

# Multispectral Image Noise Removal With Adaptive Loss and Multiple Image Priors Model

Yang Chen , Wenfei Cao , Meiqiao Bi, Jing Yao , *Member, IEEE*, and Danfeng Hong , *Senior Member, IEEE*

**Abstract**—Multispectral image (MSI) denoising is a crucial preprocessing step for various subsequent image processing tasks, including classification, recognition, and unmixing. This article proposes a novel image denoising model that integrates both noise modeling and image prior knowledge modeling. Specifically, to account for the complexity and nonuniformity of noise, a non-independent identically distributed mixture of Gaussian model is employed for noise modeling, and a weighted loss function is obtained. The weights used in the loss function are adaptively learned from noisy MSI and employed to adjust the denoising strength of each pixel. In additionally, the model leverages the prior knowledge of the image by utilizing a nonlocal low-rank matrix model that captures the spatial–spectral correlation and nonlocal spatial similarity priors of the image. Moreover, our model adopts the weighted spatial–spectral TV model to encode the local smoothness prior of the image. Both prior models are translated into regularization terms in the denoising model. The efficacy of the proposed method is demonstrated through both simulated and real image experiments.

**Index Terms**—Adaptive loss function, multispectral image (MSI) denoising, MSI priors, total variation model.

## I. INTRODUCTION

A MULTISPECTRAL image (MSI) is a 3-D image consisting of a set of 2-D images, each of which is the imaging result of a certain spectral band. MSIs are used extensively in mineral exploration, pharmaceutical counterfeiting, and food safety due to their ability to provide richer information than traditional 2-D images [1], [2], [3]. However, noise contamination is inevitable in MSIs due to sensor sensitivity, calibration errors, and physical mechanisms. The noise distribution is complex and the noise intensity varies across spectral bands [4], [5], [6], [7], leading to significant degradation in image quality and negative impacts on subsequent processing, such as classification, unmixing, and target detection. As a result, MSI noise reduction is a critical and challenging task within the MSI processing field.

In recent decades, there has been a significant advancement in MSI denoising methods. These methods can be broadly

classified into two categories: 1) deep learning-based methods and 2) traditional machine learning-based methods. The deep learning-based methods rely heavily on data and require the design of complex network structures to extract deep prior knowledge from images. For instance, Chang et al. [8] first introduced the deep convolutional neural network (CNN) for MSI denoising. Later, Yuan et al. [9] proposed a deep residual CNN with multiscale and multilevel feature representation for bandwise denoising. To encode the spatial–spectral correlation of image, Zhang et al. [10] presented a deep CNN by incorporating the spatial–spectral gradient information; Shi et al. [11] proposed a 3-D attention network with two separate branches. Pan et al. [12] and Wei et al. [13] introduced a quasi-recurrent network to capture the correlation of spatial features among the spectral domain. Dong et al. [14] presented a modified 3-D U-net architecture, and Cao et al. [15] proposed global reason network with three well-designed modules. More recently, some transformer-based approaches have been applied to HSI and have performed well [16], [17]. Although they provide outstanding denoising results, they lack theoretical support and often do not generalize well to new datasets.

In contrast, the traditional machine learning-based methods do not rely on training data and usually have good theoretical support, and demonstrate better generalization. This type of denoising models typically consists of two primary components, i.e., the loss function term and the regularization term. The loss function term measures the deviation between the ground truth image and the observed image. Most denoising models use the  $l_2$  norm as the loss function, which is simple and convex, leading to fast and efficient algorithms for obtaining the global optimal solution. However, the  $l_2$  norm loss function assumes that the noise obeys an independent identically (i.i.d.) Gaussian distribution, which deviates from the true MSI noise distribution, resulting in the lack of robustness in removing mixed noise. To address this issue, many denoising models have proposed a Gaussian noise plus sparse noise assumption, which treats the non-Gaussian noise as sparse noise and embeds it as a parameter to be learned into the model [18], [19], [20]. This improves the robustness of the algorithm, and thus, have been widely used in MSI denoising task. However, this noise model roughly treats all non-Gaussian noise as sparse noise, which still deviates from the real mixed noise distribution. To alleviate this issue, the mixture of Gaussian (MoG) distribution noise model was proposed [21], [22], which can theoretically approximate any distribution as long as there are enough Gaussian components. The MoG model derives the weighted  $l_2$  norm

Manuscript received 20 April 2023; revised 25 May 2023; accepted 8 June 2023. Date of publication 16 June 2023; date of current version 28 June 2023. This work was supported by the National Natural Science Foundation of China under Grant 12101384, Grant 12271327, Grant 62201553, and Grant 42271350. (Corresponding author: Jing Yao.)

Yang Chen and Wenfei Cao are with the School of Mathematics and Statistics, Shaanxi Normal University, Xi'an 710119, China (e-mail: yangchen9103@163.com; caowenf2015@gmail.com).

Meiqiao Bi, Jing Yao, and Danfeng Hong are with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: meiqiaobi2022@163.com; yaojing@aircas.ac.cn; hongdf@aircas.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3286974

loss function based on the maximum likelihood principle. However, the approximation ability of MoG model is often limited by the small number of Gaussian components in practice. To solve this problem, Cao et al. [23], [24] proposed the mixture of exponential power (MoEP) distribution model, which results in a weighted  $l_p$  norm loss function. In addition, Chen et al. [25] proposed a non-i.i.d. mixture of Gaussian (NMoG) noise model considering the non-i.i.d. statistical characteristics of MSI noise and came up with the weighted  $l_2$  norm function. Further, Barron [26] proposed a more general noise model that can obtain a general loss function by incorporating a set of robust loss functions.

Another critical component of denoising model is the regularization term, which is modeled based on the image prior knowledge. To date, the most significant prior knowledge that have been demonstrated include spatial and spectral correlation, nonlocal spatial similarity, and local smoothness. Extensive research has been dedicated to exploring more accurate and efficient methods for modeling these prior knowledge components. For spatial and spectral correlation prior, the low-rank matrix decomposition model [18], [27], [28], [29] and low-rank tensor decomposition model [30], [31], [32], [33], [34] have been proposed. To incorporate nonlocal spatial similarity prior, it is common practice to partition the image into small blocks, which are then grouped into subimage groups based on their similarity. The model is then applied to each subimage group separately, such as BM4D [35] applies 4-D filtering on each subimage group, Chang et al. [36] arranged each group into a matrix and then applied a low-rank matrix model. In addition, Xue et al. [32] and Zhang et al. [37] reorganized each subimage group into a 3-D tensor and 4-D tensor, respectively, and applied a low-rank tensor model on it. For local smoothness prior, a commonly used approach is to apply 2DTV band by band and then sum them together [38]. However, this method does not account for the smoothness of the image along the spectral dimension. So, the 3DTV model [39] was transferred to MSI and named spatial–spectral total variation (SSTV) [7]. In order to address the large difference in pixel scale and noise intensity across spectral bands in MSI, an adaptive SSATV model was developed by introducing weights into the SSTV model to adjust the denoising strength of each pixel [40], [41]. In addition, to protect image boundary information, Chen et al. [42] presented an adaptive SSTV model. Furthermore, Peng et al. [43], [44] confirmed that the gradient domain of an image is also low-rank, which resulting in the development of enhance TV and CTV models. To achieve improved denoising performance, the denoising model often incorporates multiple prior knowledge models. These may include a combination of spatial and spectral correlation priors with local smoothness priors [45], [46], [47], or a combination of spectral correlation priors with nonlocal similarity priors [48]. Moreover, some models integrate all the three priors completely or partially [7], [49], [50], [51], [52].

The denoising methods mentioned above, which incorporate multiple prior knowledge models, have demonstrated exceptional performance. However, to improve model efficiency, these methods often use relatively simple loss functions, such as  $l_2$  norm. Unfortunately, noise in MSI is often complex and

noni.i.d., which limits the performance of these methods since they assume independent and identically distributed statistical characteristics. In this article, we propose a novel MSI denoising method that combines noise modeling methodology with prior modeling methodology to address this issue.

We begin by assuming a non-i.i.d. noise structure for MSI and use the NMoG to model it. Based on this noise model, we derive an adaptive weighted  $l_2$  norm loss function, the weight reflects the noise intensity of each pixel, and the pixel with high noise intensity has small weight to reduce the adverse impact of strong noise on the model results. On the contrary, the pixel with low noise intensity has large weight to protect the information in the original image from being distorted. The weight information can be efficiently and adaptively learned from noisy images using the variational expectation maximization (VEM) algorithm. Next, we build the regularization term by completely integrating the previously mentioned three image priors. Specifically, we use the nonlocal low-rank matrix model to capture the nonlocal similarity and the spatial–spectral correlation prior. In addition, we adopted an edge preserving total variation model to encode the nonlocal smoothness prior of HSI. Finally, we develop an effective ADMM algorithm to solve the denoising model. We validate the effectiveness of our proposed method on both synthetic and real HSI datasets, and our results indicate that our approach achieves competitive or superior performance compared to other state-of-the-art methods.

