Hyperspectral Mixed Noise Removal via Spatial-Spectral Constrained Unsupervised Deep Image Prior

Yi-Si Luo, Xi-Le Zhao ⬤, *Member, IEEE*, Tai-Xiang Jiang ⬤, *Member, IEEE*, Yu-Bang Zheng ⬤, *Student Member, IEEE*, and Yi Chang ⬤, *Member, IEEE*

Abstract—Recently, deep learning-based methods are proposed for hyperspectral images (HSIs) denoising. Among them, unsupervised methods such as deep image prior (DIP)-based methods have received much attention because these methods do not require any training data. However, DIP-based methods suffer from the semiconvergence behavior, i.e., the iteration of DIP-based methods needs to terminate by referring to the ground-truth image at the optimal iteration point. In this article, we propose the spatial-spectral constrained deep image prior (S2DIP) for the HSI mixed noise removal. Specifically, we integrate the DIP, the spatial-spectral total variation regularization term, and the ℓ_1-norm sparse term to respectively capture the deep prior of the clean HSI, the spatial-spectral local smooth prior of the clean HSI, and the sparse prior of noise. The proposed S2DIP jointly leverages the expressive power brought from the deep convolutional neural network without any training data and exploits the HSI and noise structures via hand-crafted priors. Thus, our method avoids the semiconvergence behavior of DIP-based methods. Meanwhile, our method largely enhances the HSI denoising ability of DIP-based methods. To tackle the corresponding model, we utilize the alternating direction multiplier method algorithm. Extensive experiments demonstrate that our method outperforms model-based and deep learning-based state-of-the-art HSI denoising methods.

Index Terms—Convolutional neural networks (CNNs), denoising, hyperspectral image, spatial-spectral, unsupervised.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) contain abundant spatial and spectral information, and thus, can be utilized into various applications, such as object detection [2], [3], classification [4]–[7], and so on. Due to the limitation of the imaging system, HSIs are inevitably corrupted by mixed noise. In particular, due to the miscalibration in the detectors of the imaging system [8], [9], HSIs are often corrupted by along-track stripe noise. In this work, we consider mixed noise including Gaussian noise and other complex sparse noise, which contains impulse noise, stripe noise, and deadlines. The mixed noise seriously affects subsequent applications as well as visual qualities. Thus, HSI denoising is considered an essential technique in hyperspectral imaging.

Earlier HSI denoising methods are dominated by model-based techniques [10]–[15], such as total variation (TV)-based methods [16]–[20], sparsity-based methods [21], [22], nonlocal-based methods [23], dictionary learning-based methods [24], [25], and matrix/tensor low-rankness-based methods [16], [20], [26]–[35]. These methods exploit the intrinsic structure of HSIs via optimization-based models to deal with the HSI denoising. Algorithms like alternating direction multiplier method (ADMM) can be utilized to address these denoising models. However, due to the lack of representation abilities, model-based methods are sometimes hard to capture real-world HSIs, which contain more fine details.

Inspired by the success of deep learning in inverse problems in imaging [36]–[40], deep learning-based methods [41]–[46] have emerged for HSI denoising in last few years. Generally, most of the deep learning-based methods learn the noisy-to-denoising mapping driven by abundant training data [42], [44], which leads to good performance for HSI denoising on specific datasets, thanks to the expressive power of the deep convolutional neural network (CNN). Nevertheless, supervised deep learning-based methods critically depend on the diversity and quantity of training data. As the real noise of HSIs is complex, it is difficult to guarantee high-quality denoising results under the realist complex noise scenario of HSIs, where the underlying assumption is not held in the training data.

Recently, an unsupervised image restoration method deep image prior (DIP) [47] was proposed. The DIP employs the CNN and optimizes its learnable parameters by targeting the
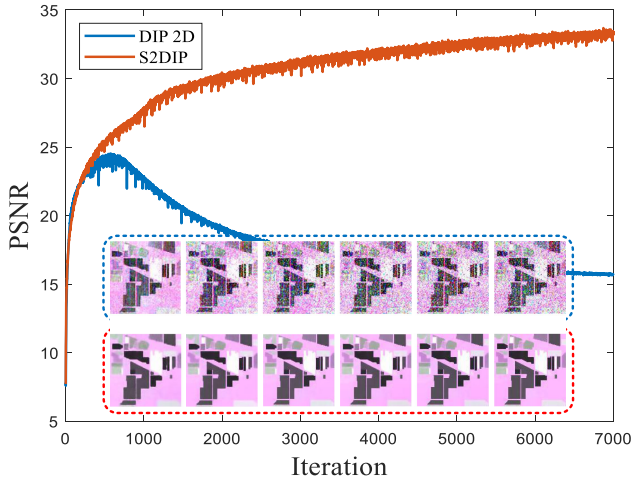
Fig. 1. History of PSNR values of the denoising results on HSI *Indian pines* by DIP 2D [1] and S2DIP with respect to the iterations. The mixed noise contains Gaussian noise, impulse noise, stripe noise, and deadlines. Six images refer to the denoising results of DIP 2D and S2DIP at tth iteration ($t = 1000, 2000, \dots, 6000$).

observed image as the network output with a randomly given input, demonstrating that the CNN itself can represent a well-reconstructed image by the iterative process without any training data. More recently, Sidorov *et al.* [1] extended the DIP into HSI restoration, which includes HSI denoising, inpainting, and super-resolution.

However, DIP-based methods have the following two limitations for the HSI mixed noise removal.

1) First, DIP-based methods suffer from the inherent semiconvergence behavior, which refers to the behavior of an iterative method that the PSNR value begins to increase at early iterations and, after a certain "optimal" iteration, the PSNR value begins to decrease (see Fig. 1). Hence, the iteration needs to terminate by referring to the ground-truth image at the point of the highest PSNR value.

2) Second, DIP-based methods, which are designed for the Gaussian noise removal, is not resilient to mixed noise. DIP-based methods can not remove the complex mixed noise due to the lack of considerations on the clean HSI prior and the noise prior.

We attribute these limitations of DIP-based methods to two aspects. First, DIP-based methods lack specific characterization on HSI intrinsic structures. HSIs have unique spatial-spectral structures, which require specific prior terms to characterize. Second, DIP-based methods lack robust noise modeling. In real applications, the noise of HSI is very complex and various. Only considering Gaussian noise and neglect other types of noise is not enough to completely remove the complex mixed noise in HSI.

In this article, we propose the spatial-spectral constrained deep image prior (S2DIP) for the HSI mixed noise removal. We organically integrate the unsupervised deep image prior with prior information on both clean HSI and hyperspectral noise. Specifically, we integrate the DIP, the spatial-spectral total variation (SSTV) regularization term, and the ℓ_1-norm sparse

term to respectively capture the deep prior of the clean HSI, the spatial-spectral local smooth prior of the clean HSI, and the sparse prior of the noise for the mixed noise removal. To tackle the corresponding model, we utilize the ADMM algorithm. Meanwhile, S2DIP allows us to easily design an automatic stopping criterion without referring to the ground-truth image. The deep image prior, the clean HSI prior, and the noise prior are complementary to each other and are organically combined to benefit each other. The proposed S2DIP greatly enhances the performance of DIP-based methods for the mixed noise removal and alleviates the inherent semiconvergence behavior, which is a sore point of the original DIP-based methods.

Compared with model-based methods, the proposed S2DIP capitalizes the expressive power from the CNN and holds higher model representation ability. Compared with supervised deep learning-based methods, the proposed S2DIP does not need training data and has a better generalization ability for diverse HSI data with various complex noise.

We summarize the contributions of this article as follows.

1) To address the limitations of unsupervised DIP-based methods for HSIs denoising, we suggest the S2DIP for the HSI mixed noise removal by exploiting the intrinsic clean HSI prior and the noise prior. Our method greatly enhances the performance of DIP-based methods for the HSI mixed noise removal and alleviates the inherent semiconvergence behavior of DIP-based methods.

2) To tackle the corresponding model, we utilize the efficient ADMM algorithm, which allows each variable, including CNN parameters, can be efficiently updated in an alternate order. In company with the algorithm, S2DIP also allows us to easily design an automatic stopping criterion without referring to the ground-truth image.

3) Extensive experiments validate that our method greatly enhances the performance of DIP-based methods for the mixed noise removal and alleviates the semiconvergence behavior of DIP-based methods. Meanwhile, comprehensive comparisons demonstrate that S2DIP outperforms state-of-the-art model-based methods (LRMR [26], LRT-DTV [27], and $\ell_0 - \ell_1$ HTV [48]) and deep learning-based methods (HSID-CNN [44], DIP 2D [1], and DIP 3D [1]).

The rest of this article is organized as follows. In Section II, we introduce some related work. In Section III, we introduce the proposed S2DIP. In Section IV, we carry out the experimental results. Sec. V provides some discussions. Finally, Section VI concludes this article.

II. RELATED WORK

A. Model-Based Methods for HSI Denoising

Traditional HSI denoising methods are dominated by model-based methods. These methods, including but not limited to linear transforms-based methods [49], [50], TV-based methods [16]–[19], [51]–[55], sparsity-based methods [21], [22], [32], [56], and matrix/tensor low-rankness-based methods [20], [26]–[31], [57]–[59], consider the HSI prior information and establish optimization model and corresponding algorithms for denoising. For instance, Zhang *et al.* [26] vectorized each band

of the HSI as a column and unfolds them as a matrix, and then considered the low-rank property on the unfolded matrix. Wang *et al.* [27] utilized the tensor decomposition, which delivered the global tensor low-rankness. The TV regularization was considered along with the tensor decomposition. More recently, mixed noise removal methods [60]–[62] and image quality improvement methods for specific degradations, e.g., destriping [8], [9], [63]–[65], were studied by using model-based techniques such as low-rankness [66] and sparsity [8]. Such methods achieve state-of-the-art performance due to the comprehensive consideration of the HSI prior information.

B. Deep Learning-Based Methods for HSI Denoising

In recent years, deep learning-based approaches for HSI denoising [42]–[45], [67]–[70] have emerged and presented state-of-the-art performance. For instance, Yuan *et al.* [44] employed the deep residual CNN to conduct HSI denoising. Dong *et al.* [42] utilized the deep 3-D encoder–decoder network for HSI denoising. Zhang *et al.* [71] proposed the deep spatio-spectral Bayesian posterior for HSI denoising. Cao *et al.* [68] used the deep spatial-spectral global reasoning network for HSI denoising. The core concept of these methods is to train a CNN with abundant pairs of training data $\{\mathcal{Y}, \mathcal{X}\}$, where $\mathcal{Y}$ denotes the dataset of noisy HSIs and $\mathcal{X}$ denotes the dataset of corresponding clean HSIs. The training process could be described as

$$\min_{\Theta} \mathcal{L}(f_\Theta(\mathcal{Y}), \mathcal{X}) \tag{1}$$

where $f_\Theta(\cdot)$ denotes a CNN with learnable parameters Θ, and $\mathcal{L}$ is the loss function. The well-trained CNN can be viewed as a noisy-to-denoising mapping. Due to the powerful nonlinear modeling ability of the deep CNN, supervised methods have achieved promising results on specific datasets and specific noisy types.

C. Deep Image Prior

Recently, Ulyanov *et al.* [47] proposed an unsupervised deep learning-based image restoration technique, named deep image prior. The DIP uses CNN to conduct image restoration without any training process. By targeting the degraded image as the network output with randomly generated network input, the CNN could remove the Gaussian noise from the degraded image with appropriate iteration steps. More recently, Sidorov *et al.* [1] extended DIP-based methods into HSI restoration.

