	31.63	33.26	34.96	35.04	31.31	31.38	32.58	30.70	31.41	34.86
	SSIM	0.190	0.560	0.844	0.894	0.931	0.954	0.955	0.898	0.892	0.919	0.882	0.897	0.955
	FSIM	0.553	0.781	0.928	0.941	0.961	0.972	0.973	0.939	0.944	0.954	0.934	0.945	0.973
	ERGAS	918.95	300.97	141.31	112.62	89.84	73.70	73.12	110.84	112.18	98.51	121.44	132.34	74.13
	SAM	51.284	24.632	11.121	8.095	7.656	6.179	5.889	6.982	8.883	7.756	9.289	12.198	5.584
	Time		9.0	24.4	190.1	2.7	4.9	74.6	5589.1	2467.9	3160.4	6854.5	118.7	107.2
mixed	PSNR	12.02	20.91	21.38	29.30	24.64	24.97	24.93	23.73	23.47	24.27	31.58	24.25	29.60
	SSIM	0.024	0.476	0.500	0.829	0.795	0.813	0.805	0.719	0.723	0.758	0.909	0.771	0.860
	FSIM	0.521	0.770	0.809	0.886	0.933	0.941	0.935	0.877	0.893	0.907	0.949	0.881	0.930
	ERGAS	939.24	421.74	408.53	140.55	307.17	306.50	305.94	322.66	303.36	318.12	108.93	239.49	169.14
	SAM	45.891	22.021	21.147	7.568	14.474	14.440	14.452	15.086	13.462	15.149	8.480	10.427	13.936
	Time		12.4	30.8	185.8	2.7	5.7	74.7	5489.2	1973.4	3834.0	6873.7	13.9	113.9
average	PSNR	18.36	29.79	32.20	33.69	35.25	36.21	36.21	33.07	33.81	34.40	33.61	35.00	37.08
	SSIM	0.342	0.741	0.806	0.899	0.920	0.940	0.939	0.886	0.893	0.903	0.906	0.916	0.948
	FSIM	0.632	0.872	0.911	0.940	0.961	0.970	0.970	0.938	0.949	0.952	0.948	0.954	0.971
	ERGAS	824.81	203.76	174.71	100.05	105.28	94.93	94.46	122.39	116.32	119.07	101.06	106.89	75.39
	SAM	41.509	14.308	11.685	6.669	7.444	6.488	6.358	7.039	7.516	8.162	7.358	7.394	5.821
	Time		7.1	24.9	186.0	2.8	6.8	88.0	4378.4	3412.6	3298.4	6844.5	97.7	153.1

image details. Using data whitening, HyRes, FastHyDe, GLF, and TLNLGS achieve better denoising results in the case of $\sigma = [30, 80]$.

In terms of quantitative comparison, the results of different methods on the WDC datasets under different noise cases are listed in Table II. We highlight the top two algorithms using two different colors (red and blue). From the results, our method considerably outperforms the compared methods in most indices and noise cases, which demonstrates that our method achieves a superior performance of preserving the image details while removing noise from the noisy images.

DS2DP obtains the best result in the case of mixed noise since it is mainly designed for mixed noise removal. Specifically, compared with the outstanding noise removal methods GLF and NGmeet, the proposed method achieves a substantial improvement in the cases of mixed noise, indicating the potency of exploring the nonlocal group sparsity. For computational efficiency, the LRTA, HyRes, and FastHyDe methods are obviously superior to other methods. However, compared with the nonlocal-based methods (KBR, LLRT, and WLRTR) and the DL-based approach (DS2DP), the time cost of the proposed method is advantageous.
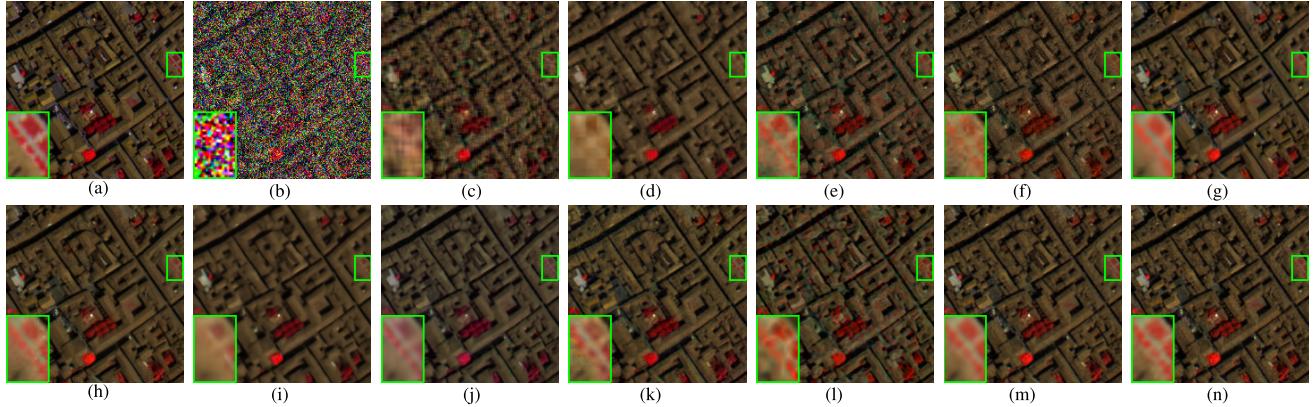
Fig. 7. Denoised results of the PaC dataset in the case $\sigma = 100$. The false color image is composed by bands (R: 79, G: 40, and B: 2). (a) Original. (b) Noisy. (c) LRTA. (d) TDL. (e) LRTDTV. (f) HyRes. (g) FastHyDe. (h) GLF. (i) KBR. (j) LLRT. (k) WLRTR. (l) DS2DP. (m) NGmeet. (n) TLNLGS.
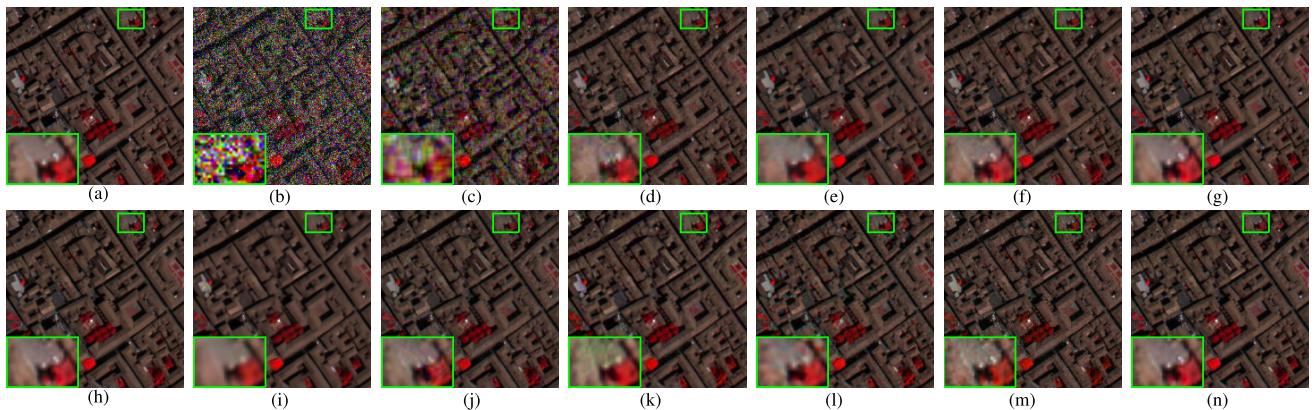
Fig. 8. Denoised results of the PaC dataset in the case $\sigma = [30, 80]$. The false color image is composed by bands (R: 78, G: 35, and B: 21). (a) Original. (b) Noisy. (c) LRTA. (d) TDL. (e) LRTDTV. (f) HyRes. (g) FastHyDe. (h) GLF. (i) KBR. (j) LLRT. (k) WLRTR. (l) DS2DP. (m) NGmeet. (n) TLNLGS.