## II. NOTATIONS AND PRELIMINARIES

A tensor is a multidimensional data represented by decorated letters, such as  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ , and its element is denoted by  $\mathcal{X}_{i_1, i_2, \dots, i_N}$ . In addition, the matrices, vectors, and scalars are represented by nonbold upper case letters  $X$ , bold lower case letters  $\mathbf{x}$ , and nonbold letters  $x$ , respectively.

The MSI data cube can be treated as a 3-D tensor  $\mathcal{X} \in \mathbb{R}^{M \times N \times S}$  with two spatial modes and a spectral mode, where  $M$ ,  $N$ , and  $S$  represent the spatial height, spatial width, and spectrum number, respectively. The MSI tensor  $\mathcal{X}$  can be unfolded into a matrix along the spectral mode, denoted as  $X_{(3)}$  with the element  $(X_{(3)})_{i,s}$  corresponding to the element  $\mathcal{X}_{i,j,s}$ , where  $l$  is given by  $(i-1)M + j$ .

The observed MSI is often corrupted by complex types of noise, including Gaussian noise, stripe noise, deadline noise, and others. To simplify the noise model, the noise was assumed to be additive, and therefore, we can express the noise degradation model as follows:

$$\mathcal{Y} = \mathcal{X} + \mathcal{E} \quad (1)$$

where  $\mathcal{Y}$  represents the noisy MSI tensor,  $\mathcal{X}$  is the clean MSI tensor, and  $\mathcal{E}$  is the noise tensor.

The distribution of  $\mathcal{E}$  is complex and unknown. To model this type of distribution, a mixture of Gaussian (MoG) distribution is often used. However, since the noise types and intensity between different bands can vary significantly, a more effective approach is to apply the MoG model to the noise of each band. This results in different model parameters for each band, and the parameters for all MoG models are generated from the same

prior distribution. This approach is referred to as NMoG and can be expressed as follows [25]:

$$p(\mathcal{E}_{ijs}) = \sum_{k=1}^K \alpha_{sk} \mathcal{N}(\mathcal{E}_{ijs} | m_{sk}, \tau_{sk}^{-1}) \quad (2)$$

where  $\alpha_{sk}$  and  $\tau_{sk}$  are the proportion and precision of the  $k$ th Gaussian component for the  $s$ th band, with the constraint that  $\sum_{k=1}^K \alpha_{sk} = 1$ ,  $K$  is the total number of Gaussian components. To capture the common properties of noise across bands, the precision  $\tau_s$  was assumed to sample from the following prior distribution:

$$\begin{aligned} m_{sk}, \tau_{sk} &\sim \mathcal{N}(m_{sk} | a_0, (b_0 \tau_{sk})^{-1}) \text{Gam}(\tau_{sk} | \eta_0, \gamma) \\ \gamma &\sim \text{Gam}(\gamma | \mu_0, \nu_0) \end{aligned} \quad (3)$$

where  $\text{Gam}(\cdot)$  represents the Gamma distribution,  $\eta_0$ ,  $\mu_0$ , and  $\nu_0$  are the hyperparameters.

### III. PROPOSED METHOD

In this section, we propose a new denoising model, which falls under the category of traditional machine learning and has the following expression:

$$\min_{\mathcal{X}} \text{Loss}(\mathcal{Y}, \mathcal{X}) + \lambda \text{R}(\mathcal{X}) \quad (4)$$

where  $\text{Loss}(\mathcal{Y}, \mathcal{X})$  presents the loss function which measures the deviation between observation MSI and the ground truth. And  $\text{R}(\mathcal{X})$  is the regularization term which encodes the prior information of MSI. We will introduce the proposed method in detail from two aspects: 1) loss function and 2) regularization term.

#### A. Adaptive Loss Function

To construct a more appropriate loss function, it is crucial to have a comprehensive understanding of the noise and then reasonably model the distribution characteristics of the noise, which in turn will help to measure the deviation between observed data and ground truth. Moreover, since the noise in each data point can vary, the loss function must be adaptive to cater to the specific noise characteristics. To meet these requirements, we have selected the NMoG noise model and used the VEM algorithm to estimate the model parameters. This has enabled us to derive an explicit loss function that is suitable for our purposes. We have simplified the NMoG model slightly by assuming that the noise is zero-mean. To solve this model, we have introduced a hidden variable  $\mathcal{Z}$  in the noise model. The NMoG model can be expressed as follows:

$$\begin{aligned} p(\mathcal{Y}_{ijs}) &= \prod_{k=1}^K \mathcal{N}(\mathcal{Y}_{ijs} | \mathcal{X}_{ijs}, \tau_{sk}^{-1})^{z_{ijsk}} \\ p(\tau_{sk}) &= \text{Gam}(\tau_{sk} | \eta_0, \gamma_{sk}) \\ p(z_{ijs}) &= \text{Mul}(z_{ijs} | \alpha_s) \\ p(d_{sk}) &= \text{Gam}(d_{sk} | \mu_0, \nu_0) \\ p(\alpha_s) &= \text{Dir}(\alpha_s | \alpha_0) \end{aligned} \quad (5)$$

where  $\text{Mul}(\cdot)$  and  $\text{Dir}(\cdot)$  represent the multinomial distribution and the Dirichlet distribution, respectively.

The maximization of marginal likelihood  $p(\mathcal{Y} | \mathcal{X})$  can be transferred to maximize the variational lower bound (ELBO) as follows [53]:

$$\text{ELBO} = \int q^*(\Theta | \mathcal{X}^{\text{old}}) \ln p(\mathcal{Y}, \Theta | \mathcal{X}) d\Theta \quad (6)$$

where  $\Theta = (\mathcal{Z}, \alpha, \tau, \Gamma)$ ,  $\mathcal{Z} = \{z_{ijs}\}$ ,  $\alpha = \{\alpha_{sk}\}$ ,  $\tau = \{\tau_{sk}\}$ , and  $\gamma = \{\gamma_{sk}\}$ . The distribution  $q^*(\Theta | \mathcal{X}^{\text{old}})$  is the estimated posterior distribution.

*V Step:* We employ the variational inference algorithm to solve the approximate posterior distribution of the parameters involved in (6). The estimated posterior distribution is assumed to have the following decomposition form:

$$q(\mathcal{Z}, \alpha, \tau, \Gamma) = \prod_{ijs} q(\mathcal{Z}_{ijs}) \prod_b q(\alpha_s) \prod_{sk} q(\tau_{sk}) q(\Gamma_{sk}).$$

The posterior distribution of parameter  $\tau_s$  can be updated using the following equation:

$$q^*(\tau_s) = \prod_k \text{Gam}(\tau_{sk} | \eta_{sk}, \gamma_{sk}) \quad (7)$$

where the parameters involved in the estimated posterior is obtained as follows:

$$\eta_{sk} = \eta_0 + \frac{1}{2} \sum_{ij} \langle \mathcal{Z}_{ijsk} \rangle \quad (8)$$

$$\gamma_{sk} = \langle \gamma_{sk} \rangle + \frac{1}{2} \left\{ \sum_{ij} \mathcal{E}_{ijs}^2 \langle \mathcal{Z}_{ijsk} \rangle \right\} \quad (9)$$

where  $\langle \cdot \rangle$  represents the expectation operator.

The posterior distribution of latent variable  $\mathcal{Z}$  is

$$q^*(\mathcal{Z}_{ijs}) = \prod_k \varrho_{ijsk}^{z_{ijsk}} \quad (10)$$

where the closed-form solution of distribution parameter  $\varrho_{ijsk}$  is as follows:

$$\begin{aligned} \varrho_{ijsk} &= \rho_{ijsk} / \sum_k \rho_{ijsk} \\ \ln \rho_{ijsk} &= \langle \ln \alpha_{sk} \rangle - \ln \sqrt{2\alpha} + \langle \ln \tau_{sk} \rangle / 2 \\ &\quad - \mathcal{E}_{ijs}^2 \langle \tau_{sk} \rangle / 2. \end{aligned} \quad (11)$$

The posterior distribution of the mixing proportion  $\alpha_s$  is as follows:

$$q^*(\alpha_s) = \prod_k \alpha_{sk}^{\alpha_{sk} - 1} \quad (12)$$

where  $\alpha_{sk} = \alpha_0 + \sum_{ij} \langle \mathcal{Z}_{ijsk} \rangle$ .

The posterior distribution of hyperparameter  $\gamma$  is

$$q^*(\gamma_{sk}) = \text{Gam}(\gamma_{sk} | \mu, \nu) \quad (13)$$

where  $\mu = \mu_0 + c_0 K S$  and  $\nu = \nu_0 + \sum_{s,k} \langle \tau_{sk} \rangle$ .

Next, we present the necessary expectations required for the above update equations, based on the current posterior distributions. The details are provided as follows:

$$\begin{aligned} \langle z_{ijsk} \rangle &= \varrho_{ijsk} \\ \langle d_{sk} \rangle &= \frac{\mu}{\nu} \end{aligned}$$

$$\begin{aligned} \langle \ln \alpha_{sk} \rangle &= \psi(\alpha_{sk}) - \psi(\hat{\alpha}) \\ \langle \tau_{sk} \rangle &= \frac{c_{sk}}{d_{sk}} \end{aligned} \quad (14)$$

where  $\psi(\cdot)$  is the digamma function.