Similar to the supervised methods, the optimization process of DIP-based methods for HSI [1] restoration is formulated as

$$\min_{\Theta} \frac{1}{N} \|f_\Theta(\mathcal{Z}) - \mathcal{Y}\|_{\ell_2}^2 \tag{2}$$

where $\mathcal{Z} \in \mathbb{R}^{H \times W \times B}$ denotes the randomly generated network input and $\mathcal{Y}$ denotes the noisy HSI. The mean square error (MSE) is adopted as the loss function, where $N = H \times W \times B$ denotes the number of total pixels. The gradient descent algorithm is adopted to iteratively optimize the CNN. Obviously, the optimization process will finally converge to a noisy HSI $f_\Theta(\mathcal{Z})$, which is almost structurally identical with the observation $\mathcal{Y}$. However, it has been discovered that the network will first fit the signal part of the observation, and then fit the noisy part [47]. Thus, the denoising of $\mathcal{Y}$ can be achieved by stopping the iteration at appropriate steps before the network fitting noise, and the denoising result is obtained by $\mathcal{X} = f_{\Theta^*}(\mathcal{Z})$ with Θ^* denotes the network parameters after appropriate iterations. This is also a key limitation of DIP-based methods, where the denoising iterative process must be stopped by referring to the ground-truth image at the optimal point (see Fig. 1) for an example of DIP 2D for HSI denoising. We refer to this limitation as semiconvergence. In this article, we avoid this issue by combining DIP with spatial-spectral priors of the clean HSI and the sparse prior of noise.

III. PROPOSED S2DIP

This section introduces the proposed S2DIP. First, we introduce the degradation model of mixed noise in HSI. Then, we give the network structure and form the optimization model of our method. To address the proposed model, we utilize the ADMM algorithm. Finally, we propose an automatic stopping criterion to be used for verifying the stability of S2DIP, which addresses the semiconvergence of DIP-based methods.

A. Degradation Model

The noise in HSI is complex and various. Only considering Gaussian noise [1] is hard to comprehensively remove the complex noise. This motivates us to improve the generalization ability of DIP-based methods by considering more robust noise modeling. In this work, we consider the noisy HSI as an addition of the clean HSI, the Gaussian noise, and the sparse noise. The degradation process can be formulated as

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

where $\mathcal{Y} \in \mathbb{R}^{H \times W \times B}$ denotes the noisy HSI, $\mathcal{X} \in \mathbb{R}^{H \times W \times B}$ denotes the underlying clean HSI, $\mathcal{N} \in \mathbb{R}^{H \times W \times B}$ denotes the Gaussian noise, and $\mathcal{S} \in \mathbb{R}^{H \times W \times B}$ denotes the sparse noise. Our method uses the deep prior of CNN to generate $\mathcal{X}$ and fully consider the spatial-spectral prior information of $\mathcal{X}$ and the sparsity of $\mathcal{S}$ for the mixed noise removal.

B. Network Architecture of S2DIP

We employ a U-Net with skip connections for HSI denoising [1], [47] (see Fig. 2). The input data are the random noise, which goes through the encoder and progressively outcomes embedded 3-D features. The embedded features go through the decoder to reconstruct the clean HSI.

In [1], the author tried both 2-D and 3-D convolution in the U-Net for HSI denoising. However, it appears that the performance of 3-D convolution is inferior to that of 2-D convolution. In this article, we use the separable 3-D convolution [72], which is shown to be more capable to encode the HSI than 3-D convolution [42]. The separable 3-D convolution represents the spatial and spectral information using 2-D and 1-D kernels, respectively, so that better representation of the HSI can be obtained, see more details in [42]. Here, we employ four layers in both encoding and decoding stages. We denote
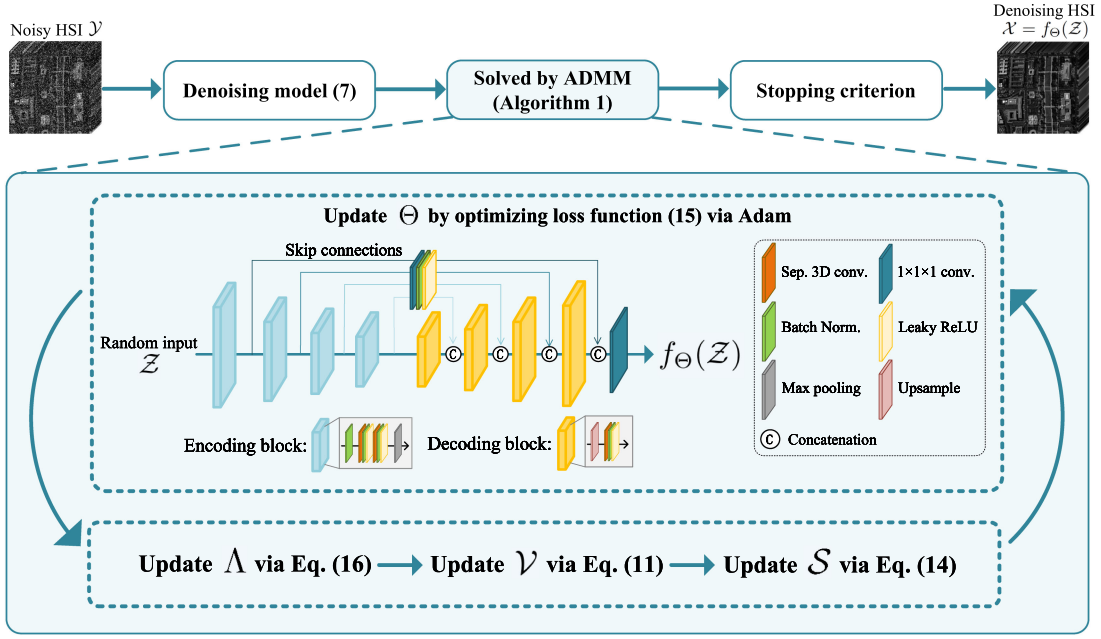
Fig. 2. Overall flowchart of the proposed S2DIP. Given the noisy HSI $\mathcal{Y}$, the proposed denoising model is formulated as (7). To tackle (7), we utilize the ADMM. In company, we design a stopping criterion without referring to the ground-truth image.

the proposed separable 3-D U-Net by $f_\Theta(\cdot)$ with Θ refers to the learnable parameters. We randomly initialize the parameters of the network. The network parameters are iteratively and unsupervisedly updated (see Section III-D and Algorithm 1). Except for the observed HSI, no extra training data are needed.

C. Optimization Model of S2DIP

1) Spatial-Spectral Constraint: The optimization model of DIP-based methods (2) aims at fitting the noisy HSI while the noise could be removed using the so-called deep prior of CNN. However, it is unavoidable for DIP-based methods that the CNN will eventually fit the noise, which causes semiconvergence. To address this issue, we propose to use the SSTV regularization to fully preserve the spatial-spectral local smoothness of the network output so that spatial-spectral consistency can be ensured. The SSTV regularization term is expected to address the semiconvergence and simultaneously improve the denoising performance.

Formally, for a three-way tensor $\mathcal{X} \in \mathbb{R}^{H \times W \times B}$, its derivative tensors $\nabla_x \mathcal{X} \in \mathbb{R}^{(H-1) \times W \times B}$, $\nabla_y \mathcal{X} \in \mathbb{R}^{H \times (W-1) \times B}$, and $\nabla_z \mathcal{X} \in \mathbb{R}^{H \times W \times (B-1)}$ are defined as

$$\begin{cases} \nabla_x \mathcal{X}(i,j,k) = \mathcal{X}(i+1,j,k) - \mathcal{X}(i,j,k) \\ \nabla_y \mathcal{X}(i,j,k) = \mathcal{X}(i,j+1,k) - \mathcal{X}(i,j,k) \\ \nabla_z \mathcal{X}(i,j,k) = \mathcal{X}(i,j,k+1) - \mathcal{X}(i,j,k) \end{cases} \quad (4)$$

where $\mathcal{X}(i,j,k)$ denotes the (i,j,k)th element of $\mathcal{X}$, and ∇_x, ∇_y, and ∇_z denote the finite derivative operators on the vertical direction, horizontal direction, and spectral direction, respectively. The TV of $\mathcal{X}$ is given by

$$\|\mathcal{X}\|_{\text{TV}} = \|\nabla_x \mathcal{X}\|_{\ell_1} + \|\nabla_y \mathcal{X}\|_{\ell_1}. \quad (5)$$
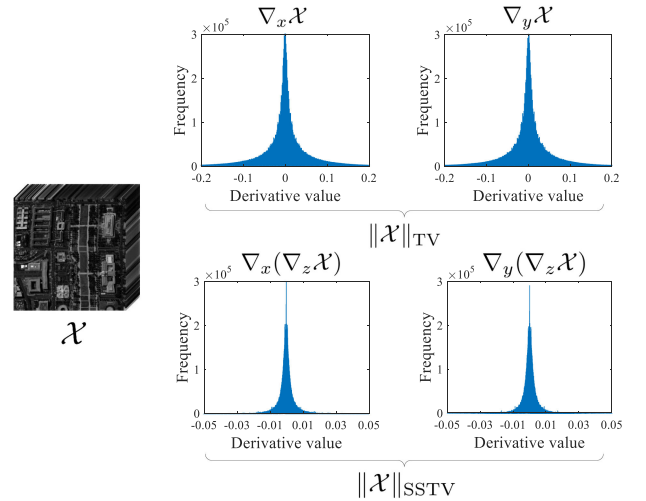
Fig. 3. Statistical distributions of the derivative values of a clean HSI.

Further considering the spatial-spectral local smoothness, the SSTV [17] of $\mathcal{X}$ is given by

$$\|\mathcal{X}\|_{\text{SSTV}} = \|\nabla_x(\nabla_z \mathcal{X})\|_{\ell_1} + \|\nabla_y(\nabla_z \mathcal{X})\|_{\ell_1}. \quad (6)$$

The TV considers the spatial local smoothness and SSTV considers the spatial-spectral local smoothness. Fig. 3 shows the statistical distributions of derivative values of a clean HSI, which illustrates the spatial and spatial-spectral local smoothness. Thus, we introduce TV and SSTV regularizations [18] in the optimization model to respectively explore the spatial and spatial-spectral local smoothness of HSIs.

2) Sparse Noise Modeling: The model of DIP-based methods (2) only considers the Gaussian noise. In real applications, the noise of HSI is complex and various. In this article, we further consider the sparse noise to improve the generalization ability of DIP-based methods for more robust denoising. The sparse noise contains impulse noise, stripe noise, and deadlines that are commonly existing in HSIs. We consider minimizing the ℓ_1-norm of $\mathcal{S}$ to enforce the sparsity on $\mathcal{S}$ so that the sparse noise can be faithfully separated from the clean HSI.

Based on the analysis of spatial-spectral constraints and sparse noise modeling, the proposed optimization model for the HSI mixed noise removal is formulated as

$$\min_{\Theta, \mathcal{S}} \|\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2 + \alpha_1 \|\mathcal{X}\|_{\text{TV}} + \alpha_2 \|\mathcal{X}\|_{\text{SSTV}}$$
$$+ \alpha_3 \|\mathcal{S}\|_{\ell_1} \tag{7}$$
$$\text{where } \mathcal{X} = f_\Theta(\mathcal{Z}).$$

Here, $\alpha_i (i = 1, 2, 3)$ are tradeoff parameters, $\|\mathcal{X}\|_{\text{TV}}$ and $\|\mathcal{X}\|_{\text{SSTV}}$ are the TV and SSTV regularizations, respectively. $\|\mathcal{S}\|_{\ell_1}$ is the ℓ_1-norm of the sparse noise. $\|\mathcal{Y} - \mathcal{X} - \mathcal{S}\|_F^2$ is the fidelity term. The CNN $f_\Theta(\cdot)$ with random input $\mathcal{Z}$ is used to represent the clean HSI $\mathcal{X}$.

In our model, the deep prior, the clean HSI prior, and the sparse noise prior are complementary to each other and are organically combined to remove the mixed noise in HSI.