4) Experimental Results on PaC Dataset: Figs. 7 and 8 show some visualization results of our method in comparison to other existing HSI denoising methods for removing noise in the PaC dataset. LRTA and TDL cannot completely remove the noises and obtain artifacts. From the enlarged results shown on the left-bottom, it can be seen that the textural details are very well restored with NGmeet and TLNLGS, while most approaches suffer from serious missing image details and incomplete noise removal. FastHyDe and GLF effectively remove noise, but with blurred details. Although NGmeet obtains satisfactory result in Fig. 7, incomplete noise removal also exists in Fig. 8. On the whole, our method can use the potential structure sparsity as a prior to remove the noise and preserve the image details.

Table III shows quantitative metrics about ours and other methods under different noise intensities on the PaC dataset. It can be observed that the proposed TLNLGS outperforms the comparison methods in most cases, which is similar to the WDC dataset. Moreover, from the result of averaging different noise cases, the proposed method also achieves better quantitative results. For the running time, HyRes and LRTA cost shorter times than other methods. The proposed method has a tradeoff between the performance and execution time.
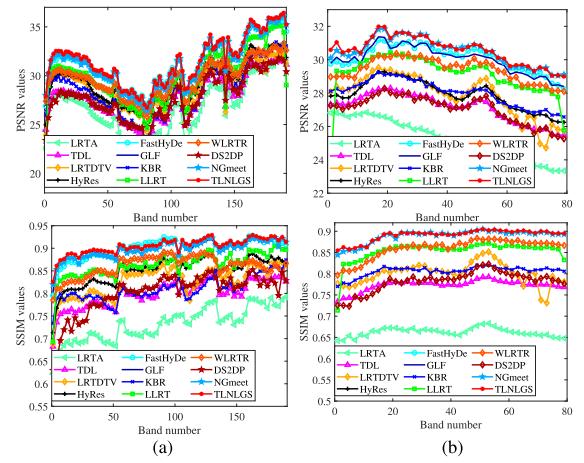
Fig. 9. PSNR and SSIM values of each band in the case of $\sigma = 100$ on the WDC and PaC datasets. (a) WDC dataset. (b) PaC dataset.

Fig. 9 presents the PSNR and SSIM values of each band in the case of $\sigma = 100$ on the WDC and PaC datasets. Clearly, the PSNR and SSIM values of LRTA are lower than one of the other denoising methods since they ignore the spatial prior information. In contrast, the proposed method achieves