*E Step:* We calculate the evidence lower bound (ELBO) of the logarithm of likelihood  $p(\mathcal{Y}|\mathcal{X})$  as (6). Note that we only focus on the components related to  $\mathcal{X}$  and others were treated as constants. Therefore, the ELBO can be denoted as a function of  $\mathcal{X}$  with respect to  $\mathcal{X}^{\text{old}}$ . We denote the function as  $Q(\mathcal{X}, \mathcal{X}^{\text{old}})$

$$\begin{aligned} &Q(\mathcal{X}, \mathcal{X}^{\text{old}}) \\ &= \int q^*(\Theta|\mathcal{X}^{\text{old}}) \ln p(\mathcal{Y}, \Theta|\mathcal{X}) d\Theta \\ &= - \sum_{i,j,s,k} \frac{1}{2} \langle \mathcal{Z}_{ijsk} \rangle \langle \tau_{sk} \rangle (\mathcal{Y}_{ijs} - \mathcal{X}_{ijs})^2 + \text{const.} \end{aligned} \quad (15)$$

*M Step:* We maximum the ELBO function  $Q(\mathcal{X}, \mathcal{X}^{\text{old}})$  with respect to  $\mathcal{X}$ , the loss function of this optimization problem can be reformulated as follows:

$$\begin{aligned} \text{Loss}(\mathcal{Y}, \mathcal{X}) &= \|\mathcal{W} \odot (\mathcal{Y} - \mathcal{X})\|_F^2 \\ \mathcal{W}_{ijs} &= \sqrt{\sum_{k=1}^K \frac{1}{2} \langle \mathcal{Z}_{ijsk} \rangle \langle \tau_{sk} \rangle} \end{aligned} \quad (16)$$

where  $\odot$  represents elementwise product.

The loss function obtained here takes the form of a weighted l2-norm. The weight is calculated adaptively based on the estimated noise from the denoising process. It can be seen that the high-intensity noise receives small weight, reducing the negative impact of strong noise on model results. Conversely, the low-intensity noise receives large weight to preserve the structure information in the original image.

### B. Multiple Image Priors Model-Based Regularization

As mentioned previously, the nonlocal spatial similarity, the spatial and spectral correlation, and the local smoothness are the most significant priors knowledge for MSI restoration. The regularization term in our denoising model will completely encode all these priors with the following form:

$$R(\mathcal{X}) = \|\mathcal{X}\|_{\text{NSC}} + \|\mathcal{X}\|_{\text{LS}} \quad (17)$$

where  $\|\mathcal{X}\|_{\text{NSC}}$  represents the regularization related to nonlocal spatial similarity and the spatial-spectral correlation priors. And  $\|\mathcal{X}\|_{\text{LS}}$  represents the regularization related to local smoothness prior.

To model the nonlocal spatial similarity prior, it is necessary to divide the images into overlapping patches for matching and grouping. In order to simultaneously encode the spatial-spectral correlation prior, the MSI is divided into overlapping full-band patches to ensure that the correlation exists in each patch. This allows the low-rank model to be applied to each group of full-band patches, enabling both nonlocal spatial similarity prior and the spatial-spectral correlation prior modeling. Note that Chang et al. [36] demonstrated that, for MSI, the low-rank property between patches is much more significant than between

bands. To simplify the model, each patch in the group is vectorized to form a matrix, which is then subjected to low-rank matrix decomposition. Therefore, the regularization related to nonlocal spatial similarity and spatial-spectral correlation priors can be expressed as follows:

$$\|\mathcal{X}\|_{\text{NSC}} = \sum_{i=1}^P \|\mathbf{P}_i \mathcal{X}\|_{w,*} \quad (18)$$

where  $\mathbf{P}_i$  represents the operator to group the similar patches for the  $i$ th patch and rearrange it into matrix,  $P$  denotes the number of patches, and  $\|\cdot\|_{w,*}$  represents the weighted nuclear norm.

The local smoothness prior is not only present in two spatial modes but also extends to the spectral mode. To capture such prior, the spatial-spectral total variation (SSTV) model is commonly used. Due to significant variations in noise intensity and pixel values, it is necessary to impose TV regularization constraints with different strengths on each pixel. To address this, we introduce a weighted SSTV model that allows for the application of varying regularization strengths on each pixel, which defined as follows:

$$\begin{aligned} \|\mathcal{X}\|_{\text{LS}} &= \sum_{i=1}^3 \|\mathcal{S}_i \odot (\mathbf{D}_i \mathcal{X})\|_1 \\ &= \|\mathcal{S} \odot (\mathbf{D} \mathcal{X})\|_1 \end{aligned} \quad (19)$$

where the operator  $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3]$  represents the 3-D first-order forward finite-difference operator. In addition, the operators  $\mathbf{D}_1$ ,  $\mathbf{D}_2$ , and  $\mathbf{D}_3$  correspond to the first-order finite-difference operator along the spatial horizontal, spatial vertical, and spectral mode, respectively. The parameter  $\mathcal{S} = [\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3]$  is the weight tensor, which is used to preserve image texture while enforcing sparsity constraint on the pixels. We can estimate the weight tensor from gradient maps of image. Although a large amount of noise has been removed from the denoised image  $\hat{\mathcal{X}}$ , the SNR is not enough to extract the texture information of the image. To alleviate the situation, we carry out low-rank approximation with rank setting as 1 to further denoise to get  $\mathcal{X}_0$  with higher SNR, so the texture information of the image can be effectively obtained from the gradient maps of  $\mathcal{X}_0$ . Therefore, the weight tensor is obtained from restored MSI during the denoising process as follows [42]:

$$\begin{aligned} \mathcal{X}_0 &= \text{LR}_1(\hat{\mathcal{X}}) \\ \mathcal{G}_i &= \frac{|\mathbf{D}_i \mathcal{X}_0|}{\max(|\mathbf{D}_i \mathcal{X}_0|)} \\ \mathcal{S}_i &= \frac{1}{\delta + \mathcal{G}_i}, \quad i = 1, 2, 3 \end{aligned} \quad (20)$$

where  $\text{LR}_1$  is the low-rank approximation operator using a low-rank matrix decomposition model with a fixed rank of 1.  $\hat{\mathcal{X}}$  represents the restored MSI data during denoising processing, and the threshold  $\delta$  is utilized to prevent undesirably large values in the weight tensor. Typically, it is set to the first quartile of  $\mathcal{G}_i$ .

### C. Denoising Algorithm

Based on the adaptive loss function and regularization term introduced above, the denoising model can be represented as follows:

$$\begin{aligned} \mathcal{X} = \arg \min_{\mathcal{X}} & \|\mathcal{W} \odot (\mathcal{Y} - \mathcal{X})\|_F^2 + \lambda_1 \|\mathcal{S} \odot (\mathbf{D}\mathcal{X})\|_1 \\ & + \lambda_2 \sum_{i=1}^P \|\mathbf{P}_i \mathcal{X}\|_{w,*} \end{aligned} \quad (21)$$

where  $\lambda_1$  and  $\lambda_2$  is the tradeoff parameters.

In order to solve this model, the auxiliary variables are introduced, and thus, (21) can be rewritten into the following optimization problem:

$$\begin{aligned} \min_{\mathcal{X}, \mathcal{B}, \mathbf{C}_i} & \|\mathcal{W} \odot (\mathcal{Y} - \mathcal{X})\|_F^2 + \lambda_1 \|\mathcal{S} \odot \mathcal{B}\|_1 + \lambda_2 \sum_{i=1}^P \|\mathbf{C}_i\|_{w,*} \\ \text{s.t.} \quad & \mathcal{A} = \mathcal{X} \\ & \mathcal{B} = \mathbf{D}\mathcal{A} \\ & \mathbf{C}_i = \mathbf{P}_i \mathcal{X}, \quad i = 1, \dots, P. \end{aligned} \quad (22)$$

The ADMM methodology can be used to minimize the optimization problem mentioned above by transforming it into an augmented Lagrangian function as follows:

$$\begin{aligned} L(\mathcal{X}, \mathcal{A}, \mathcal{B}, \{\mathbf{C}_i\}_{i=1}^P) &= \|\mathcal{W} \odot (\mathcal{Y} - \mathcal{X})\|_F^2 + \lambda_1 \|\mathcal{S} \odot \mathcal{B}\|_1 \\ &+ \lambda_2 \sum_{i=1}^P \|\mathbf{C}_i\|_{w,*} + \frac{\mu_1}{2} \|\mathcal{A} + \frac{\mathcal{L}_1}{\mu_1} - \mathcal{X}\|_F^2 \\ &+ \frac{\mu_2}{2} \|\mathbf{D}\mathcal{A} + \frac{\mathcal{L}_2}{\mu_2} - \mathcal{B}\|_F^2 + \frac{\mu_3}{2} \sum_{i=1}^P \|\mathbf{P}_i \mathcal{X} + \frac{\mathbf{L}_3}{\mu_3} - \mathbf{C}_i\|_F^2 \end{aligned} \quad (23)$$

where  $\mu_1$ ,  $\mu_2$ , and  $\mu_3$  are penalty parameters, and  $\mathcal{L}_1$ ,  $\mathcal{L}_2$ , and  $\mathbf{L}_3$  are the lagrange multipliers.

