

# Hyperspectral Image Denoising via Combined Non-Local Self-Similarity and Local Low-Rank Regularization

HAIJIN ZENG<sup>1</sup>, XIAOZHEN XIE<sup>1</sup>, WENFENG KONG<sup>1</sup>, SHUANG CUI<sup>2</sup>,  
AND JIFENG NING<sup>2</sup>

<sup>1</sup>College of Science, Northwest A&F University, Yangling 712100, China

<sup>2</sup>College of Information Engineering, Northwest A&F University, Yangling 712100, China

Corresponding author: Xiaozhen Xie (xiexzh@nwfau.edu.cn)


This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2452019073, and in part by the National Natural Science Foundation of China under Grant 61876153.

**ABSTRACT** Hyperspectral images (HSIs) are usually corrupted by various noises during the image acquisition process, e.g., Gaussian noise, impulse noise, stripes, deadlines and many others. Such complex noise severely degrades the data quality, reduces the interpretation accuracy of HSIs, and restricts the subsequent HSI applications. In this paper, a spatial non-local and local rank-constrained low-rank regularized Plug-and-Play (NLRPNP) model is presented for mixed noise removal in HSIs. Specifically, we first divide HSIs into local overlapping patches. Local rank-constrained low-rank matrix recovery is adopted to effectively separate the low-rank clean HSI patches from the sparse noise and a part of Gaussian noise, and to significantly preserve local structure and detail information in HSIs. Then the spatial non-local based denoiser is introduced to promote the non-local self-similarity and obviously depress the Gaussian noise. Without increasing the difficulty of solving optimization problems, we combine the local and non-local based methods into the Plug-and-Play framework, and develop an efficient algorithm for solving the proposed NLRPNP model by using the alternating direction method of multipliers method. Finally, several experiments are conducted in both simulated and real data conditions to illustrate the better performance of the proposed NLRPNP model than the existing state-of-the-art denoising models.

**INDEX TERMS** Hyperspectral images, denoising, plug-and-play framework, local low-rank matrix recovery, non-local regularization.

## I. INTRODUCTION

Hyperspectral images (HSIs) can provide spectral information about hundreds of continuous bands in the same scene, hence HSIs are widely used in many fields [1], [2]. However, limited by observation conditions and imaging sensors, HSIs obtained by hyperspectral imagers are usually corrupted by a variety of noises, e.g., stripes, deadlines, impulse noise, Gaussian noise and so on [3]. The various noises concurrently result in the loss of useful information, and further limit the accuracy of subsequent processing and application, such as image classification [4], target detection [5], unmixing [6] and so on. Therefore, as a pretreatment step, HSI denoising is a valuable research topic.

The associate editor coordinating the review of this manuscript and approving it for publication was Qiangqiang Yuan .

In HSI data, each spectral channel which includes the spatial information can be viewed as a grayscale image. From this perspective, lots of denoising methods for grayscale images can be directly applied to denoise HSI data band by band. Among these denoising methods, total variation (TV) regularization [7] is an efficient tool and widely used in image denoising task. Due to its good performance in preserving the spatial piecewise smooth and important edge information of images, i.e., the spatial sparsity in images, TV regularization is introduced to the HSI denoising problem [8]. Moreover, non-local regularization is another powerful approach to describe the non-local self-similarity and the intrinsic geometry structure in images. For denoising tasks, Manjón *et al.* [9] use the non-local mean to propose NLM3D model; Dabov *et al.* [10] introduce the famous block-matched 3-D filtering (BM3D) model.

However, besides the spatial information in the spatial domain, HSI data also contains the spectral information corresponding to the wavelength. In the spectral domain, different spectral bands are images from the same scene under different wavelengths. It means they are highly correlative. This is an important prior knowledge, i.e., low-rank prior of HSI data. Hence, it is a serious drawback that the above band-wise processing ignores the low-rank prior in the spectral domain.

In order to take full advantage of the low-rank prior in HSI data, early HSI restoration method [11] uses principal component analysis (PCA) to map HSI data to a series of orthogonal vectors, and relies on the shrinkage of representation coefficients to achieve the purpose of denoising. However, it is sensitive to outliers, such as impulse noise, deadlines, stripes and so on. Therefore, Candes *et al.* [12] propose robust principal component analysis (RPCA) method to solve this problem by using nuclear norm minimization to represent low-rank prior knowledge and utilizing  $L_1$  norm minimization to detect sparse outliers. Subsequently, the RPCA method is introduced into HSI denoising problems, and has achieved satisfying results, e.g., low-rank matrix recovery (LRMR) model [13], noise adjusted iterative low-rank matrix approximation (NAILRMA) model [14] and so on. Nevertheless, they lexicographically order 3-D HSIs into 2-D Casorati matrices whose columns and rows are great different. Then nuclear norm minimization based algorithms shrink singular values of the 2-D Casorati matrices equally, resulting in nonnegligible losses of local details and texture information in denoised HSIs. For avoiding this shortcoming, one way is to exploit the local low-rank prior in HSIs [8], [15], [16]. It usually divides HSIs into local 3-D patches whose Casorati matrices have the similar columns and rows. Then the RPCA based methods implemented on the local patches. For instance, Fan *et al.* [17] proposed a bilinear low rank matrix factorization (BLRMF) HSI denoising method, where the bi-nuclear quasi-norm is employed for constraining the patches low rank characteristic in HSI. RPCA based methods can reduce not only the information loss but also the dependence on the assumption that the noise is independent and identically distributed (i.i.d.). Hence, the local RPCA methods [8], [16], [17] can handle on more complex noises. Another way is to treat the 3-D HSIs as 3rd-order tensors and directly define the tensor low-rank property. Usually, low-rank tensor decomposition methods, such as the Tucker decomposition [18], PARAFAC decomposition [19], tensor singular value decomposition [20] and so on, are used to approximately represent the low-rank prior of HSIs in both the spatial and spectral domains. Based on them, the tensor RPCA based low-rank tensor recovery approach [12], [21] is introduced into HSI denoising problems and has achieved better denoising results. Moreover, in [22] and [23], models employing spatial-spectral deep prior were proposed, in which both the spatial and spectral information are simultaneously assigned to the proposed network.

For sufficiently using both the spatial and spectral prior information, many researchers attempt to integrate the

different prior information by their according different mathematical formulations. For instance, He *et al.* [8] propose a LRTV model which uses the band-wised TV regularization to represent the spatial sparsity and simultaneously uses the nuclear norm minimization of HSIs' Casorati matrices to describe the spectral low-rank prior. Hence, LRTV model can not only remove the Gaussian noise, but also depress the sparse noise, such as impulse noise, deadlines and stripes and so on. Chang *et al.* [16] design the anisotropic spatial-spectral TV norm (ASSTV), which can represent the sparse prior in both the spatial and the spectral domains. Then, Wang *et al.* [3] introduce the ASSTV regularization into low-rank tensor recovery model and propose a LRTDTV model, which greatly improves the recovery accuracy. He *et al.* [24] combine ASSTV regularization with the local low-rank matrix recovery model and propose a LLRSSTV model which can not only remove mixed noise, but also eliminate structure-related noise partly. However, due to the difficulty in the model solution and the algorithm implementation, aforementioned models just can combine a few regularization terms linearly, e.g., TV-based regularization terms and RPCA based regularization terms. There are still lots of prior information and their according regularization terms which can not be integrated into the above denoising framework, e.g., aforementioned non-local self-similarities observed in most HSI datasets [25], [26], [26], [27].

Fortunately, a recent Plug-and-Play (PnP) framework [28] has reported empirical success on a large variety of image processing [1], [2], [29], [30]. As a non-convex framework, PnP provides an approach for flexibly integrating the impressive capabilities of existing denoising priors, such as image deblurring, denoising and so on. On the other hand, compared with local or global methods in spatial domain, non-local methods can not only utilize neighborhood pixels, but also utilize distant pixels in similar patches. At the same time, non-local methods go beyond TV-based methods when dealing with many inverse imaging problems [9], [10], [31], [32], see Fig. 1 for an example. It can be easily observed that the HSI band, e.g., the 3rd-band, is heavily polluted by noise. The latest TV-based methods [3] cannot provide results with satisfactory detail and texture information, while the non-local based method [9] can achieve better results. This motivates us to introduce the non-local based method into the HSI denoising tasks by the PnP framework. Therefore, in this paper, we propose a local low-rank regularized non-local HSI denoising model in the PnP framework. Here, the local low-rank regularization can preserve the rich details in HSIs and reduce the dependence on the i.i.d. noise assumption, which is not suitable for the actual noises in HSIs. As an implicit regularizer, the non-local regularization term introduced into the PnP framework, adequately expresses the non-local self-similarity of the underlying HSIs and can be easy to be solved. By integrating the advantages of the local low-rankness and non-local self-similarity in HSIs, the proposed model is expected to be able to effectively maintain

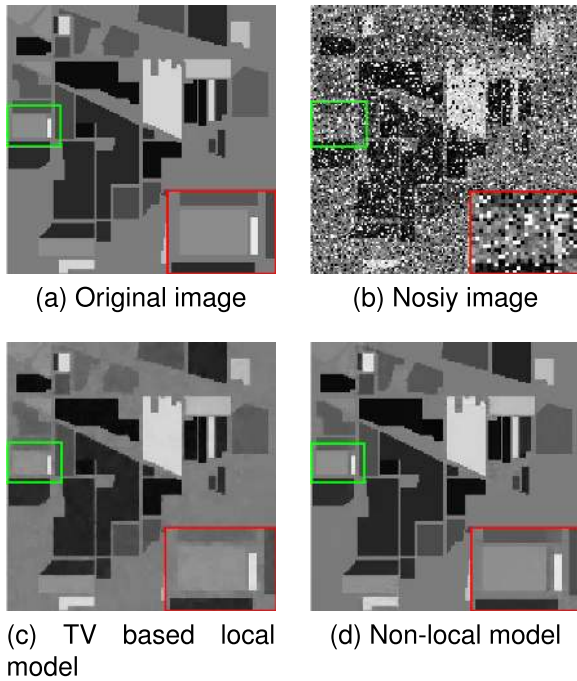


FIGURE 1. A typical restoration instance on Indian Pines data: (a-b) the original and noisy band; (c-d) denoised results by TV based model and non-local model.