D. Algorithm

To tackle the proposed model (7), we utilize the efficient ADMM algorithm. By introducing four auxiliary variables $\mathcal{V}_i (i = 1, 2, 3, 4)$, we can reformulate (7) as the equivalent optimization problem

$$\min_{\Theta, \mathcal{S}} \|\mathcal{Y} - f_\Theta(\mathcal{Z}) - \mathcal{S}\|_F^2 + \alpha_1 \|\mathcal{V}_1\|_{\ell_1} + \alpha_1 \|\mathcal{V}_2\|_{\ell_1}$$
$$+ \alpha_2 \|\mathcal{V}_3\|_{\ell_1} + \alpha_2 \|\mathcal{V}_4\|_{\ell_1} + \alpha_3 \|\mathcal{S}\|_{\ell_1}$$
$$\text{s.t. } \mathcal{V}_1 = \nabla_x(f_\Theta(\mathcal{Z})), \ \mathcal{V}_2 = \nabla_y(f_\Theta(\mathcal{Z}))$$
$$\mathcal{V}_3 = \nabla_x \nabla_z(f_\Theta(\mathcal{Z})), \ \mathcal{V}_4 = \nabla_y \nabla_z(f_\Theta(\mathcal{Z})). \tag{8}$$

The augmented Lagrangian function is given as follows by attaching Lagrangian multipliers:

$$\mathcal{L}_\mu(\Theta, \mathcal{S}, \mathcal{V}_i, \Lambda_i)$$
$$= \|\mathcal{Y} - f_\Theta(\mathcal{Z}) - \mathcal{S}\|_F^2 + \alpha_1 \|\mathcal{V}_1\|_{\ell_1} + \alpha_1 \|\mathcal{V}_2\|_{\ell_1}$$
$$+ \alpha_2 \|\mathcal{V}_3\|_{\ell_1} + \alpha_2 \|\mathcal{V}_4\|_{\ell_1} + \frac{\mu}{2} \|\nabla_x(f_\Theta(\mathcal{Z})) - \mathcal{V}_1\|_F^2$$
$$+ \frac{\mu}{2} \|\nabla_y(f_\Theta(\mathcal{Z})) - \mathcal{V}_2\|_F^2 + \frac{\mu}{2} \|\nabla_x \nabla_z(f_\Theta(\mathcal{Z})) - \mathcal{V}_3\|_F^2$$
$$+ \frac{\mu}{2} \|\nabla_y \nabla_z(f_\Theta(\mathcal{Z})) - \mathcal{V}_4\|_F^2 + \alpha_3 \|\mathcal{S}\|_{\ell_1}$$
$$+ <\Lambda_1, \nabla_x(f_\Theta(\mathcal{Z})) - \mathcal{V}_1> + <\Lambda_2, \nabla_y(f_\Theta(\mathcal{Z})) - \mathcal{V}_2>$$
$$+ <\Lambda_3, \nabla_x \nabla_z(f_\Theta(\mathcal{Z})) - \mathcal{V}_3>$$
$$+ <\Lambda_4, \nabla_y \nabla_z(f_\Theta(\mathcal{Z})) - \mathcal{V}_4> \tag{9}$$

where μ is the penalty parameter and $\Lambda_i (i = 1, 2, 3, 4)$ are the multipliers. The joint minimization problem can be decomposed into easier and smaller subproblems, followed by the update of Lagrangian multipliers.

$\mathcal{V}$ Subproblems: The $\mathcal{V}_i (i = 1, 2, 3, 4)$ subproblems are

$$\begin{cases} \min_{\mathcal{V}_1} \frac{\mu}{2} \left\|\nabla_x(f_{\Theta^t}(\mathcal{Z})) + \frac{\Lambda_1^t}{\mu} - \mathcal{V}_1\right\|_F^2 + \alpha_1 \|\mathcal{V}_1\|_{\ell_1} \\ \min_{\mathcal{V}_2} \frac{\mu}{2} \left\|\nabla_y(f_{\Theta^t}(\mathcal{Z})) + \frac{\Lambda_2^t}{\mu} - \mathcal{V}_2\right\|_F^2 + \alpha_1 \|\mathcal{V}_2\|_{\ell_1} \\ \min_{\mathcal{V}_3} \frac{\mu}{2} \left\|\nabla_x \nabla_z(f_{\Theta^t}(\mathcal{Z})) + \frac{\Lambda_3^t}{\mu} - \mathcal{V}_3\right\|_F^2 + \alpha_2 \|\mathcal{V}_3\|_{\ell_1} \\ \min_{\mathcal{V}_4} \frac{\mu}{2} \left\|\nabla_y \nabla_z(f_{\Theta^t}(\mathcal{Z})) + \frac{\Lambda_4^t}{\mu} - \mathcal{V}_4\right\|_F^2 + \alpha_2 \|\mathcal{V}_4\|_{\ell_1} \end{cases} \tag{10}$$

which can be exactly solved by

$$\begin{cases} \mathcal{V}_1^{t+1} = Soft_{\frac{\alpha_1}{\mu}}\left(\nabla_x(f_{\Theta^t}(\mathcal{Z})) + \frac{\Lambda_1^t}{\mu}\right) \\ \mathcal{V}_2^{t+1} = Soft_{\frac{\alpha_1}{\mu}}\left(\nabla_y(f_{\Theta^t}(\mathcal{Z})) + \frac{\Lambda_2^t}{\mu}\right) \\ \mathcal{V}_3^{t+1} = Soft_{\frac{\alpha_2}{\mu}}\left(\nabla_x \nabla_z(f_{\Theta^t}(\mathcal{Z})) + \frac{\Lambda_3^t}{\mu}\right) \\ \mathcal{V}_4^{t+1} = Soft_{\frac{\alpha_2}{\mu}}\left(\nabla_y \nabla_z(f_{\Theta^t}(\mathcal{Z})) + \frac{\Lambda_4^t}{\mu}\right) \end{cases} \tag{11}$$

where $Soft_v(\cdot)$ is the soft-thresholding operator defined as

$$Soft_v(\mathcal{X})(i, j, k) = \text{sign}(\mathcal{X}(i, j, k))(\max\{|\mathcal{X}(i, j, k)| - v, 0\}). \tag{12}$$

$\mathcal{S}$ Subproblem: The $\mathcal{S}$ subproblem is

$$\min_S \|\mathcal{Y} - f_{\Theta^t}(\mathcal{Z}) - \mathcal{S}\|_F^2 + \alpha_3 \|\mathcal{S}\|_{\ell_1} \tag{13}$$

which can be exactly solved by

$$\mathcal{S}^{t+1} = Soft_{2\alpha_3}(\mathcal{Y} - f_{\Theta^t}(\mathcal{Z})). \tag{14}$$

Θ Subproblem: The Θ subproblem is

$$\min_\Theta \|\mathcal{Y} - f_\Theta(\mathcal{Z}) - \mathcal{S}^t\|_F^2 + \frac{\mu}{2}\big(\|\nabla_x(f_\Theta(\mathcal{Z})) - \mathcal{D}_1^t\|_F^2$$
$$+ \|\nabla_y(f_\Theta(\mathcal{Z})) - \mathcal{D}_2^t\|_F^2 + \|\nabla_x \nabla_z(f_\Theta(\mathcal{Z})) - \mathcal{D}_3^t\|_F^2$$
$$+ \|\nabla_y \nabla_z(f_\Theta(\mathcal{Z})) - \mathcal{D}_4^t\|_F^2\big) \tag{15}$$

where $\mathcal{D}_i^t = \mathcal{V}_i^t - \frac{\Lambda_i^t}{\mu}$. Following the philosophy of DIP, the underlying clean HSI $\mathcal{X}$ is represented by the deep U-Net $f_\Theta(\cdot)$ with parameters Θ, where the random $\mathcal{Z}$ is the network input, i.e., $\mathcal{X} = f_\Theta(\mathcal{Z})$. We use adaptive moment estimation (Adam) to update Θ, i.e., update $f_\Theta(\mathcal{Z})$. Specifically, we employ one gradient descent step by using the Adam in each update. The loss functions are all F-norm-based functions, so the gradient can be easily computed in current deep learning frameworks.

Λ Updating: The Lagrange multipliers are updated as

$$\begin{cases} \Lambda_1^{t+1} = \Lambda_1^t + \mu(\nabla_x(f_{\Theta^t}(\mathcal{Z})) - \mathcal{V}_1^t) \\ \Lambda_2^{t+1} = \Lambda_2^t + \mu(\nabla_y(f_{\Theta^t}(\mathcal{Z})) - \mathcal{V}_2^t) \\ \Lambda_3^{t+1} = \Lambda_3^t + \mu(\nabla_x \nabla_z(f_{\Theta^t}(\mathcal{Z})) - \mathcal{V}_3^t) \\ \Lambda_4^{t+1} = \Lambda_4^t + \mu(\nabla_y \nabla_z(f_{\Theta^t}(\mathcal{Z})) - \mathcal{V}_4^t). \end{cases} \tag{16}$$

The ADMM algorithm is summarized in Algorithm 1. The computational complexity of $\mathcal{V}$ subproblems, $\mathcal{S}$ subproblem, and Λ updating at each iteration is $\mathcal{O}(HWB)$, where H, W, B, respectively, denote the height, width, and the number of spectral bands of the noisy HSI.

Algorithm 1: HSI Mixed Noise Removal Using S2DIP.

Input: Noisy HSI $\mathcal{Y}$, maximum iteration number $t_{max} = 7000$, tolerance $r = 0.01$;

Initialization: Randomly initialize Θ, $\Lambda_i = \mathbf{0}$, $t = 0$, $r' = r$;

1:   **while** $t \leq t_{max}$ and $r' \geq r$ **do**
2:     Update $\mathcal{V}_i$ $(i = 1, 2, 3, 4)$ via (11);
3:     Update $\mathcal{S}$ via (14);
4:     Update Θ via (15);
5:     Update Λ_i $(i = 1, 2, 3, 4)$ via (16);
6:     $r' = \frac{\|f_{\Theta^{t+1}}(\mathcal{Z}) - f_{\Theta^t}(\mathcal{Z})\|_F^2}{\|f_{\Theta^t}(\mathcal{Z})\|_F^2}$;
7:     $t = t + 1$;
8:   **end while**

Output: The denoising HSI $\mathcal{X} = f_{\Theta}(\mathcal{Z})$;

The computational complexity of Θ subproblem at each iteration for forward propagation is $\mathcal{O}(\sum_{i=1}^{D}(HWn_u[i-1]n_u[i]k_u[i] + 2HWn_d[i-1]n_d[i]k_d[i] + HWn_s[i-1]n_s[i]k_s[i]))$, where $n_u[i], n_d[i], n_s[i]$ $(i = 0, 1, 2, \ldots, D)$ denote the number of data channels at the ith layer for the upsampling, downsampling, and skip connections, respectively. $k_u[i], k_d[i], k_s[i]$ $(i = 1, 2, \ldots, D)$ denote the corresponding sizes of the convolutional kernels.

E. Automatic Stopping Criterion

We introduce an automatic stopping criterion to demonstrate that the proposed S2DIP can be automatically terminated based on its convergence property. Specifically, we use the relative error

$$\mathbf{RelErr} = \frac{\|f_{\Theta^{t+1}}(\mathcal{Z}) - f_{\Theta^t}(\mathcal{Z})\|_F^2}{\|f_{\Theta^t}(\mathcal{Z})\|_F^2} \qquad (17)$$

and the iteration step to evaluate the convergence degree. Here, $f_{\Theta^t}(\mathcal{Z}) \in \mathbb{R}^{H \times W \times B}$ denotes the tth network output. We set two parameters r and t_{max} as the tolerance of $\mathbf{RelErr}$ and the maximum number of iteration steps, respectively. The iteration will be terminated if: 1) $\mathbf{RelErr}$ of the tth iteration is lower than r; or 2) the iteration number t exceeds t_{max}. Cooperatively, $\mathbf{RelErr}$ evaluates the convergent degree of the optimization process while t_{max} guarantees the final stopping of the iteration.