The ADMM framework provides an alternative approach to optimize the augmented Lagrangian function by fixing all variables except one and optimizing it.

*Update A:* The suboptimal problem for optimizing the variable  $\mathcal{A}$  is derived from (23) by removing the terms not related to  $\mathcal{A}$

$$\min_{\mathcal{A}} \|\mathcal{X} - \frac{1}{\mu_1} \mathcal{L}_1 - \mathcal{A}\|_F^2 + \frac{\mu_2}{\mu_1} \|\mathcal{B} - \frac{1}{\mu_2} \mathcal{L}_2 - \mathbf{D}\mathcal{A}\|_F^2. \quad (24)$$

This problem can be treated as solving the linear system

$$\left( \mathbf{I} + \frac{\mu_2}{\mu_1} \mathbf{D}^* \mathbf{D} \right) \mathcal{A} = \mathcal{X} - \frac{1}{\mu_1} \mathcal{L}_1 + \frac{\mu_2}{\mu_1} \mathbf{D}^* \left( \mathcal{B} - \frac{1}{\mu_2} \mathcal{L}_2 \right) \quad (25)$$

where  $\mathbf{D}^*$  represents the adjoint operator of  $\mathbf{D}$ . The fast Fourier transform is adopted to efficiently solve this problem, and the closed-form solution to  $\mathcal{A}$  can be obtained by the following

equations:

$$\begin{cases} \mathcal{K} = \mathcal{X} - \frac{1}{\mu_1} \mathcal{L}_1 + \frac{\mu_2}{\mu_1} \mathbf{D}^* \left( \mathcal{B} - \frac{1}{\mu_2} \mathcal{L}_2 \right) \\ \mathcal{J} = |\text{fftn}(\mathbf{D}_h)|^2 + |\text{fftn}(\mathbf{D}_v)|^2 + |\text{fftn}(\mathbf{D}_s)|^2 \\ \mathcal{A}^{t+1} = \text{ifftn} \left( \frac{\text{fftn}(\mathcal{K})}{\mathbf{1} + \frac{\mu_2}{\mu_1} \mathcal{J}} \right) \end{cases} \quad (26)$$

where  $|\cdot|^2$  represents the elementwise square, the operators  $\text{fftn}(\cdot)$  and  $\text{ifftn}(\cdot)$  are the fast 3-D Fourier transform and its inverse transform, respectively, and  $\mathbf{1}$  is the tensor with all elements 1.

*Update B:* The suboptimal problem with respect to  $\mathcal{B}$  is as follows:

$$\min_{\mathcal{B}} \frac{1}{2} \|\mathbf{D}\mathcal{A} + \frac{1}{\mu_2} \mathcal{L}_2 - \mathcal{B}\|_F^2 + \frac{\lambda_1}{\mu_2} \|\mathcal{S} \odot \mathcal{B}\|_1. \quad (27)$$

This suboptimal problem can be solved by using the known *soft-thresholding* operator [54] as follows:

$$\mathbf{B}_{ijb}^{(t+1)} = \mathbf{R}_{\frac{\lambda_1 s_{ijb}}{\mu_2}} \left( \mathbf{D}\mathcal{A} + \frac{1}{\mu_2} \mathcal{L}_2 \right) \quad (28)$$

where  $\mathbf{R}_\lambda(\cdot)$  represents the *soft-thresholding* operator with parameter  $\lambda$ .

*Update C<sub>i</sub>:* The suboptimal problem concerning  $\mathbf{C}_i$  is as follows:

$$\min_{\mathbf{C}_i} \frac{1}{2} \|\mathbf{P}_i \mathcal{X} + \frac{1}{\mu_3} \mathbf{L}_3 - \mathbf{C}_i\|_F^2 + \frac{\lambda_2}{\mu_3} \|\mathbf{C}_i\|_{w,*}. \quad (29)$$

To solve this subproblem, we utilize the off-the-shelf algorithm WNNM [55]. The optimal solution of  $\mathbf{C}_i$  is deduced as the following equations:

$$\begin{aligned} (\mathbf{U}, \Sigma, \mathbf{V}) &= \text{svd} \left( \mathbf{P}_i \mathcal{X} + \frac{1}{\mu_3} \mathbf{L}_3 \right) \\ \mathbf{C}_i^{(t+1)} &= \mathbf{U} \hat{\Sigma} \mathbf{V}^T \end{aligned} \quad (30)$$

where the diagonal elements of  $\Sigma$  correspond to the singular values of  $\mathbf{P}_i \mathcal{X} + \frac{1}{\mu_3} \mathbf{L}_3$  arranged in decreasing order. In addition, the  $\hat{\Sigma}$  is the diagonal matrix, and the diagonal elements are calculated as follows:

$$\hat{\sigma}_i = \begin{cases} 0, & \text{if } \alpha_2 < 0 \\ \frac{\alpha_1 + \sqrt{\alpha_2}}{2}, & \text{if } \alpha_2 \geq 0 \end{cases} \quad (31)$$

where the parameters  $\alpha_1$  and  $\alpha_2$  are expressed as follows:

$$\begin{aligned} \alpha_1 &= \sigma_i - \epsilon \\ \alpha_2 &= (\sigma_i + \epsilon)^2 - 4\sqrt{2}. \end{aligned} \quad (32)$$

*Update X:* The subproblem with respect to  $\mathcal{X}$  can be written as

$$\begin{aligned} \min_{\mathcal{X}} & \|\mathcal{W} \odot (\mathcal{Y} - \mathcal{X})\|_F^2 + \frac{\mu_1}{2} \|\mathcal{A} + \frac{\mathcal{L}_1}{\mu_1} - \mathcal{X}\|_F^2 \\ & + \frac{\mu_3}{2} \|\mathbf{P}^{-1} \mathbf{C} - \frac{\mathbf{L}_3}{\mu_3} - \mathcal{X}\|_F^2 \end{aligned} \quad (33)$$

where  $\mathbf{C} = \{\mathbf{C}_i\}_{i=1}^P$ , and the operator  $\mathbf{P}^{-1}$  represents the inverse operator of  $\mathbf{P}$ . Its function is to arrange each patch back to its original position in the MSI while also averaging out the

**Algorithm 1:** Denoising Algorithm.

**Input:** The observed MSI tensor  $\mathcal{Y} \in \mathbb{R}^{M \times N \times S}$ , model parameters  $K, \lambda_1, \lambda_2$ , and the maximum number of iterations  $T$ .

**Output:** The denoised MSI  $\mathcal{X}$ .

**Initialization:** The hyperparameters  $(c_0, d_0, \mu_0, \nu_0)$  in noise prior. The restored MSI  $\mathcal{X}$  by PCA-based denoising method.  $\mathcal{A} = \mathcal{X}, \mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$  are tensors or matrices with all elements 0.  $t = 1$ .

- 1: **while**  $t \leq T$  **do**
- 2:   Update loss function by (7)–(16).
- 3:   Update  $\mathcal{A}, \mathcal{B}, \{C_i\}_{i=1}^P, \mathcal{X}$  by (25)–(34).
- 4:   Update  $\mathcal{S}$  by (20).
- 5:   Update lagrange multipliers  $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$  by (35).
- 6:    $t \leftarrow t + 1$ .
- 7: **end while**

overlapping portions. The solution of  $\mathcal{X}$  has the close-form as follows:

$$\mathcal{X}^{(t+1)} = \frac{2\mathcal{W}^2 \odot \mathcal{Y} + \mu_1(\mathcal{A} - \frac{1}{\mu_1}\mathcal{L}_1) + \mu_3(\mathbf{P}^{-1}\mathcal{C} - \frac{1}{\mu_3}\mathcal{L}_3)}{2\mathcal{W}^2 + \mu_1 + \mu_3}. \quad (34)$$

The multipliers are updated as follows:

$$\begin{cases} \mathcal{L}_1^{(t+1)} = \mathcal{L}_1^{(t)} + \mu_1(\mathcal{A} - \mathcal{X}) \\ \mathcal{L}_2^{(t+1)} = \mathcal{L}_2^{(t)} + \mu_2(\mathbf{D}\mathcal{A} - \mathcal{B}) \\ \mathcal{L}_3^{(t+1)} = \mathcal{L}_3^{(t)} + \frac{\mu_3}{P} \sum_{i=1}^P (\mathbf{P}_i \mathcal{X} - C_i). \end{cases} \quad (35)$$

After updating the denoised data, we proceed to update the weights involved in the loss function using (16). More information on this step can be found in Section III-A. The denoising algorithm is summarized in Algorithm 1.