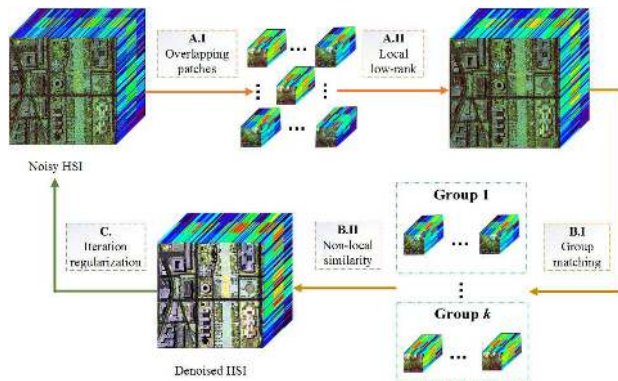


FIGURE 2. Flowchart of the proposed method. It includes three stages: A. local low-rank denoising, B. non-local low-rank denoising and C. iteration regularization.

the general structure and capture the details in the denoised HSIs.

The main contributions of this paper are summarized as follows.

- 1) A HSI denoising model combined non-local self-similarity and local low-rank regularization is proposed. To best of our knowledge, this is the first attempt to introducing both the non-local self-similarity and local low-rank prior into the HSI denoising task without increasing the difficulty of solving the model, compared to a single local low-rank or non-local low-rank model.
- 2) According to the elegant framework provided by the PnP method and the alternating direction method of multipliers (ADMM), we design an algorithm that realizes the fast solution of the proposed model.

- 3) Furthermore, in order to verify the performance of the proposed model, we conducted a lot of simulation data and real data experiments. Experiments results are reported to demonstrate the effectiveness of the proposed model and the efficiency of the numerical scheme. The rest of this paper is organized as follows. Section II gives the related works for the denoising task. The proposed model is described in Section III and its iteration solution algorithm is listed in Section IV. Section V includes experimental results and discussions. Finally, Section VI concludes the paper.

## II. RELATED WORK

### A. PROBLEM FORMULATION

In real situation, the observed HSI is corrupted by mixed-noise, which typically consists of Gaussian noise, stripes, impulse noise, deadlines and so on [14]. Let 3rd-order tensor  $\mathcal{Y} \in \mathbb{R}^{m \times n \times p}$  denotes the observed HSI, where the spatial information lies in the first two dimensions and the spectral information lies in the third dimension. Then the degradation model of the HSI can be formulated as

$$\mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N}, \quad (1)$$

where  $\mathcal{Y}, \mathcal{X}, \mathcal{S}, \mathcal{N} \in \mathbb{R}^{m \times n \times p}$ ;  $\mathcal{Y}$  denotes the observed HSI;  $\mathcal{X}$  represents the clean HSI;  $\mathcal{S}$  is the sparse noise, which consists of impulse noise, stripes, deadline and so on;  $\mathcal{N}$  is the Gaussian noise;  $m \times n$  is the spatial size of the HSI, and  $p$  is the number of spectral bands.

Under the framework of degradation model (1), HSI denoising is a process of separating the mixed noise  $\mathcal{S}, \mathcal{N}$  from the observed HSI  $\mathcal{Y}$ , and restoring the clean HSI  $\mathcal{X}$ . In mathematical theory, this is a serious ill-posed problem. The regularization method is an effective and widely used method for solving such inverse problems. It establishes the following regularized denoising framework by adding the prior information of unknown clear HSI and mixed noise, i.e.,

$$\arg \min_{\mathcal{X}} J(\mathcal{X}, \mathcal{S}, \mathcal{N}) + \beta R(\mathcal{X}), \quad (2)$$

where  $J(\mathcal{X}, \mathcal{S}, \mathcal{N})$  is a regular term to describe the distribution of different noises;  $R(\mathcal{X})$  is a regular term to represent the prior information of unknown clean HSI;  $\beta$  is a non-negative regularization parameter used to balance two regular terms. In denoising framework (2), both prior information and the formulations of regular terms are important, which determine the accuracy of the restoration results. Therefore, the research on HSI denoising mainly focuses on the exploration of prior information and the improvement of regularization formulations.

### B. PLUG-AND-PLAY FRAMEWORK

In the field of image denoising, there have been many studies using a variety of prior information and their according regularization terms. However, there are still two difficulties as following. Firstly, it is difficult to find an appropriate mathematical tool to reasonably describe the complex prior information of HSIs. Secondly, if the regularization denoising model (2) contains multiple regular terms, it will be

hard to be solved theoretically. Recently, a new PnP framework [28] provides an elegant way to overcome the second difficulty. Without increasing the difficulty of solving optimization problems, PnP can flexibly integrate impressive capabilities of multiple prior and regularization terms in one denoising model, such as image deblurring, denoising and so on. Also, it has reported empirical success on a large variety of image processing, e.g., compressive imaging [33], compression-artifact reduction [34], nonlinear inverse scattering [35]. Hence, PnP attracts more and more interest.

In the PnP framework, by using variable splitting, the regularized denoising model (2) can be compactly restated as follows

$$\arg \min_{\mathcal{X}, \mathcal{L}} J(\mathcal{X}) + \beta R(\mathcal{L}) \text{ s.t. } \mathcal{X} = \mathcal{L}.$$

According to augmented Lagrangian multiplier method, it can be rewritten as a minimization problem of the augmented Lagrangian function  $\ell(\mathcal{L}, \mathcal{X})$  as following

$$\begin{aligned} & \arg \min_{\mathcal{L}, \mathcal{X}} \ell(\mathcal{L}, \mathcal{X}) \\ & = \arg \min_{\mathcal{L}, \mathcal{X}} J(\mathcal{X}) + \beta R(\mathcal{L}) + \langle \Lambda, \mathcal{X} - \mathcal{L} \rangle + \frac{\lambda}{2} \|\mathcal{X} - \mathcal{L}\|_F^2 \\ & = \arg \min_{\mathcal{L}, \mathcal{X}} J(\mathcal{X}) + \beta R(\mathcal{L}) + \frac{\lambda}{2} \|\mathcal{X} - \mathcal{L} + \tilde{\Lambda}\|_F^2 - \frac{\lambda}{2} \|\tilde{\Lambda}\|_F^2, \end{aligned} \quad (3)$$

where  $\Lambda$  is the dual variable;  $\tilde{\Lambda} = \Lambda/\lambda$  is the scaled dual variable;  $\lambda$  is a penalty parameter;  $\|\cdot\|_F^2$  denotes the Frobenius norm. Then, the minimization problem (3) can be solved by ADMM method [36], which consists of iterations until convergence over the following three steps at the  $k$ -th iteration

$$\begin{aligned} \mathcal{X}^{(k)} &= \arg \min_{\mathcal{X}} \ell(\mathcal{L}^{(k-1)}, \mathcal{X}, \tilde{\Lambda}^{(k-1)}), \\ \mathcal{L}^{(k)} &= \arg \min_{\mathcal{L}} \ell(\mathcal{L}, \mathcal{X}^{(k)}, \tilde{\Lambda}^{(k-1)}), \\ \tilde{\Lambda}^{(k)} &= \tilde{\Lambda}^{(k-1)} + (\mathcal{X}^{(k)} - \mathcal{L}^{(k)}). \end{aligned} \quad (4)$$

By plugging (3) in (4), we have

$$\begin{aligned} \mathcal{X}^{(k)} &= \arg \min_{\mathcal{X}} J(\mathcal{X}) + \frac{\lambda}{2} \|\mathcal{X} - (\mathcal{L}^{(k-1)} + \tilde{\Lambda}^{(k-1)})\|_F^2, \\ \mathcal{L}^{(k)} &= \arg \min_{\mathcal{L}} R(\mathcal{L}) + \frac{\lambda}{2\beta} \|(\mathcal{X}^{(k)} + \tilde{\Lambda}^{(k-1)}) - \mathcal{L}\|_F^2, \\ \tilde{\Lambda}^{(k)} &= \tilde{\Lambda}^{(k-1)} + (\mathcal{X}^{(k)} - \mathcal{L}^{(k)}). \end{aligned} \quad (5)$$

Let  $\mathcal{Z} = \mathcal{X}^{(k)} + \tilde{\Lambda}^{(k-1)}$ , the minimization problem with respect to  $\mathcal{L}$  in (5) can be rewritten as

$$\arg \min_{\mathcal{L}} R(\mathcal{L}) + \frac{\lambda}{2\beta} \|\mathcal{L} - \mathcal{Z}\|_F^2. \quad (6)$$

One can find that minimization problem (6) is a classical denoising model with the input noisy data  $\mathcal{Z}$ , the positive parameter  $\lambda/\beta$ , the prior regularization term  $R(\mathcal{L})$  and the fidelity term  $\frac{1}{2} \|\mathcal{L} - \mathcal{Z}\|_F^2$ .

Let  $\sigma^2 = \beta/\lambda$ , we can abstractly denote the solution of minimization problem (6) as

$$\mathcal{L}^* = D(\mathcal{Z}, \sigma), \quad (7)$$

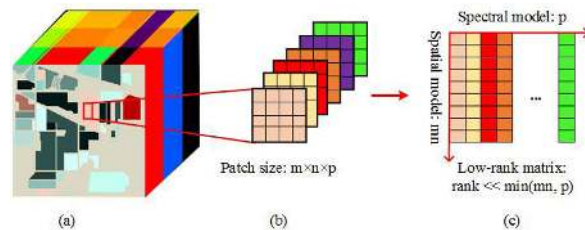


FIGURE 3. Formulation of the low-rank matrix from an HSI patch.

where  $D(\cdot, \cdot)$  is a denoising operator, which is different according to different denoising methods. For example, if the regularizer  $R(\mathcal{L})$  is a TV norm, we can use the corresponding TV-based noise reducer to solve the denoising problem. In the PnP framework,  $R(\mathcal{L})$ , as an implicit regularizer, expresses a kind of prior information about the ideal image and can be any existing denoiser. This is the main idea of the PnP framework. There are many state-of-the-art noise reducers that can be incorporated into this framework, such as BM3D [10], BM4D [31], NLM [32] and so on.