We use S2DIP* to denote the highest peak signal-to-noise ratio (PSNR) value of the proposed method and use S2DIP to denote the proposed method using the automatic stopping criterion. If the performance of S2DIP is comparable to that of S2DIP*, it means that S2DIP converges to the point with the highest PSNR value, which alleviates the semiconvergence of DIP-based methods. It is worth emphasizing that the original DIP-based methods could not use such stopping criterion since DIP-based methods suffers from semiconvergence. Using such stopping criterion for DIP-based methods would result in poor performance, as DIP always converges to a noisy HSI. For convenience, we report the highest PSNR value of DIP-based methods in the experiments.

TABLE I
SIMULATED NOISY SETTING. σ DENOTES THE STANDARD DEVIATION. p DENOTES THE SAMPLING RATE. s_1 AND s_2 DENOTE THE NUMBERS OF STRIPES AND DEADLINES IN EACH CORRUPTED BAND, RESPECTIVELY

Case	Gaussian noise		Impulse noise		Stripe		Deadline	
	added band	σ	added band	p	added band	s_1	added band	s_2
Case 1	all bands	0.2	--	--	--	--	--	--
Case 2	all bands	0.1	all bands	0.1	--	--	--	--
Case 3	all bands	0.1	all bands	0.1	40%	[6, 15]	--	--
Case 4	all bands	0.1	all bands	0.1	--	--	50%	[6, 10]
Case 5	all bands	0.1	all bands	0.1	40%	[6, 15]	50%	[6, 10]

IV. EXPERIMENTS

In this section, both simulation experiments and real data experiments are involved to verify the effectiveness of the proposed method. The competing methods include the three model-based methods and three deep learning-based methods, i.e., the matrix low-rankness-based method LRMR [26], the tensor low-rankness and total variation-based method LRTDTV [27], the ℓ_0-ℓ_1 hybrid total variation method ℓ_0-ℓ_1 HTV [48], the supervised deep CNN-based method HSID-CNN [44], and the unsupervised DIP-based methods that based on 2-D convolution and 3-D convolution (termed as DIP 2D and DIP 3D) [1], [47]. The hyperparameters of model-based methods are tuned based on their best PSNR values. The results of HSID-CNN are obtained by using the pretrained model provided by the authors.

For simulation experiments, we adopt sub-images of HSIs *WDC mall*, *Pavia*, *Indian pines*, and *Salinas*[1] with size $256 \times 256 \times 32$, $192 \times 192 \times 32$, $128 \times 128 \times 32$, and $192 \times 192 \times 32$, respectively, to generate simulated noisy HSIs. Additionally, multispectral images (MSIs) including *Trash*, *Board*[2], *Balloons*, and *Cups*[3] [73] are adopted to verify the generalization ability of our method. The size of *Trash* and *Board* is $256 \times 256 \times 32$ and the size of the *Ballons* and *Cups* is $256 \times 256 \times 31$. Due to the limitation of our computer capability (A single RTX 2080 GPU), we consider 32 bands in the experiments. For all selected images, five noisy cases are established (see Table I). We consider mixed noise including Gaussian noise, impulse noise, stripe noise, and deadlines. The stripe noise and deadlines are often caused by miscalibration in the detectors of imaging systems [8]. Based on the noisy setting in Table I, all noise is randomly performed on the ground-truth (GT) images. For cases with deadlines, we change the fidelity term of our method to $\|\mathcal{M} \odot (\mathcal{Y} - \mathcal{X}) - \mathcal{S}\|_F^2$, where $\mathcal{M}$ denotes the mask whose entries on the positions of deadlines are set to 0 and other entries are set to 1, i.e., the removal of deadlines can be seen as an inpainting problem, which can be well addressed by the deep prior of the CNN [1], [47]. Here, $\odot$ denotes the element-wise product.

Next, we introduce the hyperparameters setting. There are three hyperparameters in our model, i.e., α_1, α_2, and α_3. We

---

TABLE II
QUANTITATIVE RESULTS BY DIFFERENT METHODS ON HSIS. THE BEST VALUE IS HIGHLIGHTED BY BOLDFACE. THE SECOND-BEST VALUE IS HIGHLIGHTED BY UNDERLINED

Data	Method	Case 1			Case 2			Case 3			Case 4			Case 5			Average time (s)
		PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	
WDC mall	LRMR	26.018	0.767	0.563	29.869	0.887	0.335	27.090	0.835	0.461	29.321	0.880	0.358	26.658	0.828	0.487	16
	LRTDTV	30.658	0.899	0.218	31.625	0.917	0.236	31.433	0.905	0.271	30.965	0.904	0.272	30.615	0.889	0.311	116
	ℓ_0-ℓ_1 HTV	28.762	0.844	0.156	30.320	0.882	0.122	30.032	0.878	0.139	29.080	0.849	0.152	28.955	0.845	0.159	78
	HSID-CNN	31.181	0.918	0.210	25.222	0.792	0.234	23.237	0.740	0.306	25.424	0.791	0.244	23.391	0.737	0.313	185
	DIP 2D	27.763	0.826	0.265	24.621	0.727	0.172	22.753	0.642	0.240	24.591	0.737	0.202	22.607	0.617	0.221	390
	DIP 3D	27.784	0.828	0.264	24.714	0.744	0.210	23.306	0.689	0.244	24.553	0.731	0.199	23.116	0.675	0.225	502
	S2DIP	31.560	0.920	**0.112**	32.503	0.931	**0.102**	32.333	0.930	0.112	32.585	0.932	**0.102**	32.004	0.924	**0.111**	512
	S2DIP*	**31.740**	**0.922**	0.113	**32.886**	**0.937**	0.105	**32.620**	**0.933**	0.110	**32.662**	**0.933**	0.104	**32.442**	**0.931**	0.116	512
Pavia	LRMR	25.282	0.751	0.259	29.467	0.886	0.153	27.040	0.836	0.240	28.050	0.863	0.193	25.843	0.811	0.269	6
	LRTDTV	30.452	0.901	0.103	31.634	0.924	0.105	31.243	0.907	0.131	30.105	0.900	0.148	29.417	0.878	0.175	15
	ℓ_0-ℓ_1 HTV	28.739	0.855	0.079	30.475	0.897	**0.058**	30.044	0.893	0.070	28.757	0.862	0.088	28.516	0.858	0.102	39
	HSID-CNN	29.905	0.906	0.114	26.430	0.856	0.124	24.307	0.815	0.168	26.513	0.842	0.146	24.611	0.797	0.191	46
	DIP 2D	27.836	0.845	0.118	25.472	0.808	0.112	23.797	0.740	0.124	25.480	0.811	0.098	23.738	0.740	0.128	207
	DIP 3D	27.781	0.840	0.141	25.844	0.822	0.111	24.239	0.782	0.109	25.571	0.810	0.106	24.277	0.784	0.122	271
	S2DIP	31.570	0.928	0.062	32.790	0.943	0.058	32.554	**0.945**	0.063	32.839	0.943	0.060	32.366	0.940	0.065	275
	S2DIP*	**31.674**	**0.930**	**0.061**	**33.016**	**0.946**	0.058	**32.710**	**0.945**	0.061	**32.981**	**0.945**	0.058	**32.437**	**0.941**	**0.064**	275
Indian pines	LRMR	24.622	0.725	0.135	28.731	0.859	0.084	27.712	0.844	0.108	23.256	0.756	0.203	23.108	0.757	0.201	7
	LRTDTV	31.968	0.952	0.050	33.881	0.965	0.043	32.953	0.952	0.061	28.071	0.894	0.129	27.982	0.886	0.146	7
	ℓ_0-ℓ_1 HTV	30.386	0.948	0.070	34.880	0.980	0.039	33.502	0.975	0.050	31.764	0.975	0.072	31.379	**0.971**	0.074	19
	HSID-CNN	26.136	0.846	0.113	25.080	0.844	0.111	24.217	0.833	0.124	21.286	0.750	0.181	21.239	0.745	0.190	45
	DIP 2D	27.113	0.882	0.088	26.193	0.874	0.078	24.700	0.832	0.122	25.886	0.867	0.077	24.434	0.831	0.112	105
	DIP 3D	26.972	0.864	0.092	26.333	0.863	0.076	25.356	0.847	0.089	26.159	0.869	0.069	25.150	0.850	0.090	124
	S2DIP	32.410	0.967	0.040	35.120	**0.984**	0.035	34.009	0.983	0.040	34.584	0.983	0.037	33.522	0.978	0.043	138
	S2DIP*	**32.710**	**0.968**	**0.037**	**35.364**	**0.984**	**0.034**	**34.860**	**0.984**	0.038	**35.181**	**0.984**	0.037	**33.769**	0.978	0.042	138
Salinas	LRMR	26.088	0.652	0.148	29.786	0.808	0.098	27.287	0.757	0.172	28.354	0.782	0.166	25.422	0.717	0.228	6
	LRTDTV	33.470	0.904	0.074	34.360	0.905	0.075	33.471	0.887	0.091	32.013	0.871	0.133	31.915	0.847	0.165	15
	ℓ_0-ℓ_1 HTV	32.722	0.919	0.058	35.670	0.949	0.040	34.416	0.941	0.051	33.925	0.943	0.058	32.912	0.933	**0.069**	47
	HSID-CNN	24.602	0.713	0.244	23.273	0.673	0.266	22.040	0.632	0.291	22.824	0.653	0.287	21.270	0.596	0.317	47
	DIP 2D	31.771	0.891	0.069	27.592	0.812	0.116	25.885	0.787	0.152	27.527	0.812	0.114	25.492	0.782	0.152	208
	DIP 3D	32.729	0.916	0.058	28.141	0.840	0.109	25.983	0.807	0.134	28.211	0.846	0.104	26.185	0.806	0.141	269
	S2DIP	35.068	0.945	**0.039**	35.997	0.954	**0.036**	35.123	0.950	0.042	36.056	0.956	**0.034**	35.084	0.948	0.043	277
	S2DIP*	**35.224**	**0.946**	0.041	**36.223**	**0.956**	0.037	**35.474**	0.952	0.041	**36.212**	**0.957**	0.036	**35.259**	0.951	0.043	277
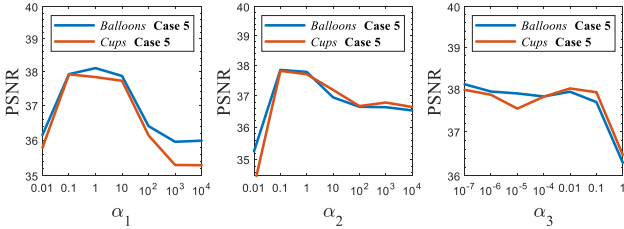
Fig. 4. PSNR values with respect to different values of α_1, α_2, and α_3.

evaluate the influence of each hyperparameter by changing it and fixing the other hyperparameters [55]. The results are shown in Fig. 4. We can observe that our method is more sensitive to α_1 and α_2. Meanwhile, our method can obtain satisfactory results for a wide range of $\alpha_i(i = 1, 2, 3)$. In the experiments, we set the hyperparameters $\alpha_1 = 0.1$, $\alpha_2 = 0.1$, and $\alpha_3 = 0.01$ for Case 2–5. For Case 1, we change $\alpha_3 = 10$ because there is no sparse noise in Case 1. In addition, we empirically set the tolerance of **RelErr** (i.e., r) as 0.01 and t_{max} as 7000, which provides good termination points.

In simulation experiments, the denoising results are numerically evaluated by PSNR, structure similarity (SSIM), and spectral angle mapper (SAM). It is worth noting that higher PSNR and SSIM values represent better performance while lower SAM values represent better performance.

In real experiments, noisy HSIs *Urban* of size $288 \times 288 \times 210$ with its band 134 to band 165 and *Indian*[4] of size $128 \times 128 \times 220$ with its band 1 to band 32 are included. The

hyperparameters of our method for real experiments are $\alpha_1 = 0.1, \alpha_2 = 0.1$, and $\alpha_3 = 0.01$.

All the experiments are conducted on the platform of Windows 10 with Intel Core i5-9400f CPU, Nvidia RTX 2080 GPU, and 16 GB RAM.