#### IV. EXPERIMENTS

To demonstrate the effectiveness of our proposed denoising method, we compared our method with several state-of-the-art denoising methods on both simulated and real MSI data, including BM4D [35], TDL [30], LRTV [38], NMoG [25], LRTDTV [47], LLRT [36], and RCTV [56]. Specifically, BM4D is a classical method that utilizes block-matching and 4-D filtering for denoising. TDL utilizes the  $l_2$ -norm loss function and spatial–spectral correlation prior model. LRTV is based on the Gaussian plus sparse noise model, combined with the low-rank matrix decomposition model. NMoG is a popular method that uses the NMoG noise model and low-rank matrix decomposition model. LRTDTV combines the Gaussian noise plus sparse noise model with the low-rank tensor decomposition prior model and SSTV regularization. LLRT uses the  $l_2$ -norm loss function and combines the nonlocal similarity, spectral correlation, and local smoothness priors model. RCTV uses the Gaussian plus sparse noise model, as well as the local smoothness and spectral correlation priors model. Overall, the comparison experiment involved a diverse range of denoising methods, allowing us to accurately evaluate the performance of our proposed method against the

current state of the arts. All experiments were implemented in Matlab R2021a on a PC with 3.4 GHz CPU and 32 GB RAM.

##### A. Simulation Experiments

This experiment aims to evaluate the performance of our proposed denoising method quantitatively. We utilized two MSI datasets in the simulation experiment: the Balloons data from the CAVE dataset<sup>1</sup> [57] with size of  $512 \times 512 \times 31$ , and the Washington DC MALL dataset<sup>2</sup> with size of  $1208 \times 307 \times 191$ . We resized the Balloons data to  $200 \times 200 \times 31$  and cropped the main part of the Washington DC MALL dataset to obtain a size of  $200 \times 200 \times 191$ . These datasets were used as clean MSI data since they do not contain significant visual noise and their gray values were normalized to  $[0,1]$ .

To simulate the real noise cases, we added four kinds of noise to the clean MSI data, including the following.

- 1) *I.i.d. Gaussian Noise:* The i.i.d. Gaussian noise  $N(0, \sigma^2)$  with the variance  $\sigma = 0.1$  is added to the clean MSI data.
- 2) *Non-i.i.d. Gaussian Noise:* The i.i.d. Gaussian noise  $N(0, \sigma^2)$  is added to each band of clean MSI data independently with different variance, making the SNR value of each band in the range of  $[5, 15]$ dB on Balloons dataset and  $[2, 12]$ dB on Washington DC Mall dataset, respectively.
- 3) *Gaussian + Stripe Noise:* The i.i.d. Gaussian noise is added to each band of clean MSI data with different intensity. Moreover, 40% bands were randomly selected to add stripe noise.
- 4) *Mixture Noise:* The previously introduced non-i.i.d. Gaussian noise and stripe noise were added to clean MSI. In addition, 40% (on Balloons dataset) and 45% (on Washington DC Mall dataset) bands were randomly selected to add the impluse noise with the percentage of impluse is from 50% to 70% (on Balloons dataset) and 90% to 100% (on Washington DC Mall dataset), respectively.

We conducted 20 repetitions of the noise addition and denoising experiments on two datasets for each noise case. In these experiments, the maximum number of iterations of our method is set to 10, and the hyperparameters  $\lambda_1$  and  $\lambda_2$  are adjusted in each noise case. The sensitivity of parameter is analyzed in Section IV-C. Five quantitative measurements were employed to evaluate the denoising performance, namely:

- 1) MPSNR, which is the means peak signal-to-noise ratio (PSNR) across bands;
- 2) MSSIM, which is the mean structural similarity (SSIM) across bands;
- 3) ERGAS, which stands for Erreur Relative Globale Adimensionnelle de Synthèse, and
- 4) SAM, which refers to the spectral angle mapper;
- 5) time, which is the experimental time cost of each method.

The average results of 20 repeated experiments on the Balloons dataset are presented in Table I. The best value of measurements are marked in bold. LLRT method performs remarkably in

<sup>1</sup>[Online]. Available: <https://www.cs.columbia.edu/CAVE/databases/multispectral/>

<sup>2</sup>[Online]. Available: <http://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>

TABLE I  
PERFORMANCE COMPARISON OF ALL COMPETING METHODS ON BALLOONS DATA

	Noisy HSI	BM4D	TDL	LRTV	NMoG	LRTDTV	LLRT	RCTV	Ours
i.i.d. Gaussian Noise									
MPSNR	20.174	39.472	38.045	35.337	30.215	37.612	40.310	38.029	<b>41.524</b>
MSSIM	0.166	0.965	0.953	0.943	0.661	0.952	<b>0.984</b>	0.628	<b>0.984</b>
ERGAS	25.840	2.839	3.310	4.701	8.185	3.511	2.660	3.334	<b>2.222</b>
SAM	0.753	0.122	0.116	0.190	0.359	0.132	<b>0.084</b>	0.123	0.085
Time	-	56	6	21	37	91	679	3	356
Non i.i.d. Gaussian Noise									
MPSNR	23.470	41.144	33.387	38.199	33.049	39.158	41.041	39.303	<b>41.169</b>
MSSIM	0.286	0.972	0.781	0.967	0.794	0.966	<b>0.986</b>	0.697	0.959
ERGAS	19.553	<b>2.433</b>	6.498	3.339	6.012	3.014	2.465	2.952	3.148
SAM	0.654	0.105	0.266	0.117	0.288	0.126	<b>0.089</b>	0.117	0.144
Time	-	59	6	24	49	88	522	2	237
Gaussian + Stripe Noise									
MPSNR	22.831	35.839	28.8505	37.897	32.766	38.358	36.263	38.017	<b>40.004</b>
MSSIM	0.272	0.884	0.590	<b>0.963</b>	0.784	0.955	0.951	0.655	0.960
ERGAS	20.862	5.964	11.339	3.461	6.269	3.326	4.262	3.561	<b>3.223</b>
SAM	0.675	0.224	0.406	0.127	0.296	0.143	0.148	0.159	<b>0.142</b>
Time	-	43	5	21	33	60	437	2	235
Mixture Noise									
MPSNR	21.240	31.272	25.574	36.543	31.861	37.174	32.703	37.057	<b>37.9350</b>
MSSIM	0.236	0.878	0.536	0.959	0.759	0.952	0.898	0.651	<b>0.969</b>
ERGAS	24.296	8.377	15.054	4.076	6.863	3.763	6.392	3.886	<b>3.445</b>
SAM	0.705	0.196	0.417	0.128	0.287	0.132	0.181	0.149	<b>0.114</b>
Time	-	43	4	21	33	62	438	2	236

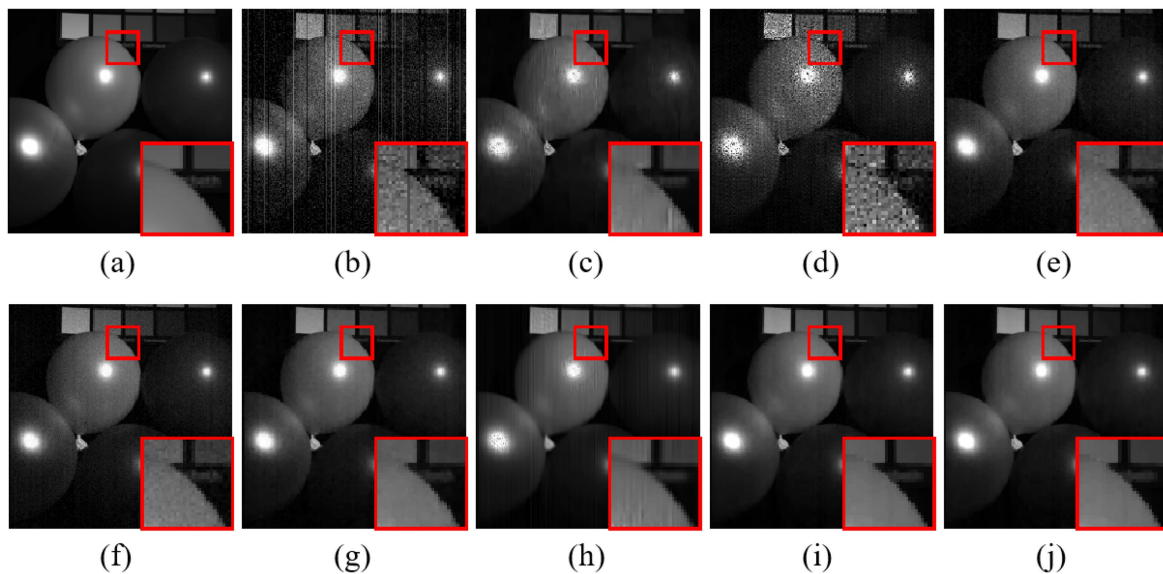


Fig. 1. Image of band 23 in Balloons dataset before and after denoising by different methods. (a) Original data. (b) Noisy data, the restored results by method. (c) BM4D. (d) TDL. (e) LRTV. (f) NMoG. (g) LRTDTV. (h) LLRT. (i) RCTV. (j) Ours.

the i.i.d. Gaussian case since their noise assumption matches the actual noise well. However, as the noise complexity increases, the performance of LLRT tends to decrease. Similarly, BM4D and TDL methods that use the  $l_2$ -norm loss function show a similar trend. On the other hand, methods utilizing the Gaussian plus sparse noise model, such as LRTV, LRTDTV, and RCTV, demonstrate more robust performance. However, their denoising performance is weaker than our method due to the better image priors model adopted in our denoising approach. Although the NMoG method performs stably under all noise conditions, but its denoising measurements are worse than others since it models only the spectral correlation prior, which may not be

the best choice for this dataset with only 31 bands. Overall, our method exhibits strong robustness and remarkable denoising performance compared to other methods.