### C. LOCAL RPCA

The spectral dimension of HSIs contains imaging results of the same spatial scene in different spectral bands, so there is high correlation between the different spectral channels, i.e., the low-rank characteristic of the spectrum in HSIs. In [13], the 3-D HSI is unfolded into a 2-D Casorati matrix, and then the rank of the constrained decomposition matrix is used to represent the low-rank prior of spectrum. Based on RPCA method [12], literature [8] uses nuclear norm minimization to approximate the low-rank prior of spectrum, and establishes the rank-constrained RPCA restoration model as follows

$$\begin{aligned} & \arg \min_{X, S} \|X\|_* + \lambda \|S\|_1, \\ & \text{s.t. } \|Y - X - S\|_F^2 \leq \varepsilon, \text{rank}(X) \leq r, \end{aligned}$$

where  $X, Y, S$  are the Casorati matrices of  $\mathcal{X}, \mathcal{Y}, \mathcal{S}$ ;  $r$  is a rank-constrained parameter;  $\varepsilon$  and  $\lambda$  are positive parameters;  $\|\cdot\|_*$  denotes the nuclear norm and  $\|\cdot\|_1$  denotes the  $L_1$  norm. The RPCA based models have been widely used for the HSI restoration problems [8], [13], [14], but they significantly lose the structure, texture and detail information in HSIs due to the global low-rank constraint.

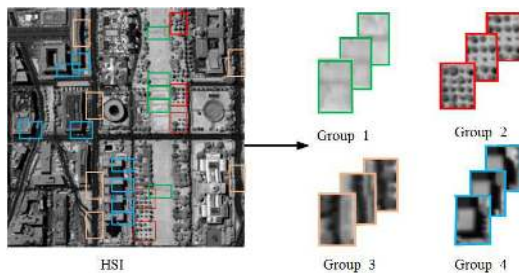
Subsequently, literatures [10], [24] find that local pixels of HSIs are more likely to belong to the same surface covering and their correlations in the spectral dimension are higher. Therefore, HSIs can be divided into many overlapping small blocks, and the RPCA based methods can be implemented on the local blocks to preserve the detail information, as shown in Fig. 3. Specifically, we define an operator  $P_{i,j} : \mathcal{X} \rightarrow \mathcal{X}_{i,j}$ . This binary operator is used to extract a block  $\mathcal{X}_{i,j} \in \mathbb{R}^{m_1 \times n_1 \times p}$  from the HSI  $\mathcal{X} \in \mathbb{R}^{m \times n \times p}$ , where the spatial size of  $m_1 \times n_1$  is centralized at pixel  $(i, j)$  of HSI data,  $m_1 \times n_1$  is approximately equal to  $p$  and  $(i, j) \in [1, m - m_1 + 1] \times [1, n - n_1 + 1]$ .  $P_{i,j}^T$  is the inverse of  $P_{i,j}$ .  $\mathbf{X}_{i,j}$  denotes the

Casorati matrix of  $\mathcal{X}_{i,j}$ . Based on these definitions, we can get the patch-based local RPCA model

$$\begin{aligned} & \arg \min_{\mathcal{X}_{i,j}, \mathcal{S}_{i,j}} \sum_{i,j} (\lambda \|\mathcal{S}_{i,j}\|_1 + \|\mathcal{X}_{i,j}\|_*) \\ & \text{s.t. } \|\mathcal{Y}_{i,j} - \mathcal{X}_{i,j} - \mathcal{S}_{i,j}\|_F^2 \leq \varepsilon, \text{rank}(\mathcal{X}_{i,j}) \leq r. \end{aligned} \quad (8)$$

### III. THE PROPOSED HSI DENOISING MODEL

Despite the good restoration performance of the patch-based local low-rank models [13], [24], they only explore the local low-rank property in HSIs and ignore the non-local self-similarity, which is an very important prior in most HSI datasets [26], [26], [27]. One can find an illustrative example of non-local self-similarity of one HSI dataset in Fig. 4. As presented in Fig. 4, each patch with obvious features has many similar patches, which can be clustered in one group. The patches in one group exhibit perfect mutual similarity, i.e., the non-local self-similarity. Then, the non-local base methods [10], [21], [31], [37] use this prior to denoise a group patches together. In addition, it has been shown in [10], [32] that non-local methods go beyond TV based methods when dealing with many inverse imaging problems. Also, one can see Fig. 1 as an example for the denoising task.



**FIGURE 4.** Simple example of the spatial non-local similarity in the 80th band image of Washington DC mall HSI dataset, where for each reference block there exist perfectly similar ones.

Motivated by the advantages of the patch based local low-rank models and non-local based methods, and without increasing the difficulty of solving optimization problems, we combine the two kinds of methods into the PnP framework and propose our non-local regularized local low-rank PnP (NLRPNP) model for HSI denoising problems. Specifically, let  $J(\cdot)$  in (3) be the local RPCA model (8), and the implicit regularizer  $R(\cdot)$  in (3) be a non-local based denoiser which is BM3D method in this paper. Based on the degradation model of the HSI (1), our denoising model can be formulated as

$$\begin{aligned} & \arg \min_{\mathcal{X}, \mathcal{L}, \mathcal{S}} \sum_{i,j} (\|\mathcal{X}_{i,j}\|_* + \lambda \|\mathcal{S}_{i,j}\|_1) + \beta R(\mathcal{L}), \\ & \text{s.t. } \mathcal{X} = \mathcal{L}, \|\mathcal{Y}_{i,j} - \mathcal{X}_{i,j} - \mathcal{S}_{i,j}\|_F^2 \leq \varepsilon, \text{rank}(\mathcal{X}_{i,j}) \leq r, \end{aligned} \quad (9)$$

where  $\lambda$  and  $\beta$  are regularization parameters. The overview of the proposed model is shown in Fig. 2.

It is worth noting that the proposed model can fully capture the local low-rankness and non-local self-similarity of

HSIs, and thus be expected to maintain the general structures and capture the local details of the HSI effectively. Specifically, the nuclear norm minimization of local patches is used to explore the local low-rank prior and to preserve the local details, while the implicit regularizer is used to constrain the non-local self-similarity of the HSI. The rank-constrained minimization of  $L_1$  norm and nuclear norm can effectively remove the sparse noise and partly depress the Gaussian noise. The minimization of Frobenius norm can eliminate the Gaussian noise. In addition, the patch based method can reduce the dependence of our model on the i.i.d. noise assumption, which is not suitable for actual noises in HSIs.

### IV. OPTIMIZATION PROCEDURE

To efficiently solve the proposed NLRPNP denoising model, we first introduce auxiliary variable  $\mathcal{J} \in \mathbb{R}^{m \times n \times p}$ . Then our model (9) can be rewritten as

$$\begin{aligned} & \arg \min_{\mathcal{X}, \mathcal{L}, \mathcal{J}, \mathcal{S}} \sum_{i,j} (\|\mathcal{X}_{i,j}\|_* + \lambda \|\mathcal{S}_{i,j}\|_1) + \beta R(\mathcal{L}) \\ & \text{s.t. } \mathcal{X}_{i,j} = \mathcal{J}_{i,j}, \mathcal{J} = \mathcal{L}, \text{rank}(\mathcal{X}_{i,j}) \leq r, \\ & \|\mathcal{Y}_{i,j} - \mathcal{X}_{i,j} - \mathcal{S}_{i,j}\|_F^2 \leq \varepsilon. \end{aligned}$$

This problem can be efficiently solved by the augmented Lagrange multiplier (ALM) method, which minimizes the following augmented Lagrangian function

$$\begin{aligned} & \arg \min_{\mathcal{X}, \mathcal{L}, \mathcal{J}, \mathcal{S}} \ell(\mathcal{X}, \mathcal{S}, \mathcal{J}, \mathcal{L}) \\ & = \arg \min_{\mathcal{X}, \mathcal{L}, \mathcal{J}, \mathcal{S}} \sum_{i,j} (\|\mathcal{X}_{i,j}\|_* + \lambda \|\mathcal{S}_{i,j}\|_1 \\ & \quad + \langle \Gamma_{i,j}^{\mathcal{Y}}, \mathcal{Y}_{i,j} - \mathcal{X}_{i,j} - \mathcal{S}_{i,j} \rangle \\ & \quad + \frac{\mu}{2} \|\mathcal{Y}_{i,j} - \mathcal{X}_{i,j} - \mathcal{S}_{i,j}\|_F^2 \\ & \quad + \langle \Gamma_{i,j}^{\mathcal{J}}, \mathcal{X}_{i,j} - \mathcal{J}_{i,j} \rangle + \frac{\mu}{2} \|\mathcal{X}_{i,j} - \mathcal{J}_{i,j}\|_F^2) \\ & \quad + \langle \Gamma, \mathcal{L} - \mathcal{J} \rangle + \frac{\mu}{2} \|\mathcal{L} - \mathcal{J}\|_F^2 + \beta R(\mathcal{L}), \\ & \text{s.t. } \text{rank}(\mathcal{X}_{i,j}) \leq r, \end{aligned}$$

where  $\Gamma_{i,j}^{\mathcal{Y}}, \Gamma_{i,j}^{\mathcal{J}}, \Gamma$  are the Lagrangian multipliers;  $\mu$  is the penalty parameter. The above minimization of the augmented Lagrangian function can be solved by the ADMM method. At the  $(k-1)$ -th iteration, its solution at the next iteration can be transformed into following two subproblems

$$\begin{aligned} (\mathcal{X}^{(k)}, \mathcal{S}^{(k)}) &= \arg \min_{\mathcal{X}, \mathcal{S}} \ell(\mathcal{X}, \mathcal{S}, \mathcal{J}^{(k-1)}) \\ & \text{s.t. } \text{rank}(\mathcal{X}_{i,j}) \leq r. \end{aligned} \quad (10)$$

$$(\mathcal{J}^{(k)}, \mathcal{L}^{(k)}) = \arg \min_{\mathcal{J}, \mathcal{L}} \ell(\mathcal{X}^{(k)}, \mathcal{J}, \mathcal{L}). \quad (11)$$