TABLE III

QUANTITATIVE RESULTS OF COMPARISON METHODS ON THE PaC DATASET UNDER DIFFERENT NOISE CASES

σ	Index	Noisy	LRTA	TDL	LRTDTV	HyRes	FastHyDe	GLF	KBR	LLRT	WLRTR	DS2DP	NGmeet	TLNLGS
2	PSNR	42.10	44.90	49.02	46.37	49.41	49.52	49.77	46.38	49.68	47.50	45.06	49.80	49.99
	SSIM	0.988	0.994	0.998	0.996	0.998	0.998	0.998	0.996	0.998	0.997	0.996	0.998	0.998
	FSIM	0.993	0.996	0.999	0.998	0.999	0.999	0.999	0.998	0.999	0.998	0.997	0.999	0.999
	ERGAS	28.96	21.08	13.25	17.91	12.73	12.57	12.24	17.79	12.36	15.90	21.25	12.15	11.84
	SAM	3.901	2.839	1.644	2.065	1.648	1.603	1.568	1.767	1.518	1.777	2.116	1.527	1.462
	Time		1.5	6.4	44.2	1.2	5.4	64.7	770.3	508.9	966.2	3894.9	38.9	127.2
10	PSNR	28.12	38.19	41.54	39.82	41.42	42.17	42.92	40.06	42.27	41.99	38.84	43.29	43.33
	SSIM	0.799	0.973	0.988	0.984	0.988	0.990	0.991	0.984	0.990	0.990	0.980	0.992	0.992
	FSIM	0.899	0.984	0.993	0.991	0.992	0.994	0.995	0.991	0.994	0.994	0.987	0.995	0.995
	ERGAS	144.80	46.02	31.32	37.80	31.79	29.06	26.66	36.76	32.23	29.96	42.73	25.50	25.45
	SAM	16.089	5.158	3.186	3.580	3.356	2.989	2.596	2.881	2.867	2.732	4.230	2.497	2.363
	Time		1.6	5.8	48.0	1.0	2.4	45.9	740.0	503.2	1038.4	3895.1	46.7	43.6
30	PSNR	18.58	32.12	34.54	34.45	34.69	35.99	36.63	34.35	35.76	35.98	32.91	37.04	37.24
	SSIM	0.368	0.910	0.949	0.943	0.948	0.962	0.968	0.947	0.962	0.964	0.923	0.971	0.972
	FSIM	0.683	0.936	0.968	0.966	0.967	0.974	0.979	0.967	0.976	0.977	0.948	0.981	0.982
	ERGAS	434.24	91.26	69.24	70.39	68.65	58.95	54.26	70.51	68.44	59.46	84.16	51.75	50.61
	SAM	34.908	7.903	5.382	5.994	5.903	5.017	4.247	4.325	4.887	4.605	7.549	4.086	3.612
	Time		2.4	5.0	50.8	0.7	2.5	41.7	871.4	495.3	1040.8	3905.0	47.5	51.1
80	PSNR	10.07	26.20	28.60	28.78	28.96	31.19	31.29	29.12	30.23	30.61	27.78	31.71	31.86
	SSIM	0.084	0.719	0.830	0.833	0.842	0.900	0.901	0.842	0.884	0.892	0.803	0.912	0.913
	FSIM	0.430	0.829	0.893	0.900	0.904	0.933	0.935	0.898	0.927	0.927	0.883	0.942	0.946
	ERGAS	1156.86	178.75	136.32	138.13	131.74	101.73	100.02	128.33	117.60	109.13	149.66	95.32	93.58
	SAM	58.334	10.500	7.646	10.958	8.257	7.493	6.334	6.305	6.465	7.737	8.172	6.156	5.209
	Time		1.8	4.9	53.5	0.6	2.6	38.3	1411.4	492.2	1058.9	3911.3	44.3	35.0
100	PSNR	8.13	25.25	27.12	27.76	27.91	30.10	30.16	28.04	29.09	29.39	27.05	30.60	30.73
	SSIM	0.054	0.662	0.773	0.795	0.806	0.878	0.877	0.807	0.854	0.860	0.782	0.891	0.894
	FSIM	0.376	0.805	0.859	0.880	0.886	0.918	0.919	0.873	0.907	0.909	0.871	0.929	0.934
	ERGAS	1446.90	199.30	161.30	157.36	147.83	114.94	113.83	145.16	132.58	125.00	162.86	108.20	106.45
	SAM	63.333	10.341	7.823	12.162	8.586	8.007	6.708	6.902	6.806	8.578	10.221	6.521	5.710
	Time		1.3	4.0	52.7	0.6	2.5	38.3	1304.9	491.1	1066.6	3887.6	57.0	33.9
[10, 100]	PSNR	14.32	21.79	22.48	29.73	34.41	36.07	36.10	29.86	28.17	28.99	29.90	27.15	36.82
	SSIM	0.235	0.488	0.531	0.828	0.945	0.963	0.965	0.868	0.789	0.818	0.873	0.780	0.970
	FSIM	0.551	0.730	0.772	0.899	0.964	0.975	0.977	0.914	0.894	0.895	0.926	0.877	0.981
	ERGAS	924.49	343.05	365.27	207.53	70.73	58.29	57.64	118.53	150.78	210.03	119.53	274.57	53.17
	SAM	52.845	27.883	29.173	20.248	6.117	5.041	4.602	6.730	13.678	18.620	7.572	23.258	3.939
	Time		3.0	6.5	55.3	0.7	2.5	40.0	1728.7	512.9	1079.2	3899.2	57.5	50.4
[30, 80]	PSNR	13.82	23.43	26.89	31.08	31.96	33.68	34.03	30.73	31.37	31.17	30.00	28.68	34.65
	SSIM	0.188	0.586	0.740	0.887	0.910	0.940	0.944	0.888	0.898	0.901	0.873	0.823	0.952
	FSIM	0.541	0.773	0.879	0.930	0.943	0.959	0.964	0.928	0.939	0.935	0.917	0.901	0.969
	ERGAS	806.84	258.98	177.34	102.29	93.25	76.41	72.97	107.09	105.33	105.13	116.80	204.16	67.89
	SAM	49.622	22.499	15.862	6.763	6.969	6.111	5.127	5.888	8.361	8.167	9.650	18.908	4.670
	Time		2.6	6.7	49.1	0.7	2.5	38.6	1614.6	582.3	1061.5	3903.5	57.0	49.5
mixed	PSNR	14.03	22.01	22.49	27.31	25.82	26.31	26.38	24.80	24.78	25.09	30.31	24.25	27.87
	SSIM	0.204	0.530	0.554	0.756	0.829	0.858	0.854	0.766	0.779	0.784	0.891	0.771	0.814
	FSIM	0.582	0.769	0.807	0.826	0.921	0.936	0.932	0.873	0.887	0.892	0.932	0.881	0.895
	ERGAS	810.88	318.17	310.42	158.62	218.32	217.03	214.65	234.40	223.87	235.62	115.19	239.49	208.16
	SAM	43.084	17.117	16.825	6.264	10.031	10.135	9.870	10.310	8.331	10.771	6.957	10.427	15.574
	Time		3.2	6.7	48.1	1.1	2.4	41.5	1569.1	567.3	1032.0	3866.8	13.9	42.8
average	PSNR	18.65	29.24	31.59	33.16	34.32	35.63	35.91	32.92	33.92	33.84	32.73	34.07	36.56
	SSIM	0.365	0.733	0.795	0.878	0.908	0.936	0.937	0.887	0.894	0.901	0.890	0.892	0.938
	FSIM	0.632	0.853	0.896	0.924	0.947	0.961	0.962	0.930	0.940	0.941	0.933	0.938	0.963
	ERGAS	719.25	182.08	158.06	111.25	96.88	83.62	81.53	107.32	105.40	111.28	101.52	126.39	77.14
	SAM	40.265	13.030	10.943	8.504	6.358	5.800	5.132	5.639	6.614	7.873	7.058	9.173	5.317
	Time		2.2	5.8	50.2	0.8	2.9	43.6	1251.3	519.2	1043.0	3895.4	45.4	54.2

better or more competitive results compared with the existing approaches, indicating the superiority of our method.

B. Experiments on HSI With Real Noise

We also illustrate the effectiveness of the proposed method by comparing it against other approaches on two HSIs with real noise. As real-world noisy HSIs generally miss the associated noise-free ground truth, we only visually and qualitatively evaluate the denoising performance on real data.

1) Real Noisy HSI Datasets: Two representative real-world data are selected for real noisy HSI datasets, including the airborne visible/infrared imaging spectrometer (AVIRIS) Indian

Pines dataset and the GaoFen-5 (GF-5) dataset. The spatial size of the Indian Pines dataset is 145×145, and the spectral dimension is 220. Due to the influence of atmosphere (water absorption), the Indian Pines dataset is seriously degraded by noise [as shown in Fig. 10(a)]. The spectral band of the original GF-5 is 330, of which some bands are missing information. Therefore, a sub-HSI with the size of $256 \times 256 \times 155$ is chosen for experiments. The GF-5 dataset is degraded by mixed noise as shown in Fig. 12(a).