As a specific example, Fig. 1 provides the visual presentation of the denoising results of one experiment under mixture noise case. To facilitate comparison, we enlarged a common region of each figure, marked by a red box. The figures demonstrate that our method visually achieves best denoising performance compared with other methods.

Table II presents the denoising results on the Washington DC mall dataset. It is worth noting that this dataset has a greater number of spectral bands and exhibits strong spectral

TABLE II  
PERFORMANCE COMPARISON OF ALL COMPETING METHODS ON WASHINGTON DCMALL DATA

	Noisy HSI	BM4D	TDL	LRTV	NMoG	LRTDTV	LLRT	RCTV	Ours
i.i.d. Gaussian Noise									
MPSNR	20.172	32.914	36.380	35.305	36.192	36.249	37.518	35.497	<b>37.854</b>
MSSIM	0.448	0.920	0.968	0.954	0.967	0.962	0.977	0.938	<b>0.979</b>
ERGAS	25.934	5.783	3.933	4.414	4.082	3.996	3.591	4.309	<b>3.343</b>
SAM	0.559	0.134	0.092	0.099	0.107	0.101	0.080	0.087	<b>0.077</b>
Time	-	307	17	154	55	355	2815	10	1554
Non i.i.d. Gaussian Noise									
MPSNR	21.445	31.234	26.623	34.634	<b>37.000</b>	36.068	35.583	35.552	36.755
MSSIM	0.464	0.817	0.670	0.933	0.952	0.946	0.943	0.940	<b>0.953</b>
ERGAS	24.638	9.169	14.635	6.275	5.867	4.491	4.803	4.448	<b>4.186</b>
SAM	0.609	0.284	0.404	0.177	0.205	0.127	0.134	<b>0.103</b>	0.110
Time	-	294	19	143	100	346	2667	8	1593
Gaussian + Stripe Noise									
MPSNR	20.805	30.147	25.602	30.434	<b>36.625</b>	35.619	34.613	35.009	36.347
MSSIM	0.488	0.853	0.691	0.848	<b>0.967</b>	0.958	0.952	0.935	0.966
ERGAS	25.697	9.359	15.938	8.569	5.059	4.541	5.302	4.814	<b>4.260</b>
SAM	0.617	0.276	0.426	0.189	0.164	0.120	0.149	<b>0.104</b>	0.108
Time	-	296	19	142	59	351	2679	9	1610
Mixture Noise									
MPSNR	19.752	28.835	24.318	33.441	<b>34.967</b>	34.572	33.754	33.862	34.291
MSSIM	0.444	0.826	0.653	0.944	<b>0.962</b>	0.949	0.950	0.925	0.954
ERGAS	28.261	10.907	18.118	7.013	5.647	5.810	5.650	5.550	<b>5.439</b>
SAM	0.655	0.289	0.466	0.195	0.165	0.175	0.130	<b>0.128</b>	0.135
Time	-	288	19	145	62	345	2649	9	2126

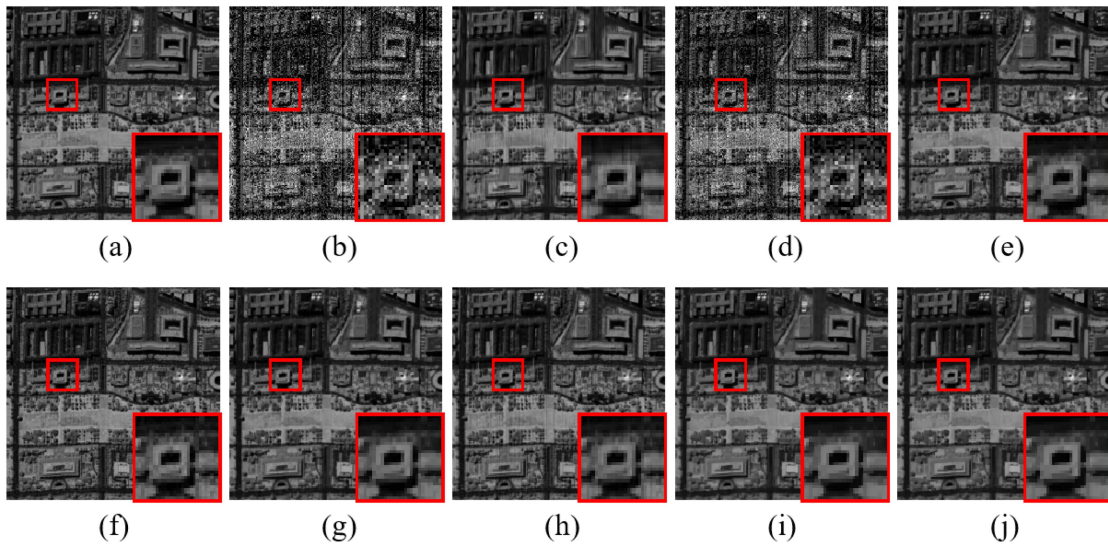


Fig. 2. Image of band 90 in Washington DC mall dataset before and after denoising by different methods. (a) Original data. (b) Noisy data, the restored results by method. (c) BM4D. (d) TDL. (e) LRTV. (f) NMoG. (g) LRTDTV. (h) LLRT. (i) RCTV. (j) Ours.

correlation, making the NMoG method remarkably effective in all noise cases. On the other hand, the i.i.d. Gaussian noise model-based method is unstable, while the Gaussian plus sparse noise model-based method is relatively more stable. Our method achieves the best or second-best performance under all noise cases. The visual comparison results of one realization are illustrated in Fig. 2. From the figure, it is evident that our method outperforms other methods in removing noise and preserving detailed information.

### B. Real Data Experiments

In this section, we present an evaluation of the performance of our method on a real MSI dataset, namely, the AVIRIS Indian

Pines dataset<sup>3</sup> with the size  $145 \times 145 \times 220$ . This dataset is corrupted by various types of noise, including Gaussian noise, stripes, atmosphere absorption, and other unknown noise. In these experiments, the maximum number of iterations of our method is set to 10, and hyperparameters are set as  $\lambda_1 = 500$  and  $\lambda_2 = 5$ .

Figs. 3–5 show the visual presentation of denoising results on bands 103, 149, and 165, respectively. It is evident that the TDL method fails to denoise this dataset effectively, and the BM4D method obviously has residual noise. This is due to the overly simplistic noise assumption and inadequate image prior

<sup>3</sup>[Online]. Available: <http://www.tec.army.mil/hypercube>



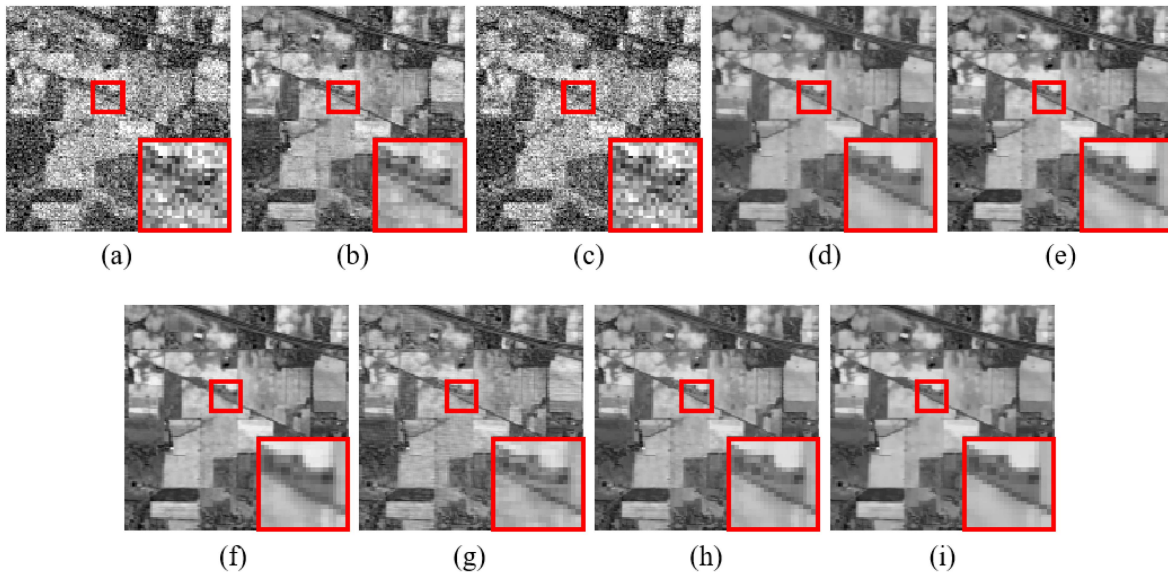


Fig. 3. Image of band 103 in AVIRIS Indian Pines dataset before and after denoising by different methods. (a) Original data, the restored results by method. (b) BM4D. (c) TDL. (d) LRTV. (e) NMoG. (f) LRTDTV. (g) LLRT. (h) RCTV. (i) Ours.