**A. OPTIMIZATION SUBPROBLEM FOR  $(\mathcal{X}, \mathcal{S})$**

With the other parameters fixed, the subproblem (10) for  $(\mathcal{X}, \mathcal{S})$  can be reformulated as

$$\begin{aligned} \arg \min_{\mathcal{S}, \mathcal{X}} \ell(\mathcal{X}, \mathcal{S}, \mathcal{J}) &= \sum_{ij} (\|\mathcal{X}_{i,j}\|_* + \lambda \|\mathcal{S}_{i,j}\|_1 \\ &+ \langle \Gamma_{i,j}^{\mathcal{Y}}, \mathcal{Y}_{i,j} - \mathcal{X}_{i,j} - \mathcal{S}_{i,j} \rangle \\ &+ \frac{\mu}{2} \|\mathcal{Y}_{i,j} - \mathcal{X}_{i,j} - \mathcal{S}_{i,j}\|_F^2 \\ &+ \langle \Gamma_{i,j}^{\mathcal{J}}, \mathcal{X}_{i,j} - \mathcal{J}_{i,j} \rangle + \frac{\mu}{2} \|\mathcal{X}_{i,j} - \mathcal{J}_{i,j}\|_F^2) \\ \text{s.t. rank}(\mathcal{X}_{i,j}) &\leq r. \end{aligned} \quad (12)$$

To solve the subproblem (12), we perform rank-constrained RPCA method on each patch separately and accumulate a weighted sum of  $(\mathcal{X}_{i,j}, \mathcal{S}_{i,j})$  to reconstruct  $(\mathcal{X}, \mathcal{S})$ . The optimization for each  $(\mathcal{X}_{i,j}, \mathcal{S}_{i,j})$  can be reformulated as

$$\begin{aligned} \arg \min_{\mathcal{S}, \mathcal{X}} \|\mathcal{X}_{i,j}\|_* + \lambda \|\mathcal{S}_{i,j}\|_1 \\ &+ \langle \Gamma_{i,j}^{\mathcal{Y}}, \mathcal{Y}_{i,j} - \mathcal{X}_{i,j} - \mathcal{S}_{i,j} \rangle \\ &+ \frac{\mu}{2} \|\mathcal{Y}_{i,j} - \mathcal{X}_{i,j} - \mathcal{S}_{i,j}\|_F^2 \\ &+ \langle \Gamma_{i,j}^{\mathcal{J}}, \mathcal{X}_{i,j} - \mathcal{J}_{i,j} \rangle + \frac{\mu}{2} \|\mathcal{X}_{i,j} - \mathcal{J}_{i,j}\|_F^2 \\ \text{s.t. rank}(\mathcal{X}_{i,j}) &\leq r. \end{aligned} \quad (13)$$

We alternately update the two variables  $\mathcal{X}_{i,j}$  and  $\mathcal{S}_{i,j}$ , then the optimization problem (13) can be separated into two simpler minimization subproblems.

a) With  $\mathcal{S}_{i,j}$  fixed, the minimization subproblem for  $\mathcal{X}_{i,j}$  can be deduced from (13) as following

$$\begin{aligned} \mathcal{X}_{i,j}^{(k)} &= \arg \min_{\mathcal{X}_{i,j}} \|\mathcal{X}_{i,j}\|_* + \frac{2\mu}{2} \left\| \mathcal{X}_{i,j} - \frac{1}{2} (\mathcal{Y}_{i,j} + \mathcal{J}_{i,j} \right. \\ &\quad \left. - \mathcal{S}_{i,j} + \frac{\Gamma_{i,j}^{\mathcal{Y}} + \Gamma_{i,j}^{\mathcal{J}}}{\mu}) \right\|_F^2 \\ \text{s.t. rank}(\mathcal{X}_{i,j}) &\leq r. \end{aligned} \quad (14)$$

For the sake of simplicity, we denote the iteration of  $\mathcal{X}_{i,j}$  in (14) as

$$\begin{aligned} \mathcal{X}_{i,j}^{(k)} &= \arg \min_{\mathcal{X}_{i,j}} \|\mathcal{X}_{i,j}\|_* + \frac{1}{2} \times 2\mu \|\mathcal{X}_{i,j} - \mathcal{U}_{i,j}\|_F^2, \\ \text{s.t. rank}(\mathcal{X}_{i,j}) &\leq r, \end{aligned}$$

where  $\mathcal{U}_{i,j} = \frac{1}{2} (\mathcal{Y}_{i,j} + \mathcal{J}_{i,j} - \mathcal{S}_{i,j} + (\Gamma_{i,j}^{\mathcal{Y}} + \Gamma_{i,j}^{\mathcal{J}}) / \mu)$ . In [38], the updating of  $\mathcal{X}_{i,j}$  has a closed-form solution which is shown in Lemma 1.

Lemma 1 [38]: Supposing  $\mathbf{W}$  is a matrix of size  $m_1 n_1 \times p$ , and the singular value of matrix  $\mathbf{W}$  of rank  $r$  is decomposed into

$$\mathbf{W} = \mathbf{U} \mathbf{E}_r \mathbf{V}^*, \mathbf{E}_r = \text{diag}(\{\sigma_i\}_{1 \leq i \leq r})$$

The singular value shrinkage operator then obeys

$$\text{SH}_\delta(\mathbf{W}) = \arg \min_{\text{rank}(\mathbf{X}) \leq r} \delta \|\mathbf{X}\|_* + \frac{1}{2} \|\mathbf{X} - \mathbf{W}\|_F^2,$$

where

$$\text{SH}_\delta(\mathbf{W}) = \mathbf{U} \text{diag} \{ \max((\sigma_i - \delta), 0) \} \mathbf{V}^*.$$

Using Lemma 1, it is easy to get

$$\mathbf{X}_{i,j}^{(k)} := \text{SH}_{\frac{1}{2\mu}}(\mathbf{U}_{i,j}), \quad (15)$$

where  $\mathbf{X}_{i,j}$  and  $\mathbf{U}_{i,j}$  are Casorati matrices of  $\mathcal{X}_{i,j}$  and  $\mathcal{U}_{i,j}$  respectively.

b) With  $\mathcal{X}_{i,j}$  fixed, the minimization subproblem for  $\mathcal{S}_{i,j}$  can be deduced from (13) as following

$$\arg \min_{\mathcal{S}_{i,j}} \lambda \|\mathcal{S}_{i,j}\|_1 + \frac{\mu}{2} \|\mathcal{S}_{i,j} - (\mathcal{Y}_{i,j} - \mathcal{X}_{i,j} + \Gamma_{i,j}^{\mathcal{Y}} / \mu)\|_F^2.$$

The solution of the above optimization subproblem can be directly obtained by the soft threshold

$$\mathcal{S}_{i,j}^{(k)} = \mathfrak{R}_{\lambda/\mu}(\mathcal{Y}_{i,j} - \mathcal{X}_{i,j} + \Gamma_{i,j}^{\mathcal{Y}} / \mu), \quad (16)$$

where

$$\mathfrak{R}_\Delta(x) = \begin{cases} x - \Delta, & \text{if } x > \Delta \\ x + \Delta, & \text{if } x < -\Delta \\ 0, & \text{otherwise.} \end{cases}$$

**B. OPTIMIZATION SUBPROBLEM FOR  $(\mathcal{J}, \mathcal{L})$**

a) With the other parameters fixed, the subproblem (11) for  $\mathcal{J}$  can be reformulated as

$$\begin{aligned} \arg \min_{\mathcal{J}} \frac{\mu}{2} \|\mathcal{J} - \mathcal{X} + \Gamma / \mu\|_2^2 \\ + \sum_{ij} \left( \frac{\mu}{2} \|\mathcal{X}_{i,j} - \mathcal{J}_{i,j} + \Gamma_{i,j}^{\mathcal{J}} / \mu\|_F^2 \right). \end{aligned}$$

It is a convex function, which has the following closed form solution

$$\begin{aligned} \mathcal{J}^{(k)} &= \left( \mathcal{X} - \Gamma / \mu + \sum_{ij} \mathbf{P}_{i,j}^T (\mathcal{X}_{i,j} + \Gamma_{i,j}^{\mathcal{J}} / \mu) \right) \\ &\quad ./ \left( 1 + \sum_{ij} \mathbf{P}_{i,j}^T \mathbf{P}_{i,j} \right). \end{aligned} \quad (17)$$

b) With the other parameters fixed, let  $\sigma = \sqrt{\beta/\mu}$ ,  $\hat{\mathcal{J}} = \mathcal{J} + \Gamma / \mu$ , the subproblem (11) for  $\mathcal{L}$  can be reformulated as

$$\begin{aligned} \mathcal{L}^{(k)} &= \arg \min_{\mathcal{L}} \frac{\mu}{2} \|\mathcal{L} - \hat{\mathcal{J}}\|_F^2 + \beta \mathbf{R}(\mathcal{L}) \\ &= \arg \min_{\mathcal{L}} \frac{\mu}{2\beta} \|\mathcal{L} - \hat{\mathcal{J}}\|_F^2 + \mathbf{R}(\mathcal{L}) \\ &= \mathbf{D}(\hat{\mathcal{J}}, \sigma), \end{aligned} \quad (18)$$

where  $\mathbf{D}$  is defined in (7) and denotes the BM3D denoising method. Note that the denoiser parameter  $\sigma$  is linked to the noise level in i.i.d. Gaussian denoising, but in our model the  $\sigma$  is linked to the general system error between  $\hat{\mathcal{J}}$  and the ground truth. Thus, in our model  $\sigma$  is treated as a tunable parameter to obtain an appropriate effect.