A. Simulation Experiments

The numerical results on HSIs are illustrated in Table II. We can discover that S2DIP* outperforms competing methods in terms of PSNR. S2DIP is also competitive with PSNR a little lower than S2DIP*, which shows that S2DIP has convergence property to stably handle the mixed noise removal under unsupervised conditions, which avoids the semiconvergence of DIP-based methods.

It is notable that the SAM values of the proposed method are also superior to competing methods. This mainly attributes to the spatial-spectral constraint of the proposed framework, where high-quality spatial-spectral correlation and spectral fidelity are ensured.

In Fig. 5, the denoising results on HSIs for Case 1 are displayed. Since Case 1 only contains Gaussian noise, most of the methods can remove the noise well and obtain good visual quality, while S2DIP achieves better PSNR values. We subsequently illustrate the denoising results on HSIs for Case 5 in Fig. 6, as Case 5 contains the most complex noise. We can see that LRMR, which delivers the low-rankness by matrix, is hard to

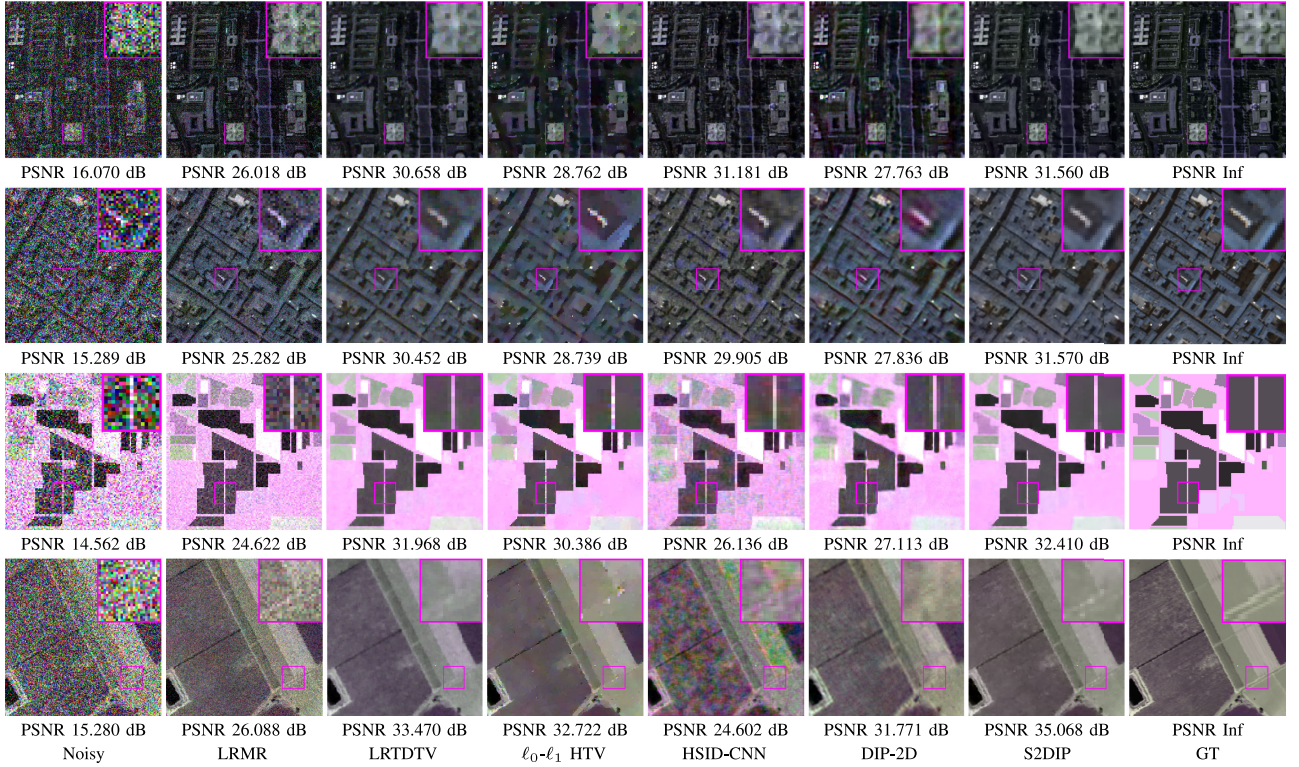[4][Online]. Available: https://purr.purdue.edu/publications/1947/1

Fig. 5. Denoising HSIs by different methods for Case 1. Each row from top to down lists *WDC mall* consisted of the 5th, 15th, and 30th bands, *Pavia* consisted of the 5th, 15th, and 30th bands, *Indian pines* consisted of the 15th, 25th, and 29th bands, and *Salinas* consisted of the 18th, 25th, and 32th bands.
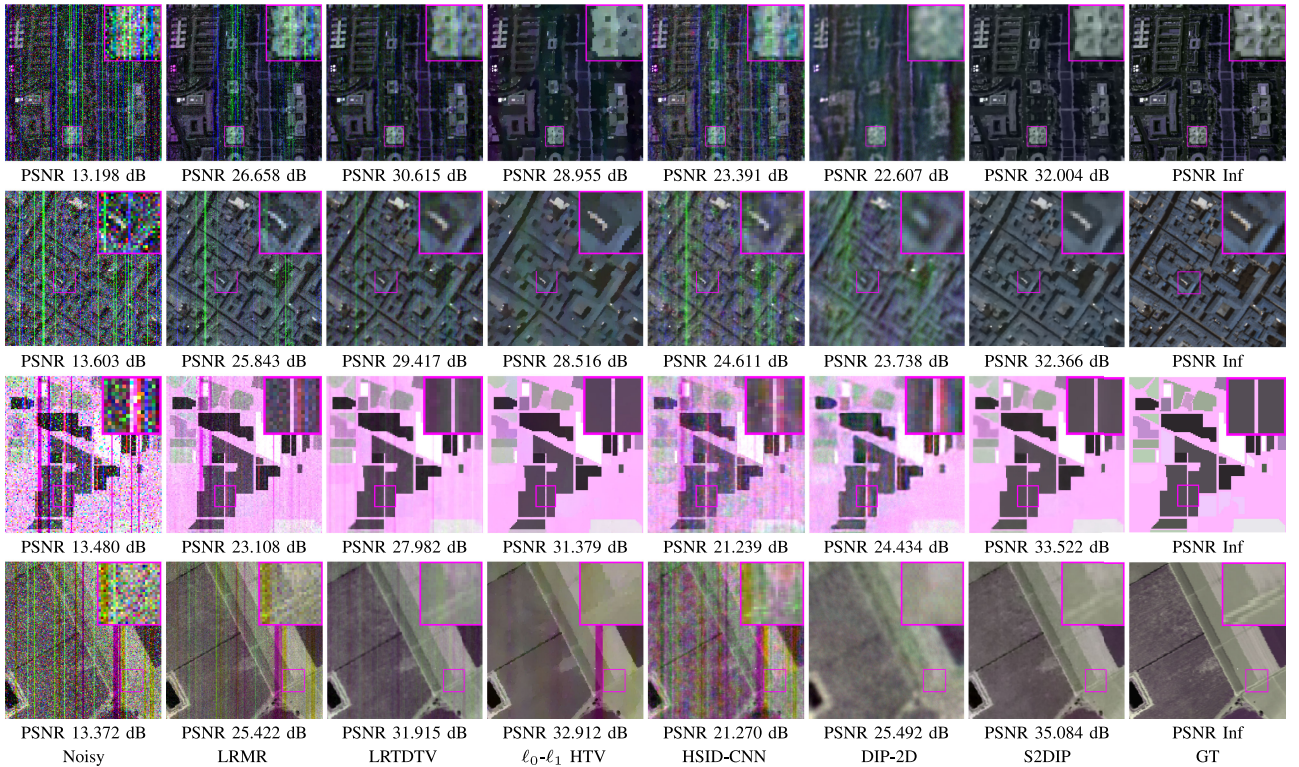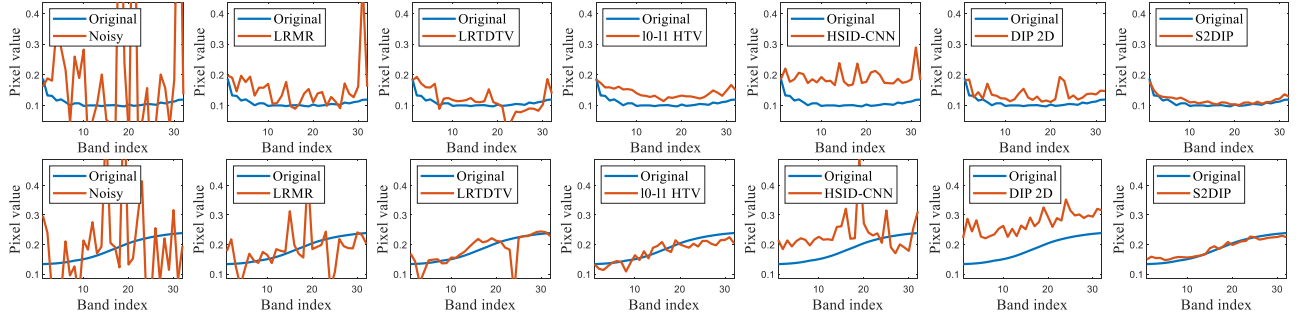

Fig. 6. Denoising HSIs by different methods for Case 5. Each row from top to down lists *WDC mall* consisted of the 5th, 15th, and 30th bands, *Pavia* consisted of the 5th, 15th, and 30th bands, *Indian pines* consisted of the 15th, 25th, and 29th bands, and *Salinas* consisted of the 18th, 25th, and 32nd bands.

Fig. 7. Spectral curves of the recovered results by different methods on HSIs *WDC mall* and *Pavia* for Case 5.
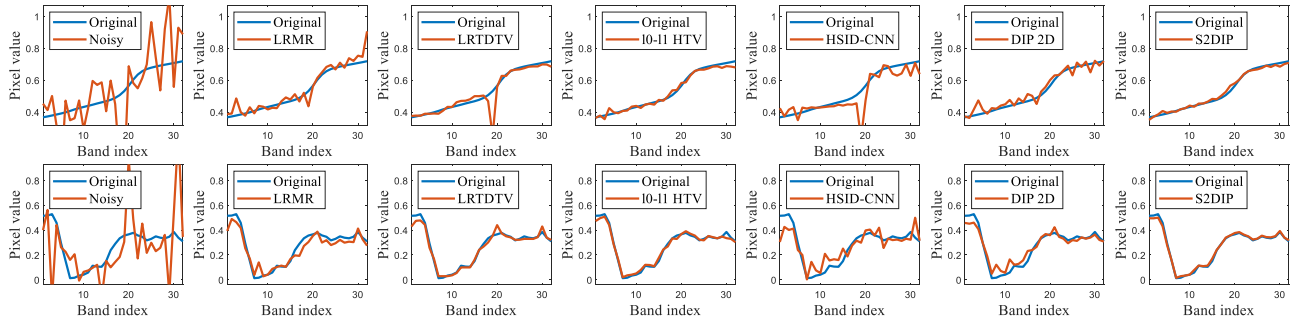
Fig. 8. Spectral curves of the recovered results by different methods on HSIs *Indian pines* and *Salinas* for Case 5.

totally remove the mixed noise. LRTDTV considers the tensor low-rankness, thus it achieves noise removal in partial bands, but also fails to totally remove the stripes and impulses. ℓ_0-ℓ_1 HTV removes the noise but produces spatial over-smoothness due to the emphasis of the TV. The supervised HSID-CNN is trained on data with only Gaussian noise. Thus, it is hard to deal with strong complex simulated noise. DIP 2D and DIP 3D are hard to fit the signal part of the observation before fitting the complex noise. Thus, DIP-based methods could not well recover clean HSIs. The proposed S2DIP achieves the best performance compared with competing methods, where the mixed noise is considerably removed and the image details are well preserved. The good visual quality of the denoising results of S2DIP can attribute to the combination of the expressive power of the CNN and the hand-crafted priors, which ensures the recovery quality.