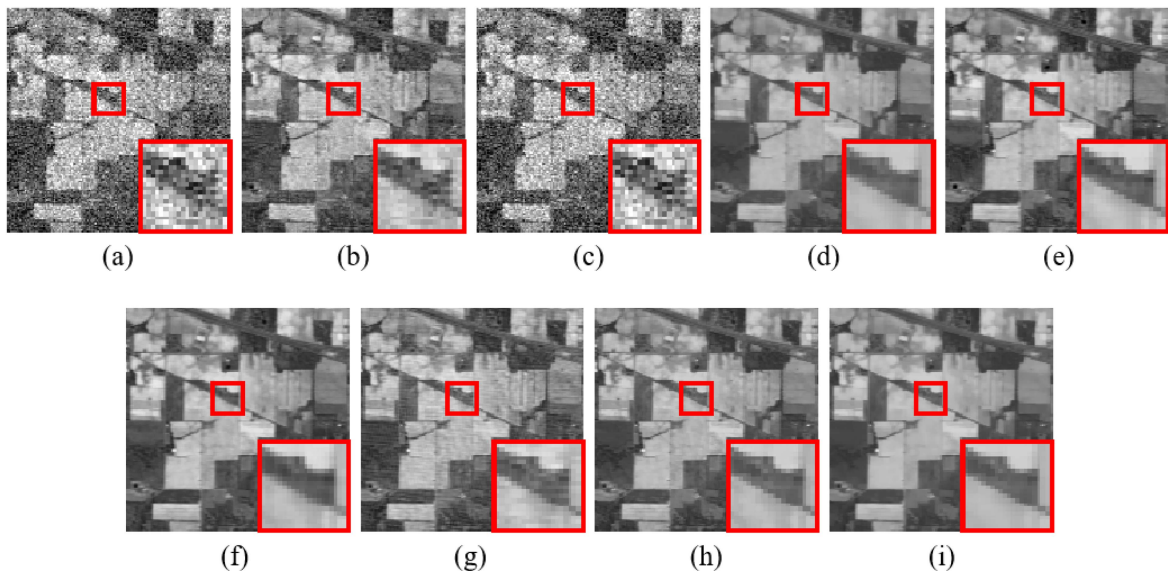


Fig. 4. Image of band 149 in AVIRIS Indian Pines dataset before and after denoising by different methods. (a) Original data, the restored results by method. (b) BM4D. (c) TDL. (d) LRTV. (e) NMoG. (f) LRTDTV. (g) LLRT. (h) RCTV. (i) Ours.

modeling. Conversely, other methods yield significant noise reduction. Compared to all the other methods, our approach exhibits the most substantial noise removal in the spatial domain and less visual image distortion.

To facilitate further comparison, Fig. 6 displays the spectral signatures of a pixel located at (24, 88) before and after restoration. The horizontal axis denotes the band number, while the vertical axis represents the digital number value of the given location. Due to the presence of noise, the curve exhibits rapid fluctuations, as shown in Fig. 6(a). After denoising, the fluctuations are substantially suppressed. In addition, we observe that

the NMoG, LLRT, and our method yield the most notable fluctuation suppression, consistent with the visual results depicted in Figs. 3–5.

### C. Parameters Setting

In the proposed denoising model, the parameters  $\lambda_1$  and  $\lambda_2$  play a crucial role in balancing the loss term and regularization terms. In our algorithm, we initially estimate the noise variance  $\sigma^2$  using the initialized  $\mathcal{X}$  and set  $\lambda_1 = \lambda_1 \sigma^2$  and  $\lambda_2 = \lambda_2 \sigma^2$ . It is important to note that the parameter selection for  $\lambda_1$  and  $\lambda_2$

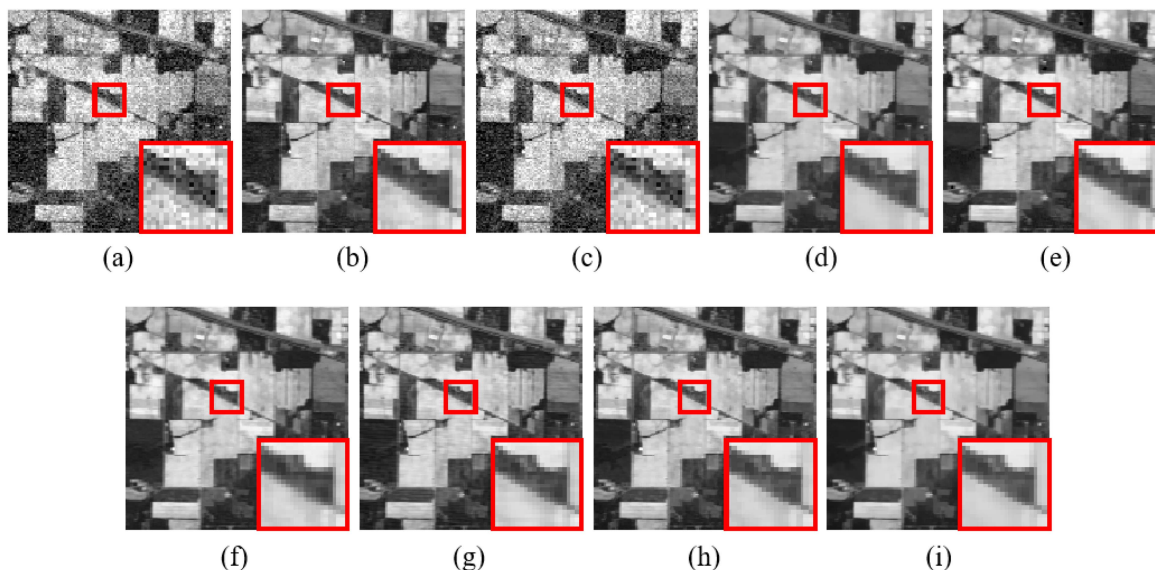


Fig. 5. Image of band 165 in AVIRIS Indian Pines dataset before and after denoising by different methods. (a) Original data, the restored results by method. (b) BM4D. (c) TDL. (d) LRTV. (e) NMoG. (f) LRTDTV. (g) LLRT. (h) RCTV. (i) Ours.

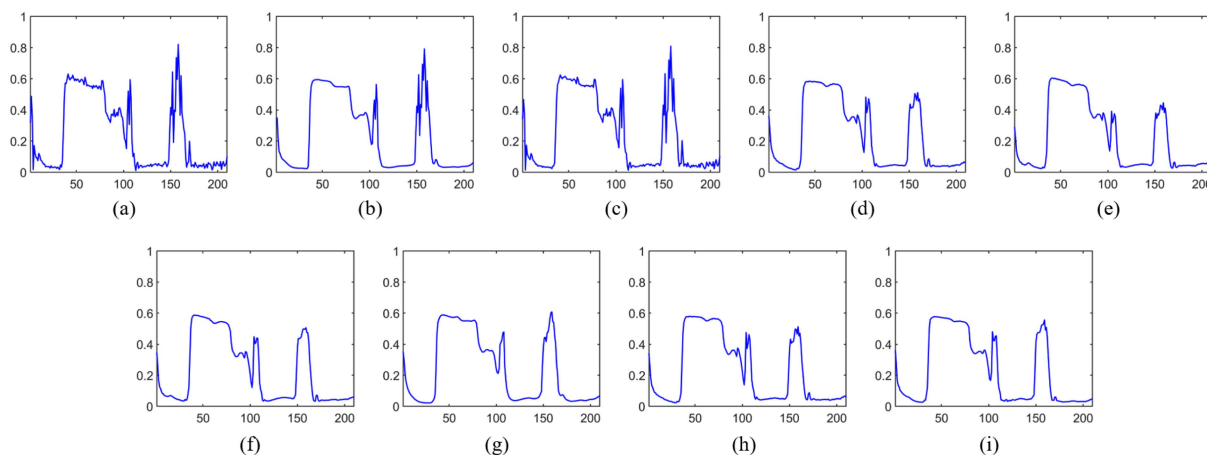


Fig. 6. Spectral signatures of point (24, 88) in AVIRIS Indian Pines dataset before and after denoising by different methods. (a) Original data, the restored results by method. (b) BM4D. (c) TDL. (d) LRTV. (e) NMoG. (f) LRTDTV. (g) LLRT. (h) RCTV. (i) Ours.

occurs before multiplying them by  $\sigma^2$ . To analyze the sensitivity of parameters, we conducted denoising experiments on the Balloons dataset under various parameter settings in the case of mixture noise. The experiment results are presented as contour maps. In Fig. 7(a), the MPSNR value is plotted against  $\lambda_1$  and  $\lambda_2$ . It is observed that MPSNR changes slowly with variations in  $\lambda_2$ , suggesting its insensitivity to this parameter. Conversely, MPSNR exhibits rapid changes with variations in  $\lambda_1$ , indicating its sensitivity to this parameter. Similarly, Fig. 7(b) shows that the value of MSSIM is also insensitive to  $\lambda_2$  but relatively sensitive to  $\lambda_1$ . However, within a large range, MSSIM remains stable at a high level. Hence,  $\lambda_2$  can be considered insensitive while  $\lambda_1$  is sensitive. Furthermore,  $\lambda_1$  represents the intensity of TV regularization. Therefore, its value can be determined

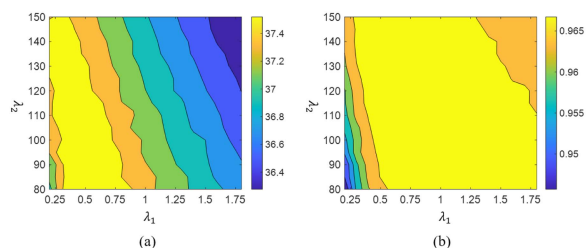


Fig. 7. Contour map of (a) MPSNR and (b) MSSIM versus  $\lambda_1$  and  $\lambda_2$ .

based on the smoothness of the image. A smoother image would require a larger value of  $\lambda_1$ , whereas a less smooth image would necessitate a smaller value of  $\lambda_1$ .