2) Experimental Results on Indian Pines Dataset: Fig. 10 shows the false color restoration results of the real Indian Pines dataset. As shown in Fig. 10(a), noise seriously influences
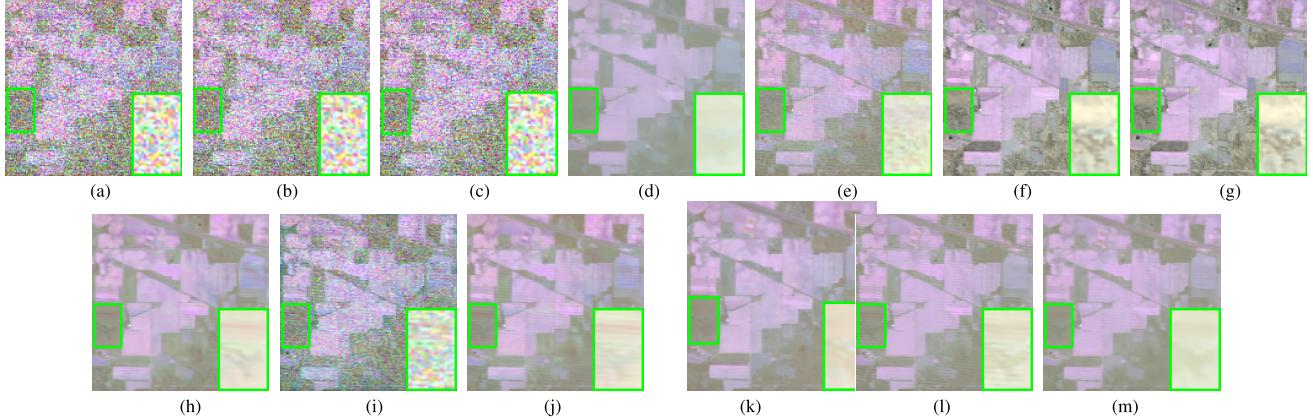
Fig. 10. Denoised results of the real Indian Pines dataset. The false color image is composed by bands (R: 220, G: 150, and B: 109). (a) Noisy. (b) LRTA. (c) TDL. (d) LRTDTV. (e) HyRes. (f) FastHyDe. (g) GLF. (h) KBR. (i) LLRT. (j) WLRTR. (k) DS2DP. (l) NGmeet. (m) TLNLGS.
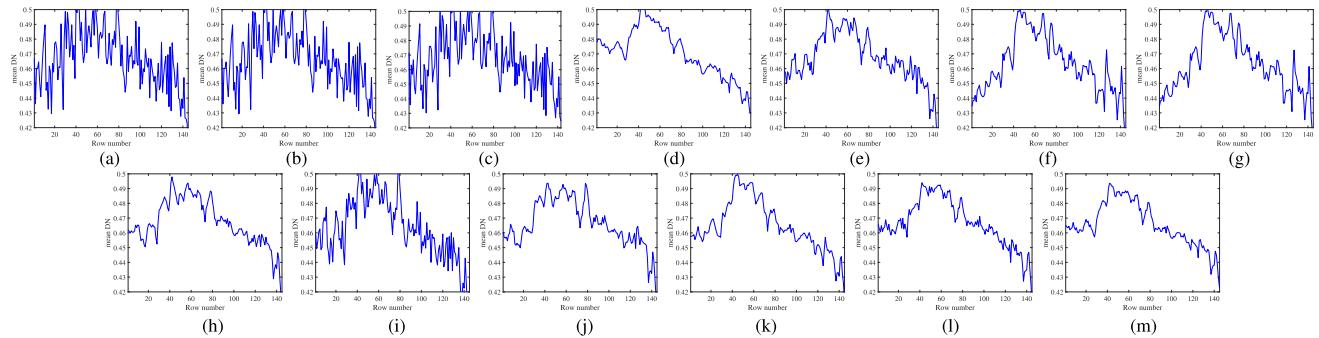
Fig. 11. Horizontal mean profiles of band 150 for the real Indian Pines dataset. (a) Noisy. (b) LRTA. (c) TDL. (d) LRTDTV. (e) HyRes. (f) FastHyDe. (g) GLF. (h) KBR. (i) LLRT. (j) WLRTR. (k) DS2DP. (l) NGmeet. (m) TLNLGS.
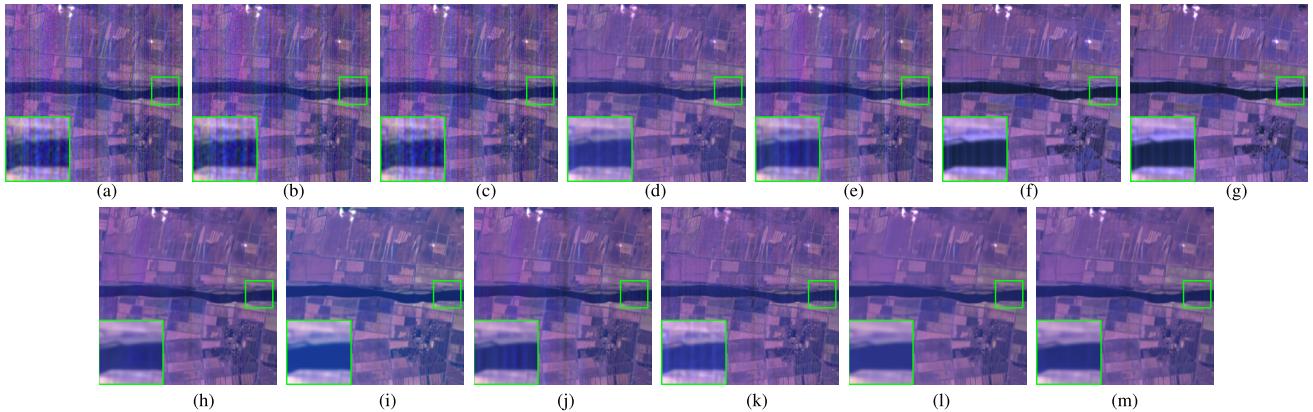
Fig. 12. Denoised results of the real GF-5 dataset. The false color image is composed by bands (R: 96, G: 151, and B: 154). (a) Noisy. (b) LRTA. (c) TDL. (d) LRTDTV. (e) HyRes. (f) FastHyDe. (g) GLF. (h) KBR. (i) LLRT. (j) WLRTR. (k) DS2DP. (l) NGmeet. (m) TLNLGS.

the image quality. LRTA, TDL, and LLRT can obtain satisfactory denoising results in simulated noise, but they fail to remove high-intensity noise in real data. Although LRTDTV removes the noise thoroughly by the combination of TV regularization and LRTA, the restored result is oversmoothed. HyRes, FastHyDe, and GLF can remove obvious noise. KBR, WLRTR, and NGmeet effectively remove the noise and preserve most of the image details. However, since they ignore the sparse prior of the nonlocal image, some stripe artifacts remain in the denoised results as shown in the enlarged box. DS2DP and our TLNLGS eliminate all noises as well as achieve better visual quality.

Fig. 11 presents the mean profiles of band 150. From the results, there are several impulses in the results of LRTA, TDL, HyRes, FastHyDe, GLF, and LLRT due to incomplete noise removal. Since the results of KBR, WLRTR, and NGmeet generate stripe artifacts, the mean profiles occur with tiny fluctuations. Moreover, the result of LRTDTV is oversmoothed. In contrast, DS2DP and the proposed method yield better illustration.

3) Experimental Results on GF-5 Dataset: Fig. 12 shows the denoising results on the real GF-5 dataset. As shown in the green enlarged close-ups, the results achieved by LRTA, TDL, and HyRes contain visible noise. LRTDTV, FastHyDe,
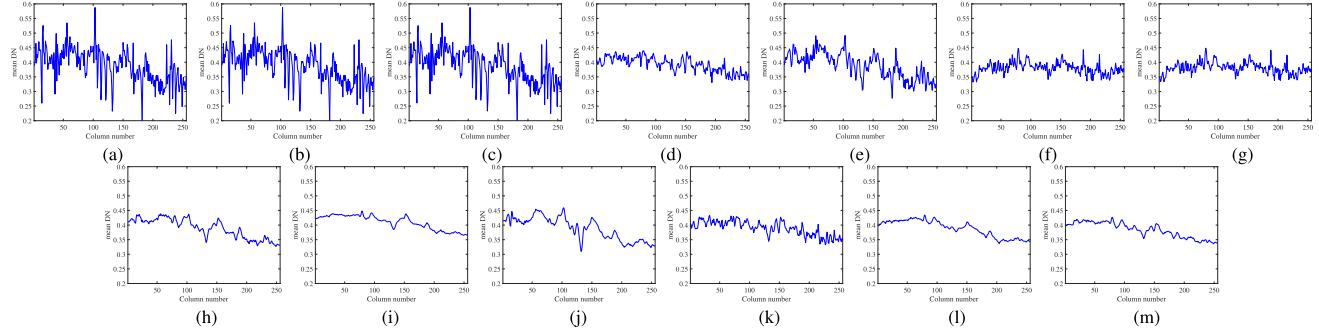
Fig. 13. Vertical mean profiles of band 154 for the real GF-5 dataset. (a) Noisy. (b) LRTA. (c) TDL. (d) LRTDTV. (e) HyRes. (f) FastHyDe. (g) GLF. (h) KBR. (i) LLRT. (j) WLRTR. (k) DS2DP. (l) NGmeet. (m) TLNLGS .