Next, we display the spectral curves of the recovered results by different methods on HSIs in Figs. 7 and 8. The less oscillating spectral curves of the denoising results by S2DIP verify that the spectral fidelity is well-preserved, which outperforms competing methods. The integration of deep prior brought by the CNN and the spatial-spectral constraint by the SSTV regularization contributes to this phenomenon.

Then, we illustrate the denoising quantitative results by different methods on MSIs in Table III. We can see that S2DIP and S2DIP* considerably outperforms competing methods, which verify the generalization ability and effectiveness of the proposed method for various data.

The denoising results by different methods on MSIs for Case 5 are displayed in Fig. 9. We can observe that S2DIP recovers the

MSI and removes the complex noise well. LRMR and LRTDTV could not totally remove the noise. ℓ_0-ℓ_1 HTV removes the noise but suffers from over-smoothness. The HSID-CNN is trained on Gaussian noisy HSIs, and thus, it is hard to deal with such mixed noise in MSIs. As compared, S2DIP has a better generalization ability for mixed noise in various types of data. DIP-based methods remove the noise but miss some image details and edges. Note that the results of DIP-based methods are selected by referring to the GT images with the highest PSNR values. In contrast, the results of S2DIP are automatically selected based on the stopping criterion. This verifies that the proposed S2DIP favorably addresses the semiconvergence of DIP-based methods and largely enhances the denoising ability of DIP-based methods.

Finally, we plot the spectral curves of the recovered MSIs by different methods in Figs. 10 and 11. We can see that S2DIP better preserves the nonlinear spectral curves. This could attribute to the integration of hand-crafted priors and the DIP, where the nonlinear modeling ability of CNN and the spatial-spectral constraint are well combined.

B. Real Experiments

The denoising results for real-world noisy HSI *Urban* are displayed in Fig. 12. We can see that S2DIP shows better performance for removing the complex noise. Specifically, S2DIP removes the noise and preserves the image details well. LRMR and LRTDTV could not totally remove the mixed noise. ℓ_0-ℓ_1 HTV and DIP-based methods have relatively over-smooth

TABLE III
QUANTITATIVE RESULTS BY DIFFERENT METHODS ON MSIs. THE BEST VALUE IS HIGHLIGHTED BY BOLDFACE. THE SECOND-BEST VALUE IS HIGHLIGHTED BY UNDERLINED

| Case		Case 1			Case 2			Case 3			Case 4			Case 5			Average time (s)
Data	Method	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	
Trash	LRMR	25.437	0.642	0.302	29.548	0.820	0.170	27.210	0.780	0.261	27.898	0.804	0.202	25.714	0.753	0.304	12
	LRTDTV	29.812	0.922	0.145	32.088	0.932	0.130	31.300	0.916	0.154	30.939	0.915	0.157	30.377	0.903	0.176	35
	ℓ_0-ℓ_1 HTV	32.391	0.947	0.095	35.508	0.971	0.067	34.618	0.968	0.078	33.954	0.966	0.096	33.628	0.965	0.099	81
	HSID-CNN	28.389	0.857	0.168	25.111	0.785	0.183	23.727	0.749	0.219	24.857	0.776	0.196	23.325	0.726	0.235	191
	DIP 2D	31.291	0.922	0.105	26.985	0.865	0.131	25.530	0.850	0.145	27.145	0.877	0.109	25.211	0.846	0.142	389
	DIP 3D	29.895	0.931	0.134	26.909	0.901	0.129	24.979	0.873	0.141	26.806	0.891	0.137	24.820	0.861	0.155	497
	S2DIP	33.352	0.971	0.078	35.549	0.978	0.064	34.891	0.976	0.068	35.488	0.978	0.064	34.694	0.975	0.073	512
	S2DIP*	33.537	0.972	0.078	35.716	0.978	0.062	35.112	0.977	0.069	35.690	0.978	0.064	34.985	0.976	0.071	512
Borad	LRMR	25.610	0.687	0.333	29.552	0.841	0.191	27.256	0.803	0.274	27.596	0.820	0.227	25.622	0.778	0.330	12
	LRTDTV	28.938	0.904	0.137	31.032	0.917	0.150	30.645	0.908	0.166	30.150	0.901	0.179	29.630	0.884	0.218	30
	ℓ_0-ℓ_1 HTV	30.676	0.908	0.112	32.953	0.940	0.091	32.395	0.935	0.092	31.450	0.925	0.104	31.340	0.923	0.111	82
	HSID-CNN	28.019	0.856	0.178	24.984	0.785	0.188	23.312	0.743	0.220	24.444	0.769	0.208	23.031	0.729	0.247	192
	DIP 2D	30.258	0.888	0.136	26.765	0.842	0.134	25.226	0.814	0.138	26.886	0.840	0.137	25.167	0.811	0.160	384
	DIP 3D	29.768	0.899	0.145	26.886	0.859	0.149	25.297	0.834	0.156	26.562	0.860	0.137	25.225	0.833	0.164	497
	S2DIP	31.811	0.939	0.102	33.600	0.956	0.085	33.272	0.955	0.091	33.390	0.956	0.085	33.188	0.953	0.090	517
	S2DIP*	31.925	0.939	0.103	33.712	0.957	0.086	33.383	0.955	0.090	33.634	0.957	0.085	33.289	0.955	0.091	517
Balloons	LRMR	28.311	0.794	0.223	27.891	0.734	0.291	27.618	0.758	0.299	26.801	0.729	0.315	28.938	0.801	0.254	11
	LRTDTV	33.928	0.949	0.235	35.406	0.948	0.193	34.788	0.937	0.203	34.359	0.944	0.210	34.154	0.933	0.226	35
	ℓ_0-ℓ_1 HTV	33.643	0.944	0.175	36.832	0.964	0.122	35.632	0.955	0.139	35.466	0.966	0.132	35.252	0.959	0.160	81
	HSID-CNN	28.452	0.839	0.317	24.586	0.680	0.338	22.761	0.625	0.366	24.130	0.669	0.353	22.659	0.624	0.408	189
	DIP 2D	33.989	0.948	0.189	27.114	0.822	0.224	25.204	0.786	0.262	27.182	0.818	0.239	25.450	0.791	0.283	414
	DIP 3D	33.343	0.940	0.217	27.257	0.817	0.242	25.110	0.773	0.254	27.033	0.824	0.241	25.319	0.786	0.275	508
	S2DIP	36.610	0.978	0.117	38.482	0.980	0.101	37.888	0.976	0.113	38.216	0.978	0.104	37.367	0.971	0.125	569
	S2DIP*	37.435	0.980	0.115	39.152	0.985	0.094	38.082	0.976	0.114	38.828	0.983	0.100	37.881	0.973	0.121	569
Cups	LRMR	27.042	0.752	0.071	28.417	0.770	0.111	29.147	0.812	0.093	26.023	0.734	0.159	28.783	0.834	0.092	11
	LRTDTV	33.804	0.963	0.079	35.908	0.961	0.064	35.601	0.959	0.066	34.547	0.950	0.076	33.867	0.942	0.082	36
	ℓ_0-ℓ_1 HTV	33.296	0.953	0.063	36.980	0.976	0.040	36.164	0.974	0.047	35.545	0.976	0.053	34.726	0.974	0.060	82
	HSID-CNN	28.739	0.854	0.109	26.626	0.827	0.124	25.617	0.809	0.140	25.418	0.795	0.155	24.483	0.770	0.175	182
	DIP 2D	32.473	0.942	0.076	29.810	0.930	0.075	27.709	0.916	0.098	28.963	0.926	0.072	27.801	0.912	0.106	407
	DIP 3D	32.254	0.940	0.074	29.403	0.929	0.075	28.183	0.910	0.093	29.440	0.927	0.077	27.844	0.920	0.081	511
	S2DIP	36.376	0.983	0.036	38.560	0.989	0.030	37.882	0.988	0.035	38.169	0.989	0.031	37.834	0.988	0.034	574
	S2DIP*	36.469	0.983	0.037	38.766	0.989	0.031	38.160	0.988	0.033	38.605	0.989	0.032	38.147	0.988	0.034	574
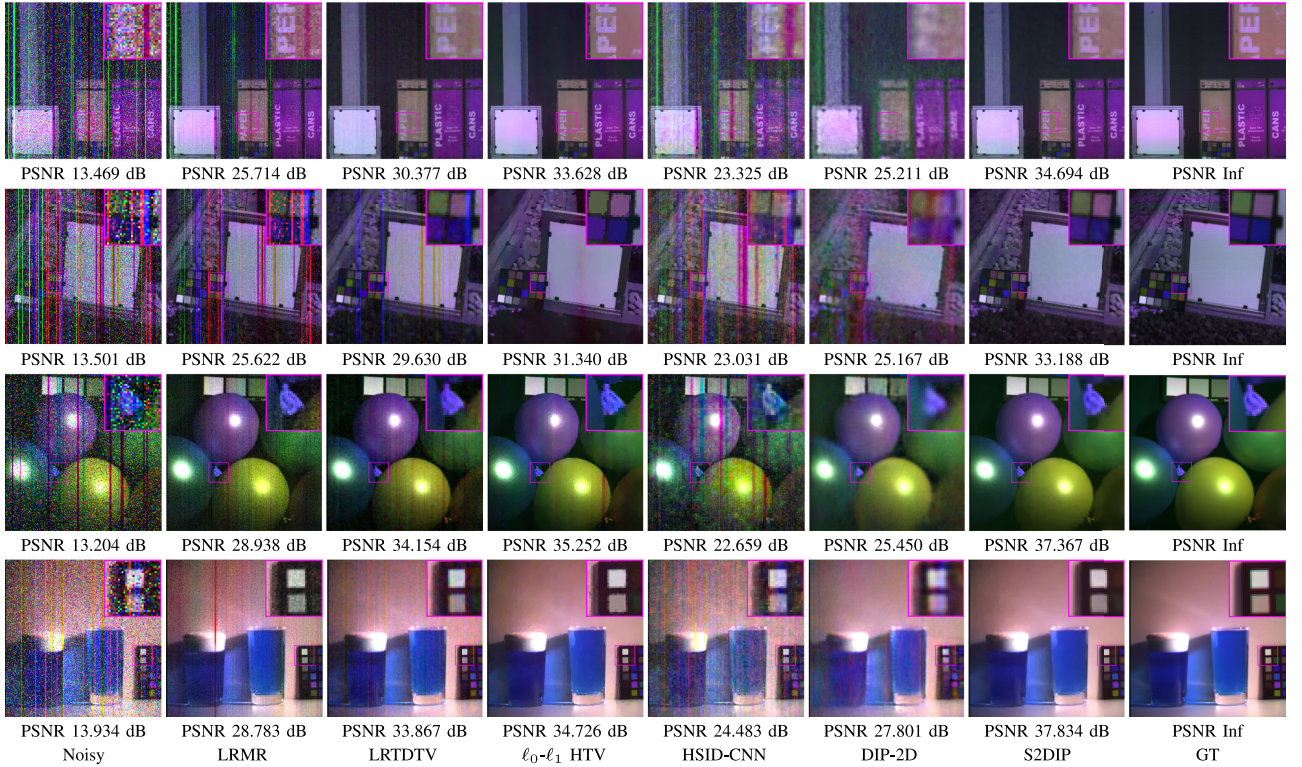
Fig. 9. Denoising MSIs by different methods for Case 5. Each row from top to down lists *Trash* consisted of the 5th, 15th, and 25th bands, *Borad* consisted of the 5th, 15th, and 25th bands, *Balloons* consisted of the 1st, 10th, and 30th bands, and *Cups* consisted of the 1st, 10th, and 30th bands.
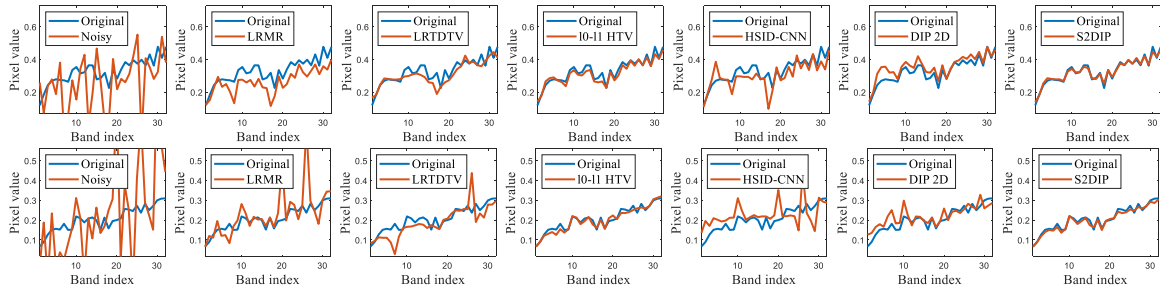
Fig. 10. Spectral curves of the recovered results by different methods on MSIs *Trash* and *Borad* for Case 5.
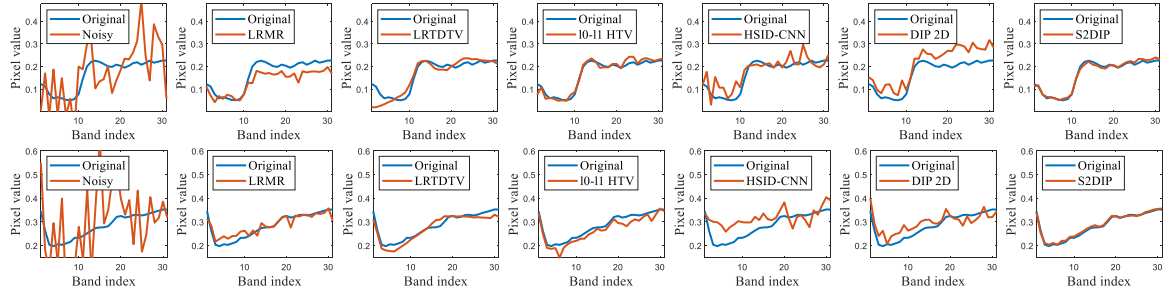
Fig. 11. Spectral curves of the recovered results by different methods on MSIs *Balloons* and *Cups* for Case 5.
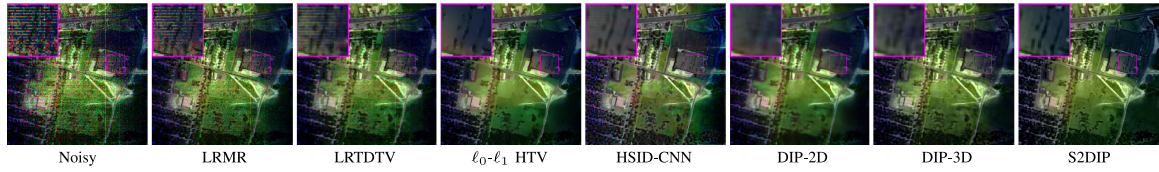
Fig. 12. Recovered results by different methods for real-world noisy HSI *Urban* (The pseudoimages consisted of the 139th, 150th, and 151th bands).
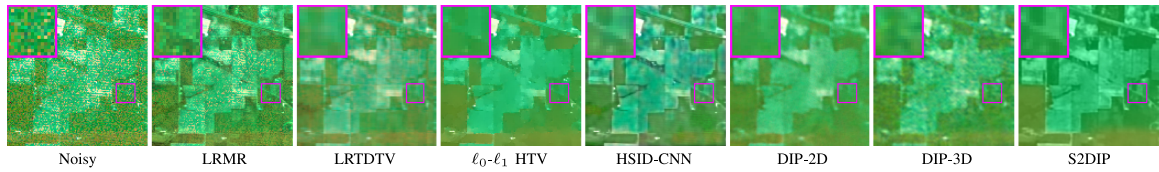
Fig. 13. Recovered results by different methods for real-world noisy HSI *Indian* (The pseudoimages consisted of the 1st, 2nd, and 3rd bands).
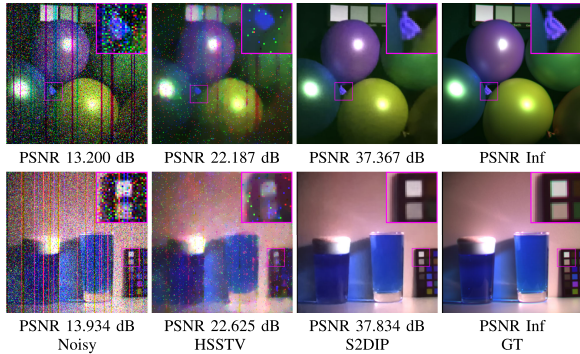
Fig. 14. Denoising MSIs *balloons* and *cups* consisted of the 1st, 10th, and 30th bands by different methods.