### C. UPDATING THE LAGRANGIAN PARAMETERS

After solving the two subproblems (10) and (11), the Lagrangian multipliers  $\Gamma_{i,j}^{\mathcal{Y}}$ ,  $\Gamma_{i,j}^{\mathcal{J}}$  and  $\Gamma$  can be updated in parallel as

$$\begin{cases} \Gamma = \Gamma + \mu(\mathcal{J} - \mathcal{L}), \\ \Gamma_{i,j}^{\mathcal{Y}} = \Gamma_{i,j}^{\mathcal{Y}} + \mu(\mathcal{Y}_{i,j} - \mathcal{X}_{i,j} - \mathcal{S}_{i,j}), \\ \Gamma_{i,j}^{\mathcal{J}} = \Gamma_{i,j}^{\mathcal{J}} + \mu(\mathcal{X}_{i,j} - \mathcal{J}_{i,j}). \end{cases} \quad (19)$$

Summarizing the optimization strategy of step-by-step iteration as above, the solution of the NLRPNP model proposed by this paper can be obtained in Algorithm 1. Further, we discuss the complexity of the proposed models. Calculating  $\mathcal{X}_{i,j}$  has a complexity of  $\mathcal{O}(pm_1^2n_1^2 + p^2m_1n_1)$ , the complexity of updating  $\mathcal{S}_{i,j}$  is  $\mathcal{O}(m_1n_1p)$ , the complexity of updating  $\mathcal{J}$  is  $\mathcal{O}(mnp + m_1n_1p)$ , the complexity of updating  $\mathcal{L}$  is  $\mathcal{O}(mnp)$ . Therefore, the total complexity of NLRPNP is  $\mathcal{O}(pm_1^2n_1^2 + p^2m_1n_1 + m_1n_1p + mnp)$ .

---

#### Algorithm 1 HSI Denoising With the NLRPNP Model

---

**Require:**  $m \times n \times p$  observed HSI  $\mathcal{Y}$ , patch size  $m_1 \times n_1$ , stopping criterion  $\varepsilon$ , regularization parameters  $\lambda$ ,  $\beta$ , and desired rank  $r$ .

**Ensure:** Denoised image  $\mathcal{X}$ ;

**Initialize:**  $\mathcal{X} = \mathcal{L} = \mathcal{S} = \mathcal{J} = 0$ ,  $\Gamma = 0$ ,  $\Gamma_{i,j}^{\mathcal{Y}} = \Gamma_{i,j}^{\mathcal{J}} = 0$ ,  $\mu = 10^{-2}$ ,  $\mu_{\max} = 10^6$ ,  $\rho = 1.5$  and  $k = 0$ ;

**Repeat until convergence**

Update all  $(\mathcal{X}_{i,j}, \mathcal{S}_{i,j})$  patches by (15) and (16) respectively;

Update  $(\mathcal{J}, \mathcal{L}, )$  by (17) and (18) respectively;

Update the Lagrangian multipliers by (19);

Update the penalty parameter by  $\mu := \min(\rho\mu, \mu_{\max})$ ;

Check the convergence condition:

$$\max \left\{ \left\| \mathcal{Y}_{i,j} - \mathcal{X}_{i,j}^{(k)} - \mathcal{S}_{i,j}^{(k)} \right\|_{\infty}, \left\| \mathcal{J}^{(k)} - \mathcal{L}^{(k)} \right\|_{\infty} \right\} \leq \varepsilon.$$


---

## V. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, the NLRPNP method is applied to both simulation and real experiments for verifying its HSI denoising performance. In order to comprehensively evaluate the performance of our proposed model, three kinds of denoising methods are used for the comparison. They are the traditional image denoising methods, e.g., NAILRMA [14], LRMR [13], LRTV [8]; local or non-local based denoising methods, e.g., NLM3D [9], BM3D [10], LRSSTV [24]; and tensor-based methods, e.g., LRTA [18], LRTDTV [3], BM4D [31]. In particular, since the BM3D and BM4D methods are only suitable for removing Gaussian noise, and most of the noise added to HSIs in the simulation experiment is mixed noise, the mixed noise is preprocessed by the RPCA method in advance for fairly comparing the effectiveness of each method.

The parameters of these comparison methods are manually adjusted to the optimal ones according to their corresponding papers. In addition, in order to facilitate the calculation,

the HSI data is normalized to  $[0, 1]$  before performing the denoising algorithms. After denoising, the denoised HSI data is converted to the original gray level. All of the simulation and real experiments are done in MATLAB R2018b on a laptop of 64GB RAM, Intel (R) Core (TM) i7-8850H CPU, @2.20GHz.

### A. SIMULATED DATA EXPERIMENTS

In this section, simulation experiments are designed and performed. In order to verify the universality of the proposed model for different data, two HSI datasets, which have been frequently used to interpret the denoising performance of different models [3], [24], are selected to carry on the simulation experiments. The first dataset is from the Washington DC Mall of the HYDICE sensor [39], and the second one is the Indian pines dataset from the USGS spectral library [39]. From the Washington DC Mall dataset, we choose a sub-blocks with a size of  $256 \times 256 \times 191$  for experiments. And from the USGS Indian pines dataset, we select a sub-blocks with a size of  $145 \times 145 \times 224$  for experiments. Fig. 6 lists the two selected HSIs.

To simulate the noisy HSI data observed in reality, we add several types of noise to the two selected HSIs under 8 different cases to test the performance of all compared denoising methods, both in visual quality and quantitative perspective. We list the details of the 8 noise cases as following:

Case 1): In this case, we add Gaussian noise and impulse noise with the same intensity to different bands. The mean value of Gaussian noise is zero and its variance is 0.05. The percentage of impulse noise is 0.1.

Case 2): In this case, for different bands, the noise intensity of Gaussian and impulse noise is equal too. However, the noise intensity is stronger than case 1). The variance of Gaussian white noise is 0.075, and the percentage of impulse noise is 0.15.

Case 3): In this case, similarly, the same intensity noise is added to all the bands. However, the noise intensity is stronger than case 2). Specifically, the variance of Gaussian white noise is increased to 0.1, while the percentage of impulse noise is increased to 0.2.

Case 4): In this case, we only add Gaussian white noise with 0.1 variance to the clean HSI, to verify the removal performance of the proposed model for a single Gaussian noise.

Case 5): In practice, the noise intensity in each band is also different in real HSIs, and the HSIs are not only contaminated by a single noise. To simulate this case, we also add Gaussian noise and impulse noise into HSIs. However, the variance of Gaussian white noise and the percentages of impulse noise in each band are randomly selected from 0 to 0.2.

Case 6): Based on Case 5), deadlines are additionally added from band 70 to band 100 in Washington DC Mall dataset, and from band 111 to band 130 in USGS Indian Pines dataset. The number of deadlines in each band is randomly selected from 3 to 10, and the pixel width of deadlines is randomly generated from 1 to 3.

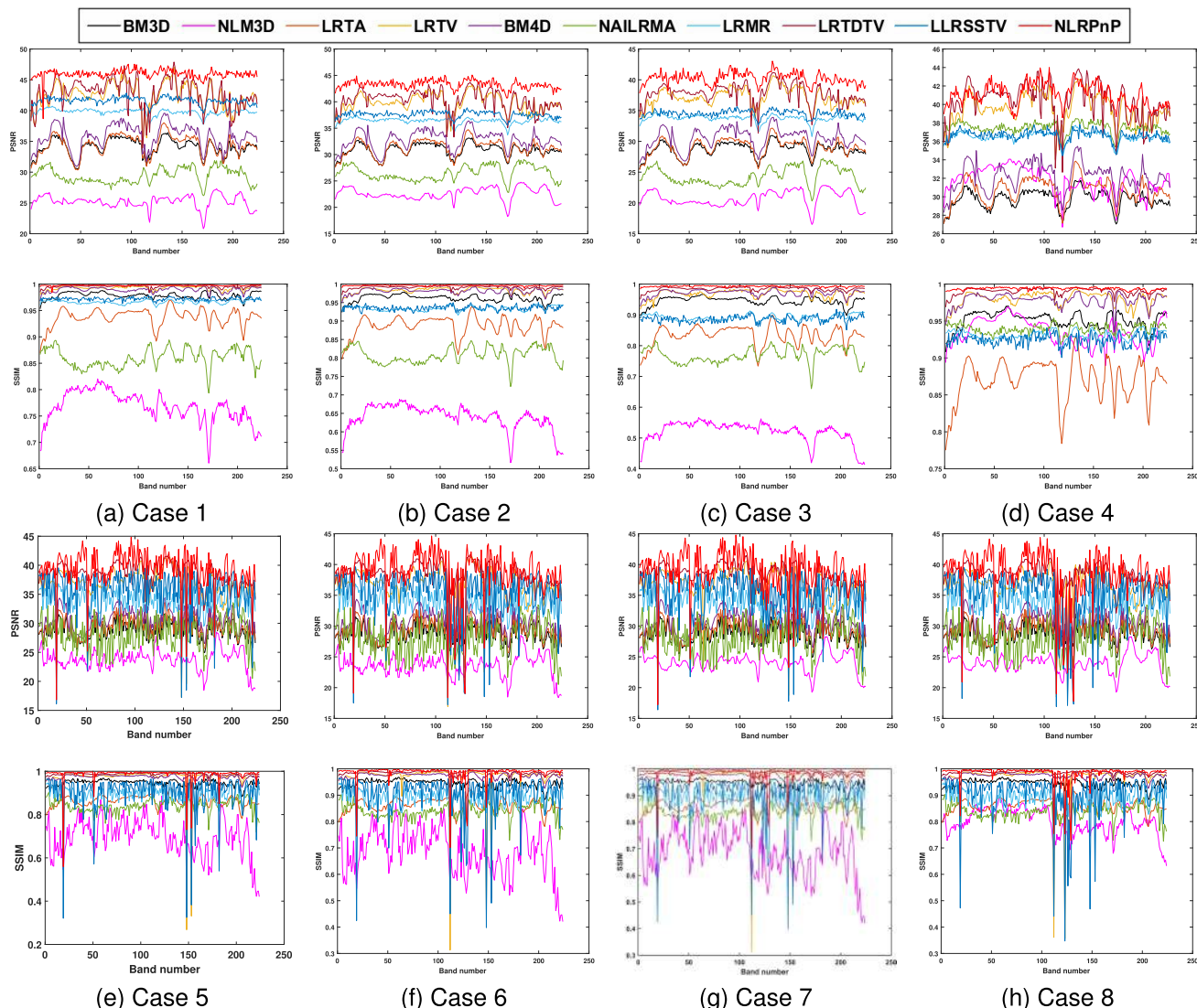


FIGURE 5. Each column shows the PSNR, SSIM values of each band of all methods under the 8 noise cases in USGS Indian Pines dataset.