TABLE IV
RUNNING TIME IN SECONDS OF THE TESTING METHODS FOR REAL DATASETS

Dataset	LRTA	TDL	LRTDTV	HyRes	FastHyDe	GLF	KBR	LLRT	WLRTR	DS2DP	NGmeet	TLNLGS
Indian Pines	1.1	8.5	65.5	2.3	2.2	27.3	1413.9	1426.3	2737.6	5574.3	32.5	33.2
GF-5	1.8	32.8	125.2	6.1	5.4	84.4	2552.0	2903.6	4046.8	5934.8	74.6	76.0

GLF, KBR, WLRTR, and DS2DP obtain satisfactory visual effect, removing the noise and preserving the image details, but the stripe noise exists in the enlarged box. Although LLRT removes the noise completely, the spectral information is distorted. NGmeet removes the noise by smoothing the image detail. The proposed TLNLGS method effectively removes the mixture noise and preserves the image structures.

Fig. 13 shows the mean profiles of band 154 on the GF-5 dataset. Since LRTA and TDL cannot remove noise, the mean profiles obtained by these methods are similar to the noisy band. Due to the effects of stripes, there are several fluctuations in the curves of the LRTDTV, HyRes, FastHyDe, GLF, WLRTR, and DS2DP methods. After denoising, these kinds of fluctuations can be eliminated by other methods, indicating better denoising performances. In summary, our method can effectively remove noise and achieve competitive denoising results compared with the state-of-the-art methods.

Table IV reports the running time of testing methods for these two real datasets. LRTA, FastHyRes, and HyRes have obvious speed advantages than other methods. However, the proposed method achieves better denoising results than the LRTA, TDL, HyRes, and FastHyDe methods with a low computational cost. Therefore, the proposed method has a tradeoff between the results and time cost.

C. Discussion

In this section, we first test the ablation study to illustrate the effectiveness of exploring the nonlocal group sparsity of reduced image under transform learning. Second, fixed transform is used to demonstrate the necessity of adaptive transform. Third, we provide the sensitivity analysis of several parameters adopted in Algorithm 1, and then discuss how they can be appropriately tuned in the experiment. Fourth, comparison with the DL-based methods is presented to show the competitiveness of the proposed method. Finally, the

TABLE V
QUANTITATIVE RESULT COMPARISONS OF ABLATION STUDY ON THE PAC DATASET

σ	Index	TLNLNN	TLNLSR	TNNLUT	TLNLGS
10	PSNR	41.79	39.47	42.51	**43.33**
	SSIM	0.988	0.983	0.990	**0.992**
	FSIM	0.993	0.990	0.994	**0.995**
	ERGAS	30.54	39.25	28.13	**25.45**
	SAM	3.369	3.280	2.680	**2.363**
100	PSNR	30.56	24.18	30.17	**30.73**
	SSIM	0.890	0.712	0.881	**0.894**
	FSIM	0.932	0.854	0.925	**0.934**
	ERGAS	109.07	226.06	113.84	**106.45**
	SAM	6.077	14.082	6.986	**5.710**
[10, 100]	PSNR	35.74	34.11	36.17	**36.82**
	SSIM	0.961	0.932	0.965	**0.970**
	FSIM	0.976	0.965	0.976	**0.981**
	ERGAS	60.33	82.54	57.46	**53.17**
	SAM	4.218	7.701	5.124	**3.939**

numerical convergence is devolved. These discussions are based on the PaC dataset under noise intensities $\sigma = 10, 100$, [10, 100].

1) Ablation Study: To further analyze the superiority of imposing group sparse regularization and different transforms $\mathbf{W}_i$, we conduct an ablation study by replacing group sparse regularization as the weighted nuclear norm and sparse regularization, respectively, as well as using a unified transform $\mathbf{W}$. The substitutes of the weighted nuclear norm, sparse regularization, and unified transform are denoted as TLNLNN, TLNLSR, and TLNLUT, respectively. Table V lists the quantitative result comparisons of the ablation study on the PaC dataset under three different noise variances. Since TLNLSR ignores high strong correlation of the grouped image, the results are lower than TLNLNN which considers the nonlocal self-similarity of the reduced image. Although TLNLUT uses
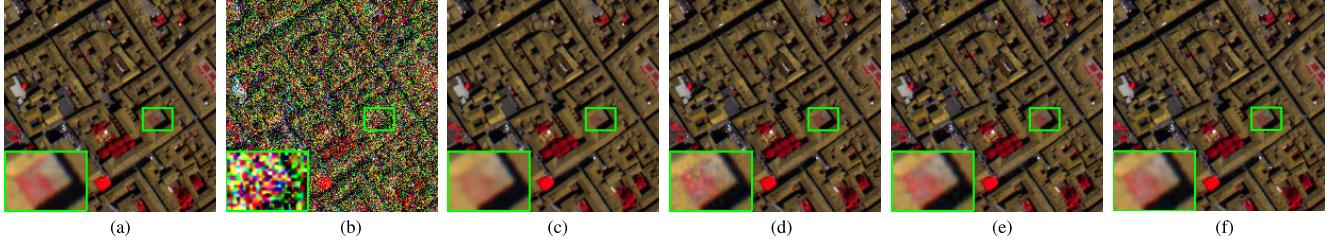
Fig. 14. Denoised results of the PaC dataset in the case $\sigma = [10, 100]$. The false color image is composed by bands (R: 68, G: 39, and B: 1). (a) Original. (b) Noisy. (c) TLNLNN. (d) TLNLSR. (e) TLNLUT. (f) TLNLGS.