results. HSID-CNN achieves considerable results but the results remain some stripe noise.

The denoising results for *Indian* are illustrated in Fig. 13. We can see that LRTDTV and DIP 2D both remove the strong noise but miss some image details. HSID-CNN removes the complex noise and achieves good results. S2DIP also successfully and reasonably restores the image structure, where the complex noise is removed and the image details and colors are well preserved.

The good performance of S2DIP for real noise removal is mainly due to the combination of deep prior and hand-crafted priors. The natural image structure can be well preserved by the DIP and the mixed noise can be fully removed by the spatial-spectral constraint and the robust noise modeling.

We can find that in real experiments, the performance of HSID-CNN is much better than that of simulation experiments. This is because that the number of noisy bands in real noisy HSIs is relatively small. HSID-CNN can explore the spatial-spectral information from other clean bands for denoising. The proposed

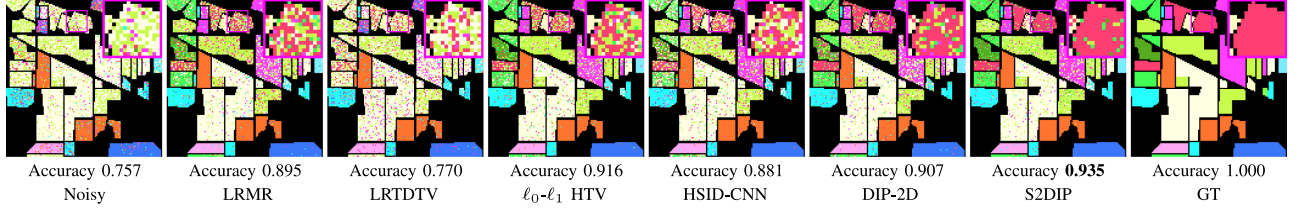| Accuracy 0.757	Accuracy 0.895	Accuracy 0.770	Accuracy 0.916	Accuracy 0.881	Accuracy 0.907	Accuracy **0.935**	Accuracy 1.000
Noisy	LRMR	LRTDTV	ℓ_0-ℓ_1 HTV	HSID-CNN	DIP-2D	S2DIP	GT

Fig. 15. HSI classification results using SVM on the denoising results by different denoising methods. Different colors refer to different classifications. The best accuracy rate is highlighted by boldface.
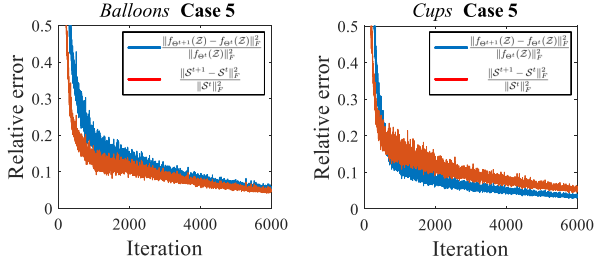

Fig. 16. Relative error of variables with respect to iteration.

TABLE IV
QUANTITATIVE RESULTS BY DIFFERENT METHODS. S2DIP WO. SPARSITY DENOTES THE PROPOSED METHOD WITHOUT THE SPARSE TERM. S2DIP WO. SSTV DENOTES THE PROPOSED METHOD WITHOUT THE TV AND SSTV REGULARIZATION TERM

Method	*Balloons* **Case 5**			*Cups* **Case 5**		
	PSNR	SSIM	SAM	PSNR	SSIM	SAM
S2DIP wo. sparsity	26.116	0.816	0.258	28.952	0.955	0.076
S2DIP wo. SSTV	34.775	0.949	0.167	33.729	0.958	0.059
S2DIP	**37.367**	**0.971**	**0.125**	**37.834**	**0.988**	**0.034**

unsupervised S2DIP could achieve even better results than the supervised HSID-CNN for real noise removal, which illustrates the effectiveness of our method in real scenarios.

V. DISCUSSIONS

A. Effectiveness of Deep Prior

Our method simultaneously uses the unsupervised deep prior, the spatial-spectral image prior, and the sparse noise prior. In this section, we illustrate the effectiveness of the deep prior. Specifically, we remove the deep prior of CNN in the proposed model (7) to clarify its influence. Here, the underlying clean HSI could be updated via ADMM. Note that the proposed method degenerates to the hybrid spatial-spectral total variation (HSSTV) [18] method without the deep prior. The HSSTV considers the TV and SSTV on the clean HSI and remove the sparse noise via minimizing its ℓ_1-norm. The difference between HSSTV and the proposed method is that we additionally use the DIP while HSSTV just employs TV and SSTV regularizers in a traditional optimization model.

We conduct the comparison between HSSTV and S2DIP on *Balloons* and *Cups* for Case 5. Fig. 14 shows the denoising results. We can see that S2DIP removes the complex noise and preserves the image details and edges well. In contrast, HSSTV could not completely remove the impulse noise and stripes, while the image details are also missed. This verifies the effectiveness of the deep prior, which can capture the natural image structure. S2DIP faithfully combines the deep prior of CNN and the hand-crafted prior, taking both advantages of the high representation ability of CNN and the spatial-spectral constraint delivered by the SSTV in an unsupervised manner.

B. Effectiveness of Hand-Crafted Priors

To address the semiconvergence of DIP-based methods and improve their denoising performance, we suggest two hand-crafted priors in DIP, i.e., the SSTV regularization term and the sparse term. This section verifies their effectiveness. Specifically, we compare the proposed method with and without the SSTV regularization term and the sparse term to clarify their influence. The results are illustrated in Table IV. We can see that S2DIP considerably outperforms S2DIP wo. SSTV and S2DIP wo. sparsity, which verifies the effectiveness of the hand-crafted priors in the DIP framework. The deep prior, the spatial-spectral local smooth prior, and the sparse noise prior are organically combined to handle the mixed noise removal in HSI. Notably, S2DIP wo. sparsity has relatively weak performance, which shows that the sparse modeling of the complex noise is very essential. In fact, the mixed noise in Case 5 contains many sparse outliers, which can not be fully removed if only considering Gaussian noise. Thus, the sparse term can largely improve the effectiveness of the unsupervised DIP-based methods for the HSI mixed noise removal.

C. Influence on Subsequent Applications

This section verifies that using S2DIP for HSI denoising can improve the performance of subsequent applications. We consider the HSI classification [6] using the support vector machine (SVM). We use the denoising results of the real-world noisy HSI *Indian* by different methods to conduct the HSI classification. 10% of the samples are set as training sets while others are testing sets. The results are shown in Fig. 15. We can see that the classification result on the denoising HSI by S2DIP has the best accuracy, which verifies the effectiveness and superiority of S2DIP over competing denoising methods.

D. Convergence Analysis

To test the convergence of the ADMM Algorithm 1, we plot the relative error of variables with respect to the iteration number in Fig. 16. The downward trend of the curves verifies the convergence behavior of our method, which addresses the semiconvergence of DIP-based methods. This is due to the SSTV and the sparse term which respectively considers the HSI spatial-spectral local smoothness and the sparsity of noise. Thus, the proposed method can accurately separate the noise and the clean HSI, resulting in stable convergence.