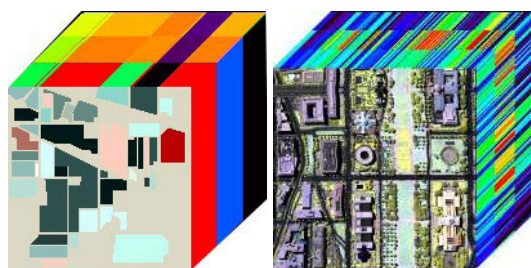


FIGURE 6. Datasets used in the simulated experiment (a) USGS Indian Pines dataset (R: 46, G: 88, B: 91). (b) HYDICE Washington DC Mall (R: 60, G: 80, B: 120).

Case 7): Based on Case 5), some stripes are additionally added from band 60 to band 90 in Washington DC Mall dataset, and from band 121 to band 140 in USGS Indian Pines dataset. The number of stripes in each band is randomly selected from 20 to 40.

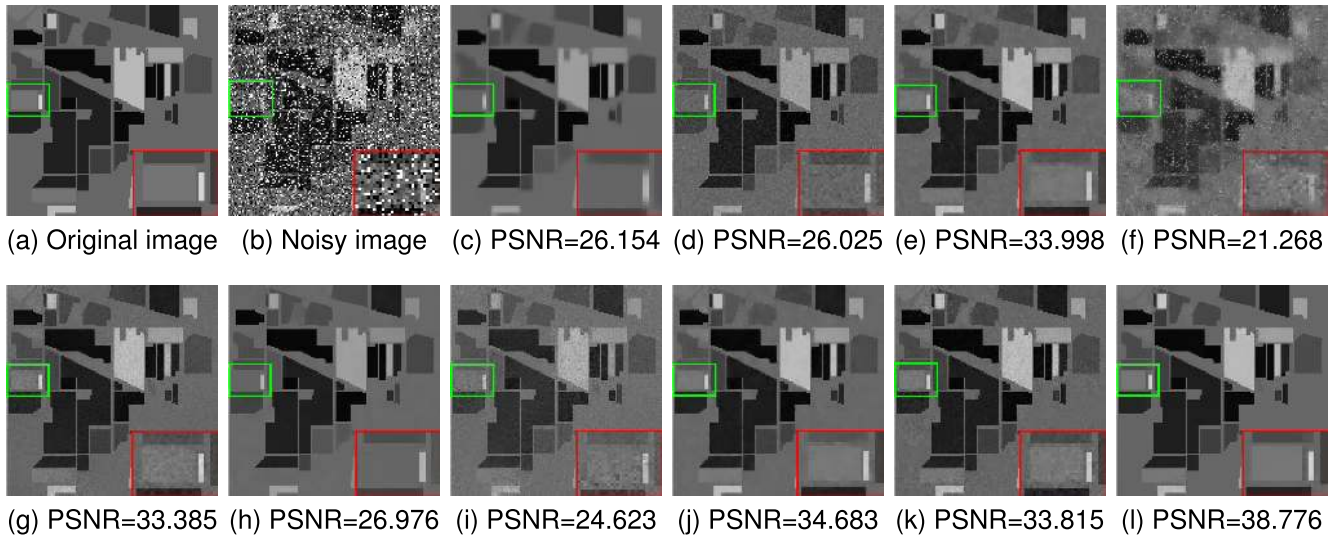
Case 8): In this case, the Gaussian noise and impulse noise in Case 5), deadlines in Case 6) and stripe noise in Case 7) are simultaneously added to the clean HSIs.

### 1) VISUAL QUALITY COMPARISON

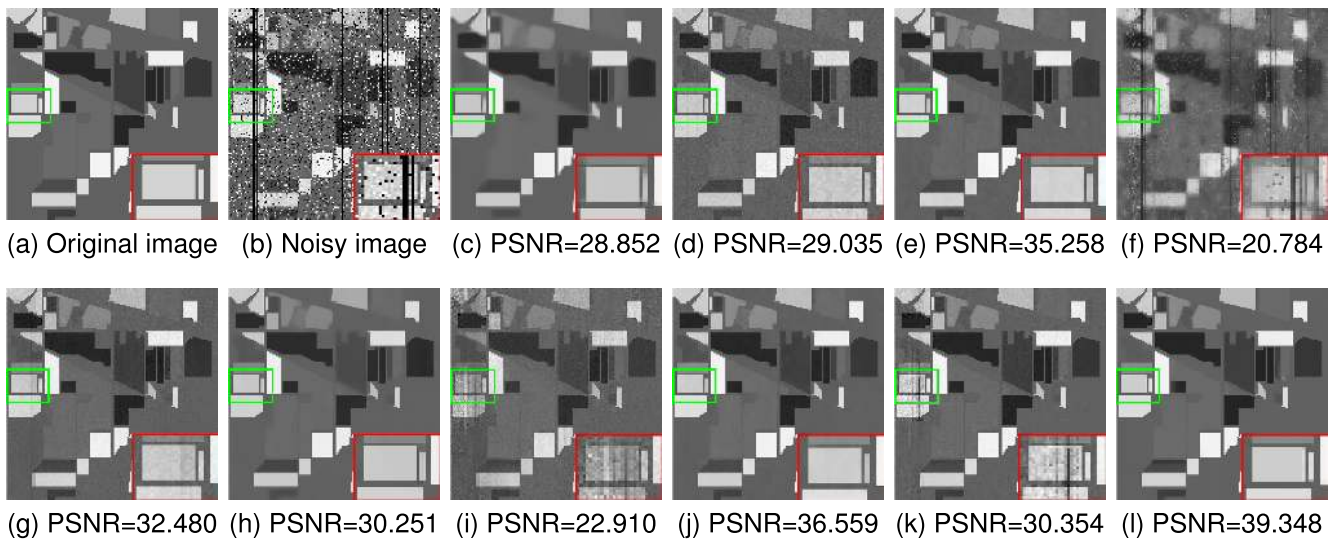
Since an HSI often has dozens or even hundreds of bands, only a part of the representative bands is displayed in the experimental analysis. Firstly, we show different denoised bands in Washington DC Mall dataset and USGS Indian Pines dataset in Fig. 7, Fig. 8, Fig. 11 and Fig. 12. Then, in order to better compare the denoising results of different models visually, the same area with an obvious contrast of the selected band is marked with a green box, and then enlarged in a red box.

As shown in the figures, due to the mixed noise, it is obvious that the original clear HSIs are greatly polluted in image recognition and overall quality. After denoising by





**FIGURE 7.** Band 3 of the USGS Indian Pines dataset before and after denoising via the different methods under noise case 3. (a) Original image of band 6, (b) noisy image. Image denoising results of (c) BM3D, (d) LRTA, (e) LRTV, (f) NLM3D, (g) LRMR, (h) BM4D, (i) NAILRMA, (j) LRTDTV, (k) LLRSSTV, (l) NLRnP.



**FIGURE 8.** Band 116 of the USGS Indian Pines dataset before and after denoising via the different methods under noise case 6. (a) Original image of band 6, (b) noisy image. Image denoising results of (c) BM3D, (d) LRTA, (e) LRTV, (f) NLM3D, (g) LRMR, (h) BM4D, (i) NAILRMA, (j) LRTDTV, (k) LLRSSTV, (l) NLRnP.

various methods, most of the noise in HSIs is removed, but the denoising effect of each method is obviously different. Specifically, NAILRMA is more suitable for low intensity Gaussian noise removal. BM3D and BM4D have the similar denoised images in vision and can achieve better Gauss noise suppression effect, which is consistent with the quantitative evaluation index. However, they can not remove strong impulse noise. In some denoised bands, there is still a phenomenon of image smoothing and the loss of texture information is obvious. NLM3D is not good at depressing mixed noise. Its denoised HSIs are visually distorted, still include some sparse noise and obvious blurring phenomenon at the edge zone. Although LRMR can remove the Gaussian and impulse noise, residual deadlines and strip noise remain in the denoised HSIs. LLRSSTV achieves a good denoising

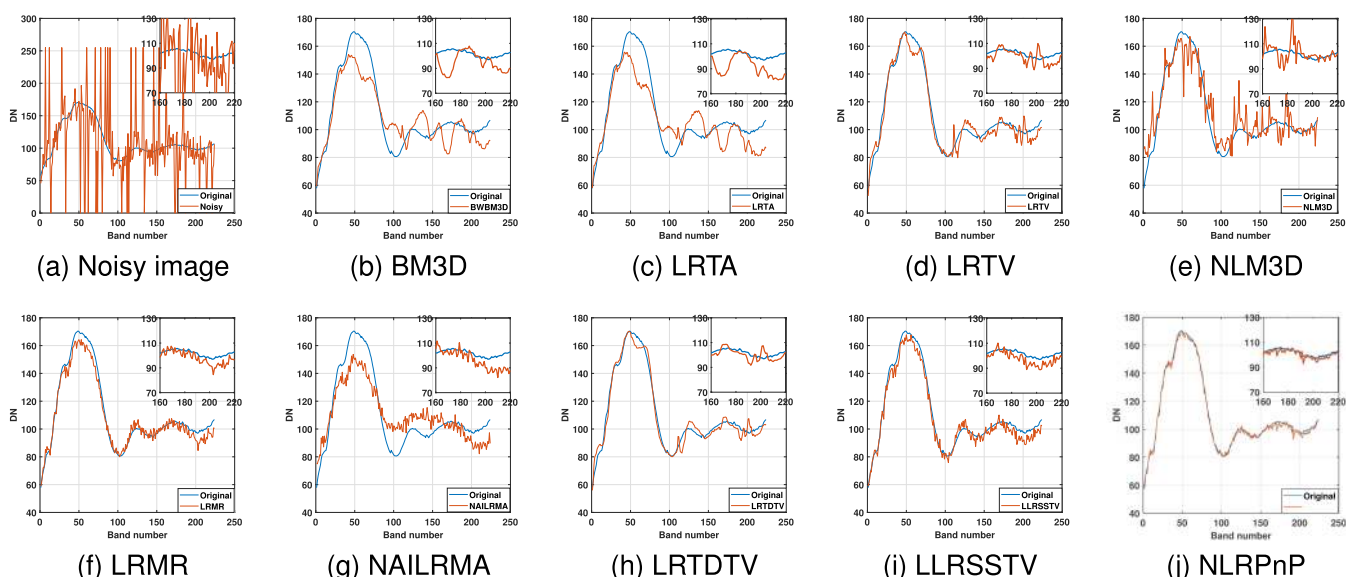
effect. However, since the spatial non-local self-similarity of HSIs is not explored, there are still some deadlines or stripes residues in the denoised HSIs. LRTDTV also achieves a good denoising effect. But, compared with our non-local regularized local low-rank PnP method, the denoised results in the region with rich texture information are not good enough. In summary, compared with the selected denoised methods, the proposed method can obtain the best denoising results in visual comparison.