## V. CONCLUSION

In this article, we propose a novel MSI denoising model with an adaptive loss function and multiple image prior model-based regularization. Specifically, we apply the NMoG distribution to model the complex and unknown noise distribution, and then obtain the weighted  $l_2$ -norm loss function by solving the noise model with VAE algorithm. The weight involved in the loss function is adaptively and efficiently learned from observed MSI. In addition, we model the nonlocal spatial similarity, spatial and spectral correlation, and local smoothness priors comprehensively and integrate them into the regularization term. Finally, we conduct the simulation and real data experiments to demonstrate that the proposed method can effectively reduce the noise compared with the state-of-the-art methods.

## REFERENCES

- [1] A. Pelagotti, A. Del Mastio, A. De Rosa, and A. Piva, "Multispectral imaging of paintings," *IEEE Signal Process. Mag.*, vol. 25, no. 4, pp. 27–36, Jul. 2008.
- [2] R. L. Kettig and D. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," *IEEE Trans. Geosci. Electron.*, vol. 14, no. 1, pp. 19–26, Jan. 1976.
- [3] X. Su et al., "A review of pharmaceutical robot based on hyperspectral technology," *J. Intell. Robot. Syst.*, vol. 105, no. 4, 2022, Art. no. 75.
- [4] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2069, Jul. 2008.
- [5] P. Ready and P. Wintz, "Information extraction, SNR improvement, and data compression in multispectral imagery," *IEEE Trans. Commun.*, vol. 21, no. 10, pp. 1123–1131, Oct. 1973.
- [6] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising with a spatial-spectral view fusion strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2314–2325, May 2014.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.*, vol. 59, no. 12, pp. 10348–10363, Dec. 2021.
- [12] E. Pan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "SQAD: Spatial-spectral quasi-attention recurrent network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.
- [13] 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.
- [14] 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.
- [15] 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, 2021.
- [16] W. He, Q. Yao, N. Yokoya, T. Uezato, H. Zhang, and L. Zhang, "Spectrum-aware and transferable architecture search for hyperspectral image restoration," in *Proc. Eur. Conf. Comput. Vis.*, pp. 19–37, 2022.
- [17] M. Li, Y. Fu, and Y. Zhang, "Spatial-spectral transformer for hyperspectral image denoising," 2022, *arXiv:2211.14090*.
- [18] 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.
- [19] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4589–4604, Oct. 2017.
- [20] 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.
- [21] D. Meng and F. Torre, "Robust matrix factorization with unknown noise," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 1337–1344.
- [22] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang, "Robust principal component analysis with complex noise," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 55–63.
- [23] X. Cao et al., "Low-rank matrix factorization under general mixture noise distributions," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1493–1501.
- [24] X. Cao, Q. Zhao, D. Meng, Y. Chen, and Z. Xu, "Robust low-rank matrix factorization under general mixture noise distributions," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4677–4690, Oct. 2016.
- [25] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-iid noise structure," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1054–1066, Mar. 2018.
- [26] J. T. Barron, "A general and adaptive robust loss function," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4331–4339.
- [27] C.-I. Chang and Q. Du, "Interference and noise-adjusted principal components analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2387–2396, Sep. 1999.
- [28] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973–980, Mar. 2011.
- [29] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE Trans. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, Y. Chen, and W. He, "Double-factor-regularized low-rank tensor factorization for mixed noise removal in hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8450–8464, Dec. 2020.
- [34] 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.
- [35] K. E. M. Maggioni, V. Katkovnik, 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] S. H. Chan, R. Khoshabeh, K. B. Gibson, P. E. Gill, and T. Q. Nguyen, "An augmented Lagrangian method for total variation video restoration," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3097–3111, Nov. 2011.
- [40] 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.
- [41] C. Jiang, H. Zhang, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image denoising with a combined spatial and spectral weighted hyperspectral total variation model," *Can. J. Remote Sens.*, vol. 42, no. 1, pp. 53–72, 2016.

- [42] Y. Chen, W. Cao, L. Pang, and X. Cao, "Hyperspectral image denoising with weighted nonlocal low-rank model and adaptive total variation regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5544115.
- [43] 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.
- [44] J. Peng, Y. Wang, H. Zhang, J. Wang, and D. Meng, "Exact decomposition of joint low rankness and local smoothness plus sparse matrices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5766–5781, May 2023.
- [45] 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.
- [46] 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.
- [47] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. 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.
- [48] X. Bai, F. Xu, L. Zhou, Y. Xing, L. Bai, and J. Zhou, "Nonlocal similarity based nonnegative Tucker decomposition for hyperspectral image denoising," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 701–712, Mar. 2018.
- [49] L. Sun, B. Jeon, Z. Wu, and L. Xiao, "Hyperspectral denoising via cross total variation-regularized unidirectional nonlocal low-rank tensor approximation," in *Proc. IEEE 25th Int. Conf. Image Process.*, 2018, pp. 2900–2904.
- [50] Y. Chang, L. Yan, X.-L. Zhao, H. Fang, Z. Zhang, and S. Zhong, "Weighted low-rank tensor restoration for hyperspectral image restoration," *IEEE Trans. Cybern.*, vol. 50, no. 11, pp. 4558–4572, 2020.
- [51] X. Kong et al., "Hyperspectral image denoising based on nonlocal low-rank and tv regularization," *Remote Sens.*, vol. 12, no. 12, 2020, Art. no. 1956.
- [52] 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, pp. 1–16, 2021.
- [53] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.
- [54] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [55] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183–208, 2017.
- [56] J. Peng, H. Wang, X. Cao, X. Liu, X. Rui, and D. Meng, "Fast noise removal in hyperspectral images via representative coefficient total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022.
- [57] 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, 2010.



**Wenfei Cao** received the Ph.D. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 2015.

He is currently an Associated Professor with the School of Mathematics and Statistics, Shaanxi Normal University, Xi'an, China. His research interests include matrix/tensor estimation, deep learning, and image processing.



**Meiqiao Bi** received the M.S. degree in computer technology from the Hebei University of Technology, Tianjin, China, in 2016.

She is currently a Visiting Student with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. Her research interests include remote sensing scene classification, computer vision, machine learning, and image processing.



**Jing Yao** (Member, IEEE) received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 2021.

He is currently an Assistant Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. From 2019 to 2020, he was a Visiting Student with Signal Processing in Earth Observation, Technical University of Munich, Munich, Germany, and at the Remote Sensing Technology Institute, German Aerospace Center, Oberpfaffenhofen, Germany. His research interests

include hyperspectral and multimodal remote sensing image analysis, mainly including optimization and deep learning-based methods for image processing and interpretation tasks.

Dr. Yao was the recipient of the Jose Bioucas Dias Award for recognizing the outstanding paper at the Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing, in 2021. He also serves as a Guest Editor for *Remote Sensing*.



**Danfeng Hong** (Senior Member, IEEE) received the M.Sc. degree (summa cum laude) in computer vision from the College of Information Engineering, Qingdao University, Qingdao, China, in 2015, the Dr.-Ing. degree (summa cum laude) in signal processing in earth observation from Technical University of Munich, Munich, Germany, in 2019.

He is currently a Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. Before joining CAS, he has been a Research Scientist and led a Spectral Vision Working Group with the Remote Sensing Technology Institute, German Aerospace Center, Oberpfaffenhofen, Germany. He was also an Adjunct Scientist with GIPSA-lab, Grenoble INP, CNRS, Université Grenoble Alpes, Grenoble, France. His research interests include artificial intelligence, remote sensing Big Data analysis, multimodal interpretation, and their applications in Earth Vision.

Dr. Hong was the recipient of the Best Reviewer Award of the IEEE TGRS in 2021 and 2022, and the Best Reviewer Award of the IEEE JSTARS in 2022, the Jose Bioucas Dias Award for recognizing the outstanding paper at WHISPERS in 2021, the Remote Sensing Young Investigator Award in 2022, the IEEE GRSS Early Career Award in 2022, and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters) in 2022. He is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, an Editorial Board Member of *Remote Sensing*, an Editorial Advisory Board Member of the ISPRS *Journal of Photogrammetry and Remote Sensing*.



**Yang Chen** received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 2020.

She is currently an Assistant Professor with the School of Mathematics and Statistics, Shaanxi Normal University, Xi'an, China. From 2017 to 2019, she was a Visiting Scholar with University of Illinois at Urbana-Champaign, Urbana, IL, USA. Her research interests include model-based deep learning and image processing.