VI. CONCLUSION

In this article, we propose the S2DIP for the hyperspectral mixed noise removal. The SSTV regularization is considered to fully explore the HSI spatial-spectral local smooth prior to improve the effectiveness of unsupervised DIP-based methods. The ℓ_1-norm sparse term is considered to remove the sparse noise. To address the proposed model, we utilize the ADMM algorithm. Experimental results on synthetic data and real data demonstrate the effectiveness of the proposed method, which outperforms state-of-the-art methods. Meanwhile, the proposed method favorably alleviates the semiconvergence problem, which always exists in the DIP framework. We will combine the unsupervised CNN with model-based methods for more applications, especially for multidimensional image processing.

REFERENCES

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

[2] G. Wang, Y. Zhang, B. He, and K. T. Chong, "A framework of target detection in hyperspectral imagery based on blind source extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 835–844, Feb. 2016.

[3] Y. Liu, G. Gao, and Y. Gu, "Tensor matched subspace detector for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 1967–1974, Apr. 2017.

[4] S. Jia, L. Shen, J. Zhu, and Q. Li, "A 3-D gabor phase-based coding and matching framework for hyperspectral imagery classification," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1176–1188, Apr. 2018.

[5] N. Akhtar and A. Mian, "Nonparametric coupled Bayesian dictionary and classifier learning for hyperspectral classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4038–4050, Sep. 2018.

[6] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4604–4616, Jul. 2020.

[7] X. Zhao *et al.*, "Joint classification of hyperspectral and lidar data using hierarchical random walk and deep CNN architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7355–7370, Oct. 2020.

[8] N. Liu, W. Li, R. Tao, and J. E. Fowler, "Wavelet-domain low-rank/group-sparse destriping for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10310–10321, Dec. 2019.

[9] N. Liu, L. Li, W. Li, R. Tao, J. E. Fowler, and J. Chanussot, "Hyperspectral restoration and fusion with multispectral imagery via low-rank tensor-approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7817–7830, Sep. 2021.

[10] Y. Wang, L. Lin, Q. Zhao, T. Yue, D. Meng, and Y. Leung, "Compressive sensing of hyperspectral images via joint tensor tucker decomposition and weighted total variation regularization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2457–2461, Dec. 2017.

[11] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017.

[12] J. Yao, D. Hong, L. Xu, D. Meng, J. Chanussot, and Z. Xu, "Sparsity-enhanced convolutional decomposition: A novel tensor-based paradigm for blind hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2021.3069845.

[13] K. Wang, Y. Wang, X.-L. Zhao, J. C.-W. Chan, Z. Xu, and D. Meng, "Hyperspectral and multispectral image fusion via nonlocal low-rank tensor decomposition and spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7654–7671, Nov. 2020.

[14] J. Yang, X. Zhao, T. Ji, T. Ma, and T. Huang, "Low-rank tensor train for tensor robust principal component analysis," *Appl. Math. Comput.*, vol. 367, 2020, Art. no. 124783.

[15] X. Zhao, W. Xu, T. Jiang, Y. Wang, and M. K. Ng, "Deep plug-and-play prior for low-rank tensor completion," *Neurocomputing*, vol. 400, pp. 137–149, 2020.

[16] 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.

[17] 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.

[18] S. Takeyama, S. Ono, and I. Kumazawa, "Mixed noise removal for hyperspectral images using hybrid spatio-spectral total variation," in *Proc. IEEE Int. Conf. Image Process.*, 2019, pp. 3128–3132.

[19] J. Liu, Y. Sun, X. Xu, and U. S. Kamilov, "Image restoration using total variation regularized deep image prior," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 7715–7719.

[20] 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.

[21] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 2272–2279.

[22] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[23] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Jan. 2013.

[24] X. Gong, W. Chen, and J. Chen, "A low-rank tensor dictionary learning method for hyperspectral image denoising," *IEEE Trans. Signal Process.*, vol. 68, pp. 1168–1180, 2020.

[25] 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.*, 2014, pp. 2949–2956.

[26] 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.

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

[28] 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. 5901–5909.

[29] A. Karami, M. Yazdi, and A. Zolghadre 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.

[30] 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.

[31] 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.*, to be published, doi: 10.1109/TGRS.2020.3046038.

[32] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3040221.

[33] H. Zhang, X.-L. Zhao, T.-X. Jiang, N. Micahel, and T.-Z. Huang, "Multi-scale features tensor train minimization for multi-dimensional images recovery and recognition," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2021.3108847.

[34] T.-X. Jiang, M. K. Ng, X.-L. Zhao, and T.-Z. Huang, "Framelet representation of tensor nuclear norm for third-order tensor completion," *IEEE Trans. Image Process.*, vol. 29, pp. 7233–7244, 2020.

[35] T.-X. Jiang, X.-L. Zhao, H. Zhang, and M. K. Ng, "Dictionary learning with low-rank coding coefficients for tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2021.3104837.

[36] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.

[37] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," *Adv. Neural Inf. Process. Syst.*, vol. 29, pp. 2802–2810, 2016.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[39] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, Nov. 2018.

[40] Y.-T. Wang, X.-L. Zhao, T.-X. Jiang, L.-J. Deng, Y. Chang, and T.-Z. Huang, "Rain streaks removal for single image via kernel-guided convolutional neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3664–3676, Aug. 2021.

[41] P. Zhong and R. Wang, "Jointly learning the hybrid CRF and MLR model for simultaneous denoising and classification of hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1319–1334, Jul. 2014.

[42] W. Dong, H. Wang, F. Wu, G. Shi, and X. Li, "Deep spatial-spectral representation learning for hyperspectral image denoising," *IEEE Trans. Comput. Imag.*, vol. 5, no. 4, pp. 635–648, Dec. 2019.

[43] 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.

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

[45] Y. Chang, M. Chen, L. Yan, X.-L. Zhao, Y. Li, and S. Zhong, "Toward universal stripe removal via wavelet-based deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2880–2897, Apr. 2020.

[46] Y.-C. Miao, X.-L. Zhao, X. Fu, and J.-L. Wang, "Hyperspectral denoising using unsupervised disentangled spatio-spectral deep priors," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2021.3106380.

[47] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 9446–9454.

[48] M. Wang, Q. Wang, J. Chanussot, and D. Hong, "ℓ_0-ℓ_1 hybrid total variation regularization and its applications on hyperspectral image mixed noise removal and compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2021.3055516.

[49] R. Tao, X. Zhao, W. Li, H.-C. Li, and Q. Du, "Hyperspectral anomaly detection by fractional fourier entropy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4920–4929, Dec. 2019.

[50] C. Zhao, C. Li, and S. Feng, "A spectral-spatial method based on fractional fourier transform and collaborative representation for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 7, pp. 1259–1263, Jul. 2021.

[51] 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.

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

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

[54] J. Peng, Q. Xie, Q. Zhao, Y. Wang, L. Yee, and D. Meng, "Enhanced 3DTV regularization and its applications on HSI denoising and compressed sensing," *IEEE Trans. Image Process.*, vol. 29, pp. 7889–7903, 2020.

[55] T. Hu, W. Li, N. Liu, R. Tao, F. Zhang, and P. Scheunders, "Hyperspectral image restoration using adaptive anisotropy total variation and nuclear norms," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1516–1533, Feb. 2021.

[56] 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.

[57] H. Zhang, J. Cai, W. He, H. Shen, and L. Zhang, "Double low-rank matrix decomposition for hyperspectral image denoising and destriping," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2021.3061148.

[58] Y.-Y. Liu, X.-L. Zhao, Y.-B. Zheng, T.-H. Ma, and H. Zhang, "Hyperspectral image restoration by tensor fibered rank constrained optimization and plug-and-play regularization," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3045169.

[59] 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, Feb. 2020.

[60] M. Wang, Q. Wang, J. Chanussot, and D. Li, "Hyperspectral image mixed noise removal based on multidirectional low-rank modeling and spatial-spectral total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 488–507, Jan. 2021.

[61] 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, Mar. 2020.

[62] 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.*, to be published, doi: 10.1109/TGRS.2021.3085779.

[63] Y. Chen, T.-Z. Huang, and X.-L. Zhao, "Destriping of multispectral remote sensing image using low-rank tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4950–4967, Dec. 2018.

[64] J. Jia *et al.*, "Destriping algorithms based on statistics and spatial filtering for visible-to-thermal infrared pushbroom hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 4077–4091, Jun. 2019.

[65] J.-H. Yang, X.-L. Zhao, T.-H. Ma, Y. Chen, T.-Z. Huang, and M. Ding, "Remote sensing images destriping using unidirectional hybrid total variation and nonconvex low-rank regularization," *J. Comput. Appl. Math.*, vol. 363, pp. 124–144, 2020.

[66] 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.

[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.*, to be published, doi: 10.1109/TGRS.2021.3069241.

[69] Q. Shi, X. Tang, T. Yang, R. Liu, and L. Zhang, "Hyperspectral image denoising using a 3-D attention denoising network," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3045273.

[70] Q. Zhang, F. Sun, Q. Yuan, J. Li, H. Shen, and L. Zhang, "Combined the data-driven with model-driven strategy: A novel framework for mixed noise removal in hyperspectral image," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 2667–2670.

[71] Q. Zhang, Q. Yuan, J. Li, F. Sun, and L. Zhang, "Deep spatio-spectral bayesian posterior for hyperspectral image non-i.i.d. noise removal," *ISPRS J. Photogrammetry Remote Sens.*, vol. 164, pp. 125–137, 2020.

[72] Z. Qiu, T. Yao, and T. Mei, "Learning spatio-temporal representation with pseudo-3D residual networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5534–5542.

[73] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.

Yi-Si Luo is currently pursuing the bachelor's degree in information and computing sciences with the University of Electronic Science and Technology of China (UESTC), Chengdu, China.

His research interests include tensor modeling and image processing.

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

He worked as a Postdoc with Hong Kong Baptist University, Hong Kong from 2013 to 2014 with Prof. M. Ng. He worked as a Visiting Scholar with the University of Lisbon, Lisbon, Portugal, from 2016 to 2017 with Prof. J. Bioucas Dias. He is currently a Professor with the School of Mathematical Sciences, UESTC. His research interests include sparse and low-rank modeling for the low-level inverse problems of multidimensional images.

Yu-Bang Zheng (Student Member, IEEE) received the B.S. degree in information and computing science from the Anhui University of Finance and Economics, Bengbu, China, in 2017. He is currently pursuing the Ph.D. degree in mathematics with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China.

His current research interests include tensor modeling and computing, tensor learning, and high-dimensional data processing.
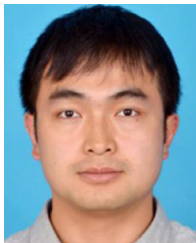
Tai-Xiang Jiang (Member, IEEE) received the Ph.D. degree in mathematics from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2019. He was a co-training Ph.D. student with the University of Lisbon, Lisbon, Portugal, supervised by Prof. Jose M. Bioucas-Dias from 2017 to 2018.

He was the Research Assistant with the Hong Kong Baptist University, Hong Kong, supported by Prof. Michael K. Ng in 2019. He is currently an Associated Professor with the School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, China. His research interests include sparse and low-rank modeling and tensor decomposition for multidimensional image processing, especially on the low-level inverse problems for multidimensional images.
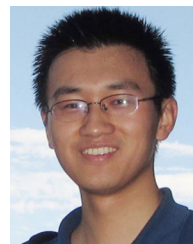
Yi Chang (Member, IEEE) received the B.S. degree in automation from the University of Electronic Science and Technology of China, Chengdu, China, in 2011, the M.S. degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology, Wuhan, China, in 2014, and the Ph.D. degree in pattern recognition and intelligence system with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, China, in 2019.

From 2014 to 2015, he was a Research Assistant with the Peking University, Beijing, China. He was a Research Intern with the machine learning group, Tencent Youtu Lab, Shenzhen, China. Currently, he works as a Postdoc with AI center, PengCheng Laboratory, Shenzhen, China. His research interests include image enhancement/restoration, and multispectral image processing.