## 2) QUANTITATIVE COMPARISON

In order to further compare the effectiveness of the proposed algorithm, the peak signal-to-noise ratio (PSNR) [40], structural similarity index (SSIM) [41], erreur relative globale adimensionnelle de synthèse (ERGAS) [42], the mean

**TABLE 1. Quantitative assessment of the denoising results in the simulated experiment of USGS Indian Pines dataset. Boldface means the best and underline means the second best.**

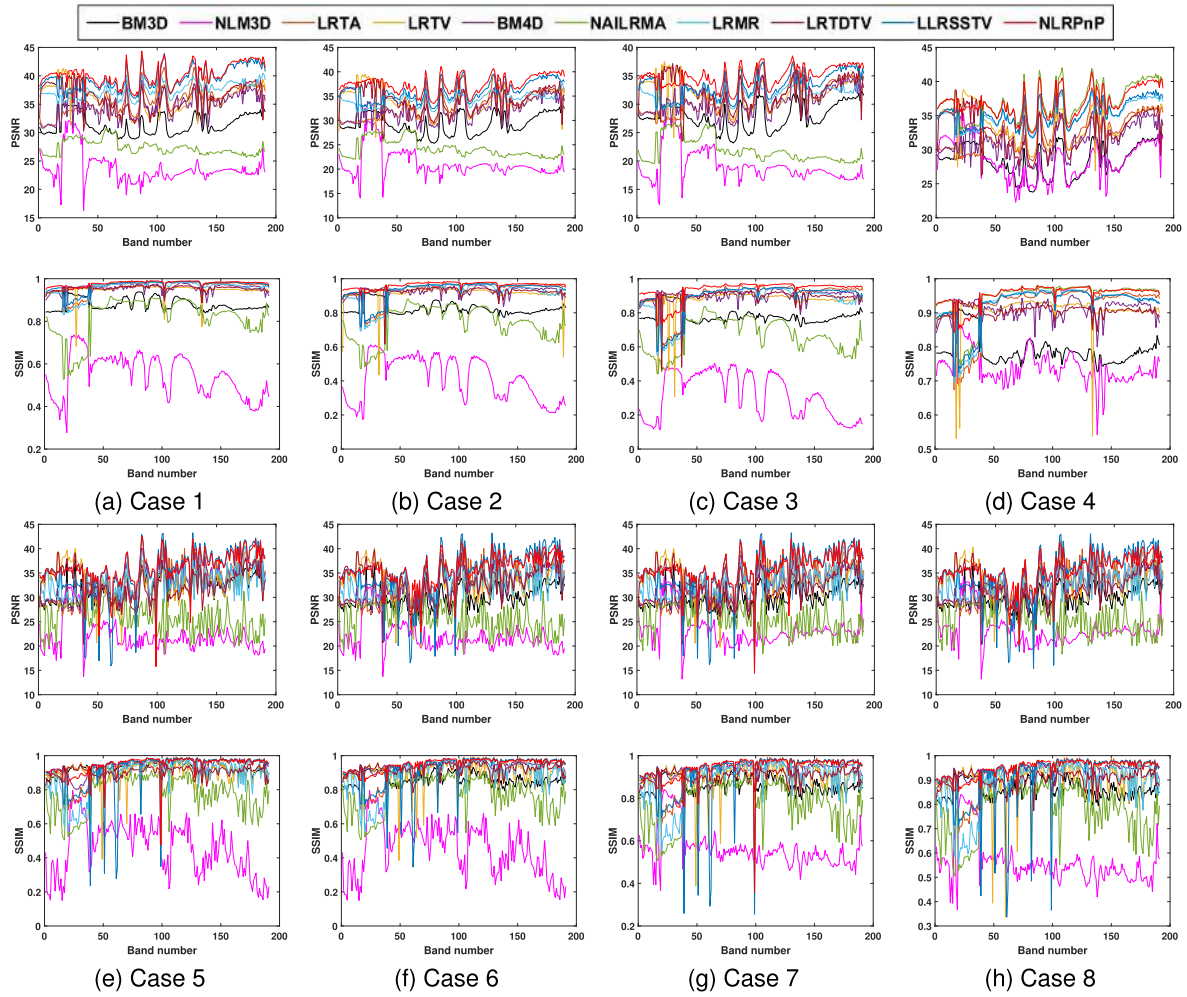
Noise Case	Level	Evaluation index	BM3D	NLM3D	LRTA	LRTV	BM4D	NAI-LRMA	LRMR	LR-TDTV	LLR-SSTV	NLR-PnP
Case 1	G=0.05 P=0.1	MPSNR	34.084	25.184	34.319	42.522	35.793	29.402	39.800	<u>42.799</u>	41.631	<b>45.977</b>
		MSSIM	0.978	0.768	0.937	0.992	0.989	0.860	0.966	<u>0.996</u>	0.972	<b>0.998</b>
		MFSIM	0.978	0.777	0.945	0.990	0.992	0.866	0.966	<u>0.995</u>	0.972	<b>0.996</b>
		ERGAS	47.561	131.426	47.031	19.032	40.226	80.189	23.856	<u>18.944</u>	19.456	<b>11.815</b>
		MSAD	1.528	5.001	1.786	0.706	1.355	0.988	<u>0.685</u>	0.781	<b>0.391</b>	
Case 2	G=0.075 P=0.15	MPSNR	30.889	22.593	31.283	39.833	32.764	26.504	36.396	<u>40.339</u>	37.598	<b>43.287</b>
		MSSIM	0.962	0.640	0.887	0.983	0.981	0.808	0.934	<u>0.992</u>	0.935	<b>0.996</b>
		MFSIM	0.960	0.690	0.902	0.977	0.985	0.826	0.935	<u>0.989</u>	0.938	<b>0.993</b>
		ERGAS	68.332	177.256	66.114	25.287	56.366	112.551	35.360	<u>24.391</u>	30.953	<b>16.217</b>
		MSAD	2.223	6.974	2.520	0.976	1.940	4.285	1.443	<u>0.888</u>	1.242	<b>0.585</b>
Case 3	G=0.1 P=0.2	MPSNR	28.676	20.712	29.031	37.187	30.498	24.313	33.752	<u>37.776</u>	34.522	<b>40.259</b>
		MSSIM	0.945	0.520	0.833	0.968	0.972	0.769	0.892	<u>0.984</u>	0.887	<b>0.993</b>
		MFSIM	0.942	0.612	0.857	0.957	0.977	0.798	0.898	<u>0.979</u>	0.894	<b>0.986</b>
		ERGAS	88.071	220.815	85.230	33.852	72.744	145.637	48.006	<u>32.048</u>	43.960	<b>23.130</b>
		MSAD	2.897	8.893	3.224	1.313	2.517	5.545	1.933	<u>1.161</u>	1.767	<b>0.869</b>
Case 4	Gaussian	MPSNR	29.733	31.782	30.658	39.533	32.134	37.451	36.474	<u>40.300</u>	36.654	<b>41.051</b>
		MSSIM	0.951	0.935	0.870	0.980	0.979	0.939	0.931	<u>0.991</u>	0.925	<b>0.994</b>
		MFSIM	0.946	0.916	0.888	0.974	0.983	0.938	0.929	<u>0.988</u>	0.925	<b>0.988</b>
		ERGAS	77.665	62.489	70.763	26.098	60.241	31.419	35.125	<u>24.729</u>	34.426	<b>21.354</b>
		MSAD	2.487	2.068	2.706	1.040	2.090	1.221	1.440	<u>0.932</u>	1.378	<b>0.728</b>
Case 5	Gaussian + impulse	MPSNR	28.779	23.713	29.499	37.028	30.673	28.170	33.977	<u>38.569</u>	35.652	<b>39.375</b>
		MSSIM	0.946	0.694	0.863	0.973	0.975	0.841	0.893	<u>0.987</u>	0.904	<b>0.987</b>
		MFSIM	0.944	0.718	0.883	0.968	0.980	0.857	0.900	<u>0.983</u>	0.910	<b>0.984</b>
		ERGAS	88.129	162.831	81.718	42.777	71.677	101.763	48.901	<b>29.702</b>	58.268	<u>38.407</u>
		MSAD	2.948	6.355	3.068	1.803	2.476	3.961	2.018	<b>1.068</b>	2.503	<u>1.461</u>
Case 6	Gaussian +impulse +deadline	MPSNR	28.715	23.492	29.415	36.879	30.582	27.824	33.769	<u>38.463</u>	35.281	<b>38.809</b>
		MSSIM	0.946	0.684	0.862	0.970	0.974	0.839	0.892	<b>0.987</b>	0.906	<u>0.986</u>
		MFSIM	0.943	0.711	0.882	0.966	0.980	0.853	0.899	<u>0.983</u>	0.911	<b>0.983</b>
		ERGAS	88.864	166.535	82.559	<u>42.503</u>	72.495	105.784	50.486	<b>30.050</b>	63.738	44.913
		MSAD	2.942	6.537	3.072	1.785	2.465	4.103	2.095	<u>1.091</u>	2.713	<b>1.724</b>
Case 7	Gaussian +impulse +stripe	MPSNR	28.748	24.106	29.439	36.961	30.630	28.139	33.787	<u>38.441</u>	35.426	<b>39.147</b>
		MSSIM	0.946	0.803	0.861	0.970	0.974	0.841	0.891	<u>0.985</u>	0.902	<b>0.987</b>
		MFSIM	0.943	0.794	0.881	0.964	0.980	0.856	0.898	<u>0.982</u>	0.908	<b>0.983</b>
		ERGAS	88.463	152.609	82.226	39.613	72.026	101.968	49.938	<b>29.960</b>	57.622	<u>37.886</u>
		MSAD	2.955	5.401	3.084	1.624	2.480	3.977	2.079	<b>1.083</b>	2.477	<u>1.472</u>
Case 8	Gaussian +impulse +deadline +stripe	MPSNR	28.678	23.938	29.355	36.765	30.531	27.778	33.595	<u>38.311</u>	35.118	<b>38.495</b>
		MSSIM	0.946	0.797	0.861	0.971	0.974	0.837	0.890	<u>0.984</u>	0.904	<b>0.987</b>
		MFSIM	0.943	0.789	0.880	0.966	0.980	0.851	0.897	<u>0.980</u>	0.909	<b>0.982</b>
		ERGAS	89.208	155.480	83.072	<u>42.531</u>	72.883	106.225	51.567	<b>30.385</b>	66.425	45.976
		MSAD	2.958	5.526	3.096	<u>1.772</u>	2.482	4.133	2.157	<b>1.109</b>	2.836	1.792