TABLE VI
QUANTITATIVE RESULT COMPARISONS OF FIXED DCT TRANSFORM ON THE PAC DATASET

σ	method	PSNR	SSIM	FSIM	ERGAS	SAM
10	DCTNLGS	39.76	0.983	0.990	38.03	3.655
	TLNLGS	**43.33**	**0.992**	**0.995**	**25.45**	**2.363**
100	DCTNLGS	27.64	0.798	0.884	152.45	9.266
	TLNLGS	**30.73**	**0.894**	**0.934**	**106.45**	**5.710**
[10, 100]	DCTNLGS	30.15	0.876	0.926	114.14	7.625
	TLNLGS	**36.82**	**0.970**	**0.981**	**53.17**	**3.939**

Fig. 15. Sensitivity analysis of all the parameters involved in the proposed model. (a) Parameter λ. (b) Parameter $\tau = \sigma\theta$. (c) Patch size p. (d) Patch number d.

unified transform **W**, it simultaneously explores the nonlocal self-similarity and local smooth structure of the reduced image. Therefore, TLNLUT achieves better results than TLNLNN and TLNLSR in most cases. Compared with the variants, TLNLGS obtains the best quantitative results. Fig. 14 presents the visual comparison of Gaussian noise with $\sigma = [10, 100]$. From the results, we can observe that the proposed TLNLGS achieves obvious improvement in terms of removing noises and preserving the image details. In summary, the quantitative and qualitative results demonstrate that these two important components contribute significantly to the superiority of the proposed method.

2) Necessity of Adaptive Transform: In the proposed model, the adaptive transform matrices are learned from the data itself to explore the group sparsity of grouped images. To illustrate the necessity of adaptive transforms, we use fixed DCT as the transform matrix (denoted as DCTNLGS method) to replace the adaptive transform. Table VI lists the quantitative result comparisons on the PaC dataset under three different noise variances. From the results, we can find that the strategy of adaptive transform is significantly better than that of the fixed transform matrix.

3) Sensitivity Analysis of Parameter λ: The regularization parameter λ is used to balance the fidelity term and the regularization term. Fig. 15(a) shows the change in PSNR value with different λ values in the PaC dataset under three noise cases. It can be observed that the denoising results gradually become better with the increase in λ and tend to stabilize when λ increased to 0.1. Therefore, we empirically set $\lambda = 0.1$ in all the experiments.

4) Sensitivity Analysis of Parameter τ: The parameter τ is used to constrain group sparsity regularization, which is an important parameter. In our work, we set the parameter
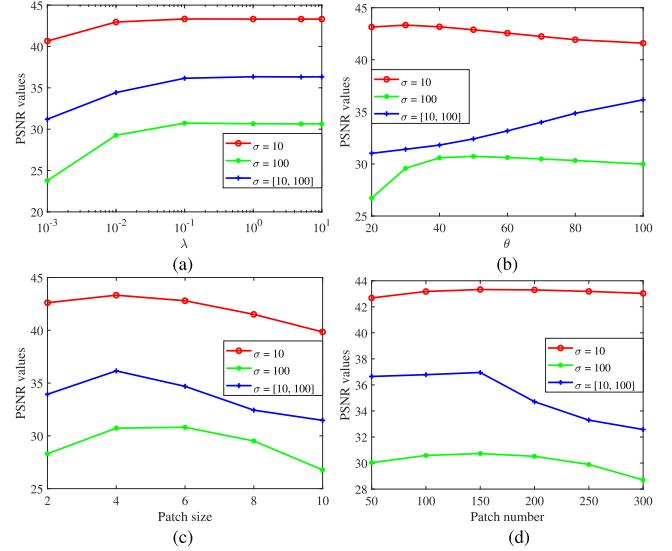
τ as $\sigma\theta$, where σ is the noise level and changes during iteration [41], and θ is the tuning parameter. Fig. 15(b) presents the PSNR change curve with different parameter θ. It is observed that θ needs to be set as lower values on low noise variance, but when the noise variance increases, a larger θ should be selected. From the sensitivity analysis, we empirically set $\theta \in [20, 100]$ in all the experiments.

5) Sensitivity Analysis of Patch Size p: The proposed method has been considered the nonlocal self-similarity of the reduced image, and thus, the patch size also affects the denoising results. The sensitivity analysis of the patch size b is presented in Fig. 15(c). It is observed that this parameter needs to be carefully selected, and satisfactory results in all the cases are achieved when p is set as 4. Therefore, we empirically set $p = 4$ in all the experiments.

6) Sensitivity Analysis of Patch Number d: Fig. 15(d) plots the sensitivity analysis of patch number d, and it is obvious that the highest PSNRs in three cases are achieved by the proposed method when the patch number increased to 150, which inspired us to fix $d = 150$ in all the experiments.

Through the sensitivity analysis of several parameters, we set parameters $\lambda = 0.1$, $p = 4$, and $d = 150$ in all the experiments. Since the parameter $\tau = \sigma\theta$ (where σ is estimated in each iteration [41]) is a threshold value for
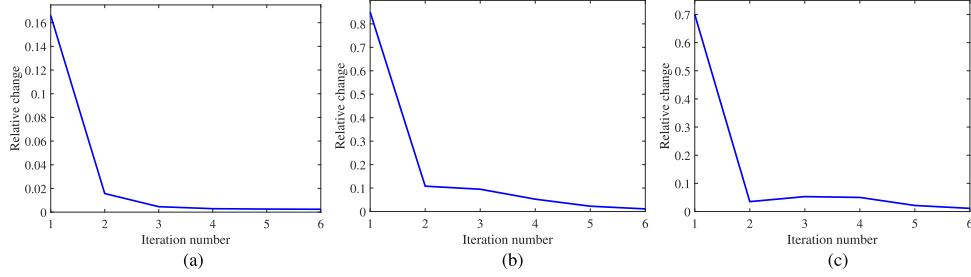
Fig. 16. Relative change values ($\|\mathcal{X}^{k+1} - \mathcal{X}^k\|_F / \|\mathcal{X}^k\|_F$) versus iteration number of the proposed method. (a) $\sigma = 10$. (b) $\sigma = 100$. (c) $\sigma = [10, 100]$.

TABLE VII

SETTING OF PARAMETER θ OF THE PROPOSED METHOD IN SIMULATED EXPERIMENTS

σ	2	10	30	80	100	[10, 100]	[30, 80]	mixed
θ	30	30	30	40	50	100	100	20

TABLE VIII

QUANTITATIVE RESULT COMPARISONS BETWEEN THE PROPOSED TLNLGS WITH TWO SUPERVISED DL-BASED METHODS ON THE WDC DATASET

σ	method	PSNR	FSIM	SAM
5	QRNN3D	43.42	0.982	2.005
	T3SC	43.85	0.981	**1.891**
	TLNLGS	**45.15**	**0.997**	1.891
25	QRNN3D	35.04	0.961	4.068
	T3SC	36.74	0.967	**3.437**
	TLNLGS	**38.42**	**0.988**	3.437
50	QRNN3D	31.72	0.949	5.042
	T3SC	33.12	0.954	**4.412**
	TLNLGS	**34.82**	**0.975**	5.042
100	QRNN3D	27.41	0.910	7.448
	T3SC	29.48	0.933	**5.615**
	TLNLGS	**31.08**	**0.948**	6.761
[0, 95]	QRNN3D	35.84	0.961	4.125
	T3SC	37.20	0.969	**3.438**
	TLNLGS	**38.46**	**0.989**	4.240

group sparsity coding, it needs to be selected according to the noise intensity. In the simulated experiments, the parameter θ is listed in Table VII. Empirically, the Gaussian noise with different standard deviation σ for each band in real dataset. Therefore, when the dataset (e.g., Indian Pines) is obviously degraded by Gaussian noise, we set the parameters as $\lambda = 0.1$, $p = 4$, $d = 150$, and $\theta = 100$. For the GF-5 dataset, it is degraded by mixed noise, and thus, the parameters are set as $\lambda = 0.1$, $p = 4$, $d = 150$, and $\theta = 20$.

7) Comparison With Supervised Deep Learning-Based Methods: Supervised DL has also been extensively presented to HSI denoising [67], [68], [69], [70]. To show the competitiveness of the proposed method with the supervised DL methods, two state-of-the-art methods 3-D quasi-recurrent neural network (QRNN3D) [69] and trainable spectral-spatial sparse coding (T3SC) [70] are chosen for comparison. For a fair comparison, the experimental settings including the ground-truth WDC data ($200 \times 200 \times 191$) and Gaussian noise intensities

($\sigma = 5, 25, 50, 100, [0, 95]$) keep the same as that of T3SC in their article. The quantitative result comparison between the proposed TLNLGS with two supervised DL methods is listed in Table VIII. It can be seen that the proposed TLNLGS outperforms the DL-based methods in terms of PSNR and FSIM, and the T3SC method obtains better SAM results. Therefore, compared with the supervised DL methods, our method can also achieve competitive results.

8) Numerical Convergence: The solved BCD algorithm can converge to a critical point. Fig. 16 shows the relative change value versus the iteration number of the proposed method on the PaC dataset. From the results, we can observe that the relative change value monotonically decreases and gradually tends to be zeros as the iteration number increases, indicating the numerical convergence of the proposed method.

VI. CONCLUSION

In this article, we proposed a novel HSI denoising method using the transform learning scheme with subspace representation. The group sparsifying transform learning is used to explore a discriminatively intrinsic nonlocal prior on the reduced image, which can not only reflect the nonlocal self-similarity and local spatial smooth structure of HSI without adding more constraints but also improve the ability of adaptive sparse representation than the fixed transform domain. Moreover, to capture the global spectral correlation of HSI and avoid the heavy computation burden with the spectral band increase, subspace representation is introduced to project the HSI onto a relatively low-dimensional subspace. The proposed method overcomes the barriers of the existing HSI denoising approaches that directly designed different regularizations on the original domain or fixed domain, which cannot explore the characteristic of transformation. The experiments on the simulated and real HSI noisy cases demonstrated the superiority of the proposed method via comparing with the state-of-the-art approaches.

It should be pointed out that the suggested idea on the exploration of nonlocal group sparsity can also be adapted to other HSI tasks such as superresolution and compressive reconstruction. Another interesting future direction is to adopt the DL methods to learn the transform matrix [71].

REFERENCES

[1] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[2] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatiospectral information fusion and PLS regression," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1127–1137, Mar. 2015.

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

[4] 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.

[5] L. Zhang and L. Zhang, "Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 270–294, Jun. 2022.

[6] I. Atkinson, F. Kamalabadi, and D. L. Jones, "Wavelet-based hyperspectral image estimation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 2, Jul. 2003, pp. 743–745.

[7] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial–spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 397–408, Feb. 2006.

[8] 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.

[9] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2949–2956.

[10] 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.

[11] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop*, Oct. 2019, pp. 3844–3851.

[12] Y.-C. Miao, X.-L. Zhao, X. Fu, J.-L. Wang, and Y.-B. Zheng, "Hyperspectral denoising using unsupervised disentangled spatiospectral deep priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.

[13] B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi, and J. Chanussot, "Noise reduction in hyperspectral imagery: Overview and application," *Remote Sens.*, vol. 10, no. 3, p. 482, 2018.

[14] B. Rasti, Y. Chang, E. Dalsasso, L. Denis, and P. Ghamisi, "Image restoration for remote sensing: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 201–230, Jun. 2022.

[15] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image denoising using first order spectral roughness penalty in wavelet domain," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2458–2467, Jun. 2014.

[16] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 499–515, Apr. 2013.

[17] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.

[18] H. K. Aggarwal and A. Majumdar, "Hyperspectral image denoising using spatio-spectral total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 442–446, Mar. 2016.

[19] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[20] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, and G. He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5366–5380, Sep. 2017.

[21] 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.

[22] A. Karami, M. Yazdi, and A. Z. Asli, "Noise reduction of hyperspectral images using kernel non-negative tucker decomposition," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 487–493, Jun. 2011.

[23] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3556–3570, Aug. 2020.

[24] 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.

[25] 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.

[26] 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 J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 713–729, Mar. 2018.

[27] Y. Wang, J. Peng, Q. Zhao, D. Meng, Y. Leung, and X.-L. Zhao, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1227–1243, Apr. 2018.

[28] H. Fan, C. Li, Y. Guo, G. Kuang, and J. Ma, "Spatial–spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6196–6213, Oct. 2018.

[29] T.-X. Jiang, L. Zhuang, T.-Z. Huang, X.-L. Zhao, and J. M. Bioucas-Dias, "Adaptive hyperspectral mixed noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.

[30] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5174–5189, Jul. 2019.

[31] 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.

[32] 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.

[33] Y. Chang, L. Yan, and S. Zhong, "Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4260–4268.

[34] Y. Chang, L. Yan, X. Zhao, H. Fang, Z. Zhang, and S. Zhong, "Weighted low-rank tensor recovery for hyperspectral image restoration," *IEEE Trans. Cybern.*, vol. 50, no. 11, pp. 4558–4572, Nov. 2020.

[35] Y. Chen, W. He, N. Yokoya, T.-Z. Huang, and X.-L. Zhao, "Nonlocal tensor-ring decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1348–1362, Oct. 2020.

[36] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[37] B. Rasti, J. R. Sveinsson, and M. O. Ulfarsson, "Wavelet-based sparse reduced-rank regression for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6688–6698, Oct. 2014.

[38] 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.

[39] B. Rasti, M. O. Ulfarsson, and P. Ghamisi, "Automatic hyperspectral image restoration using sparse and low-rank modeling," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2335–2339, Dec. 2017.

[40] 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.

[41] W. He *et al.*, "Non-local meets global: An integrated paradigm for hyperspectral image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 2089–2107, Apr. 2020.

[42] L. Zhuang, M. K. Ng, X. Fu, and J. M. Bioucas-Dias, "Hy-demosaicing: Hyperspectral blind reconstruction from spectral subsampling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5515815.

[43] L. Zhuang and M. K. Ng, "Hyperspectral mixed noise removal by ℓ_1-norm-based subspace representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1143–1157, 2020.

[44] Y. Chen, J. Zeng, W. He, X.-L. Zhao, and T.-Z. Huang, "Hyperspectral and multispectral image fusion using factor smoothed tensor ring decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5515417.

[45] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.

[46] H. Kong, C. Lu, and Z. Lin, "Tensor Q-rank: New data dependent definition of tensor rank," *Mach. Learn.*, no. 110, pp. 1867–1900, Jun. 2021.

[47] S. Ravishankar and Y. Bresler, "Sparsifying transform learning with efficient optimal updates and convergence guarantees," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2389–2404, May 2015.

[48] B. Wen, S. Ravishankar, and Y. Bresler, "Structured overcomplete sparsifying transform learning with convergence guarantees and applications," *Int. J. Comput. Vis.*, vol. 114, nos. 2–3, pp. 137–167, 2015.

[49] S. Ravishankar, A. Lahiri, C. Blocker, and J. A. Fessler, "Deep dictionary-transform learning for image reconstruction," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1208–1212.

[50] B. Wen, Y. Li, and Y. Bresler, "Image recovery via transform learning and low-rank modeling: The power of complementary regularizers," *IEEE Trans. Image Process.*, vol. 29, pp. 5310–5323, 2020.

[51] W. He, N. Yokoya, and X. Yuan, "Fast hyperspectral image recovery of dual-camera compressive hyperspectral imaging via noniterative subspace-based fusion," *IEEE Trans. Image Process.*, vol. 30, pp. 7170–7183, 2021.

[52] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.

[53] J. Xue, Y.-Q. Zhao, Y. Bu, W. Liao, J. C.-W. Chan, and W. Philips, "Spatial–spectral structured sparse low-rank representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 3084–3097, 2021.

[54] Y. Chen, T.-Z. Huang, W. He, X.-L. Zhao, H. Zhang, and J. Zeng, "Hyperspectral image denoising using factor group sparsity-regularized nonconvex low-rank approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5515916.

[55] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, Mar. 1997.

[56] J. Xue, Y. Zhao, Y. Bu, J. C.-W. Chan, and S. G. Kong, "When Laplacian scale mixture meets three-layer transform: A parametric tensor sparsity for tensor completion," *IEEE Trans. Cybern.*, early access, Jan. 26, 2022, doi: 10.1109/TCYB.2021.3140148.

[57] J. Xue, Y. Zhao, S. Huang, W. Liao, J. C.-W. Chan, and S. G. Kong, "Multilayer sparsity-based tensor decomposition for low-rank tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 18, 2021, doi: 10.1109/TNNLS.2021.3083931.

[58] A. Zelinski and V. Goyal, "Denoising hyperspectral imagery and recovering junk bands using wavelets and sparse approximation," in *Proc. IEEE Int. Symp. Geosci. Remote Sens.*, Jul. 2006, pp. 387–390.

[59] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Wavelet based hyperspectral image restoration using spatial and spectral penalties," *Proc. SPIE*, vol. 8892, pp. 135–142, Oct. 2013.

[60] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[61] J. Zeng, T. T.-K. Lau, S.-B. Lin, and Y. Yao, "Global convergence of block coordinate descent in deep learning," in *Proc. ICML*, 2019, pp. 7313–7323.

[62] Y. Fu *et al.*, "Exploring structural sparsity of deep networks via inverse scale spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Apr. 22, 2022, doi: 10.1109/TPAMI.2022.3168881.

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

[64] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[65] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[66] L. Wald, *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*. Presses des MINES, 2002.

[67] 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 Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7317–7329, Oct. 2019.

[68] X. Cao, X. Fu, C. Xu, and D. Meng, "Deep spatial–spectral global reasoning network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.

[69] 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.

[70] T. Bodrito, A. Zouaoui, J. Chanussot, and J. Mairal, "A trainable spectral-spatial sparse coding model for hyperspectral image restoration," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 5430–5442.

[71] Y.-S. Luo, X.-L. Zhao, T.-X. Jiang, Y. Chang, M. K. Ng, and C. Li, "Self-supervised nonlinear transform-based tensor nuclear norm for multidimensional image recovery," *IEEE Trans. Image Process.*, vol. 31, pp. 3793–3808, 2022, doi: 10.1109/TIP.2022.3176220.

Yong Chen (Member, IEEE) received the B.S. degree from the School of Science, East China University of Technology, Nanchang, China, in 2015, and the Ph.D. degree from the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2020.

He is currently working with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang. From 2018 to 2019, he was a Research Intern with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. His research interests include remote sensing image processing, low-rank matrix/tensor representation, and deep learning (DL). More information can be found on his homepage at: https://chenyong1993.github.io/yongchen.github.io/.

Wei He (Senior Member, IEEE) received the B.S. degree from the School of Mathematics and Statistics, Wuhan University, Wuhan, China, in 2012, and the Ph.D. degree in surveying, mapping and remote sensing (LIESMARS) from Wuhan University, in 2017.

He is currently a Full Professor with LIESMARS, Wuhan University. From 2018 to 2020, he was a Researcher with the Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan, where he was a Research Scientist from 2020 to 2021. His research interests include image quality improvement, remote sensing image processing, low-rank representation, and deep learning (DL).

Xi-Le Zhao (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 with Prof. Michael K. Ng, Hong Kong Baptist University, Hong Kong, from 2013 to 2014. He worked as a Visiting Scholar with Prof. Jose Bioucas Dias, 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 models, algorithms, and theories for low-level inverse problems of multidimensional images. More information can be found on his homepage at: https://zhaoxile.github.io/.

Ting-Zhu Huang received the B.S., M.S., and Ph.D. degrees in computational mathematics from the Department of Mathematics, Xian Jiaotong University, Xian, China, in 1986, 1992, and 2001, respectively.

He is currently a Professor with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China. His research interests include scientific computation and applications, numerical algorithms for image processing, numerical linear algebra, preconditioning technologies, and matrix analysis with applications.

Prof. Huang is an Editor for 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.

Hui Lin (Senior Member, IEEE) received the Diploma degree in aero-photogrammetry engineering from the Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 1980, the M.Sc. degree in cartography and remote sensing from the Chinese Academy of Sciences, Beijing, China, in 1983, and the M.A. and Ph.D. degrees in geographic information systems (GIS) from the State University of New York at Buffalo, Buffalo, NY, USA, in 1987 and 1992, respectively.

He is a Professor and the Dean of the School of Geography and Environment, Jiangxi Normal University, Nanchang, China. He is an Academician of International Eurasian Academy of Sciences and Fellow of British Academy of Social Science. He currently serves as the Vice Chairperson, China National Committee of International Society of Digital Earth. His major research interests include satellite remote sensing for cloudy and rainy environments, virtual geographic environments, and geographic information system applications.

Jinshan Zeng received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 2015.

He is currently a Distinguished Professor with the School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, China, and serves as the Director of the Department of Data Science and Big Data. He has published over 40 articles in high-impact journals and conferences such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), JMLR, the IEEE TRANSACTIONS ON SIGNAL PROCESSING (TSP), ICML, and AAAI. He has had two articles coauthored with collaborators that received the International Consortium of Chinese Mathematicians (ICCM) Best Paper Award (2018, 2020). His research interests include nonconvex optimization, machine learning [in particular deep learning (DL)], and remote sensing.