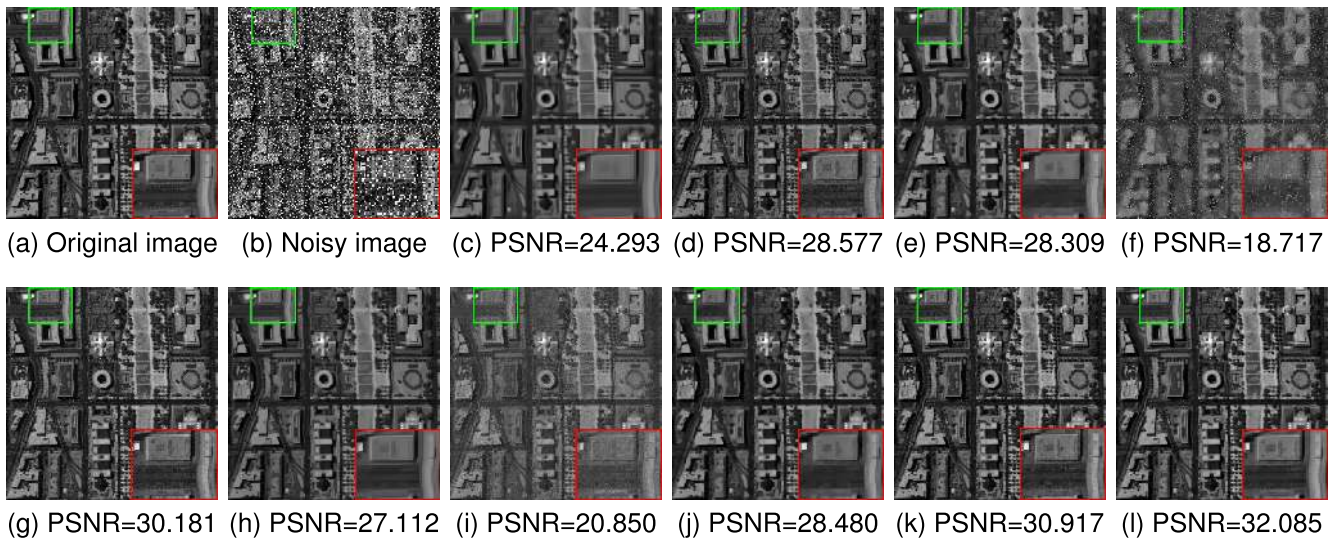
**FIGURE 9. The reflectance of special pixel of USGS Indian Pines in (40,10) under noise case 2.**

spectral angle distance (MSAD) and feature similarity (FSIM) [43] index are adopted to give a quantitative assessment. First, the indices in each band of the denoised HSI

are calculated, and then the indices of all bands are averaged to establish the final numerical evaluation criteria, MPSNR, MSSIM, MSAD, MERGAS and MFSIM. Larger MPSNR,



**FIGURE 10.** Each column shows the PSNR, SSIM values of each band of all methods under the 8 noise cases in Washington DC Mall dataset.



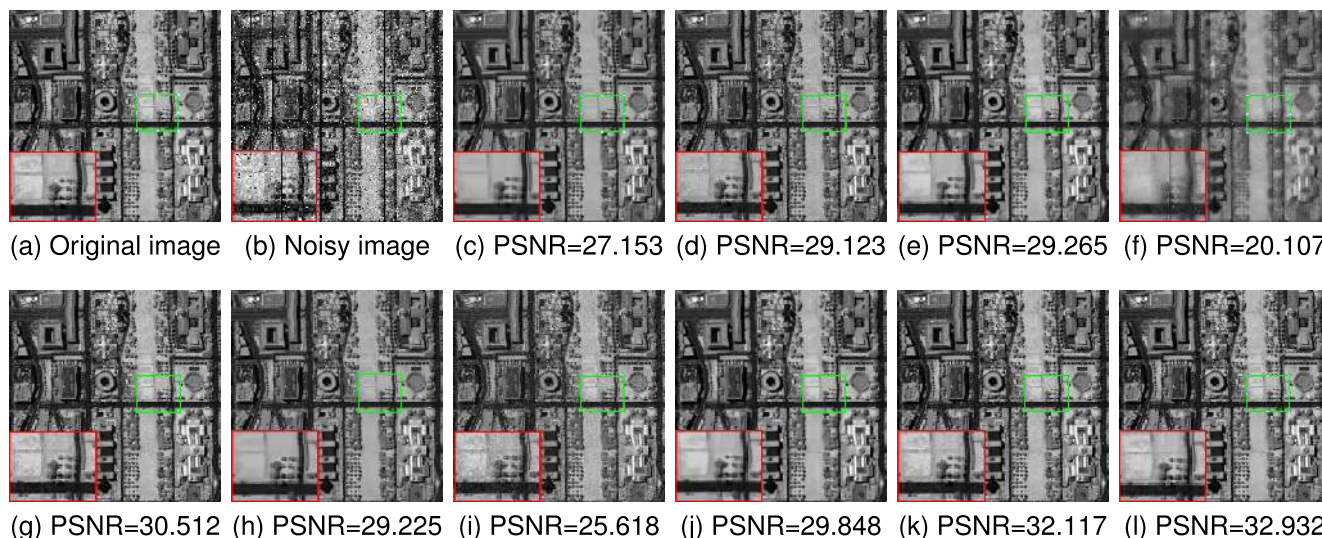
**FIGURE 11.** Denoising result of Bnad 112 of Washington DC Mall under noise case 3. (a) Original image of band 6, (b) noisy image. Image denoising results of (c) BM3D, (d) LRTA, (e) LRTV, (f) NLM3D, (g) LRMR, (h) BM4D, (i) NAILRMA, (j) LRTDTV, (k) LLRSSTV, (l) NLRPnP.

MSSIM and MFSIM, and smaller MERGAS, MSAD indicate that the better the denoising performance of the corresponding model is.

Under the conditions of 8 noise cases and 10 denoising methods, Table 2 shows the values of indices for quantitative assessment in Washington DC Mall dataset, and Table 1

**TABLE 2.** Quantitative assessment of the denoising results in the simulated experiment of Washington DC Mall dataset. Boldface means the best and underline means the second best.

Noise Case	Level	Evaluation index	BM3D	NLM3D	LRTA	LRTV	BM4D	NAI-LRMA	LRMR	LR-TDTV	LLR-SSTV	NLR-PnP
Case 1	G=0.05 P=0.1	MPSNR	31.413	23.712	35.788	36.256	34.737	27.346	37.490	36.193	<u>39.475</u>	<b>40.240</b>
		MSSIM	0.879	0.547	0.962	0.949	0.955	0.819	0.955	0.955	<u>0.968</u>	<b>0.978</b>
		MFSIM	0.922	0.787	0.978	0.966	0.971	0.928	0.977	0.973	<u>0.984</u>	<b>0.987</b>
		ERGAS	111.382	299.442	70.182	63.012	78.872	191.150	53.102	62.601	41.148	<b>38.432</b>
		MSAD	3.767	12.699	2.451	2.406	2.781	8.015	2.563	2.695	<u>2.011</u>	<b>1.778</b>
Case 2	G=0.075 P=0.15	MPSNR	29.299	21.115	33.236	33.970	32.222	24.212	34.343	34.123	<u>36.035</u>	<b>37.512</b>
		MSSIM	0.825	0.417	0.932	0.914	0.927	0.739	0.917	0.928	<u>0.934</u>	<b>0.963</b>
		MFSIM	0.882	0.738	0.962	0.942	0.952	0.896	0.959	0.955	<u>0.967</u>	<b>0.978</b>
		ERGAS	139.244	426.203	90.453	82.724	101.402	285.845	76.745	79.934	<u>60.359</u>	<b>52.697</b>
		MSAD	4.883	17.346	3.324	3.242	3.726	11.670	3.668	3.506	<u>2.938</u>	<b>2.321</b>
Case 3	G=0.1 P=0.2	MPSNR	27.926	19.128	31.264	32.067	30.435	21.925	31.943	32.431	<u>33.357</u>	<b>34.956</b>
		MSSIM	0.782	0.318	0.897	0.876	0.898	0.665	0.872	<u>0.896</u>	0.892	<b>0.934</b>
		MFSIM	0.849	0.694	0.944	0.917	0.932	0.866	0.937	0.934	<u>0.945</u>	<b>0.963</b>
		ERGAS	161.101	550.276	110.605	100.841	121.782	381.207	100.822	97.598	<u>81.418</u>	<b>69.340</b>
		MSAD	5.783	21.400	4.183	3.992	4.574	15.124	4.797	4.334	<u>3.955</u>	<b>3.010</b>
Case 4	Gaussian	MPSNR	28.488	28.492	32.715	33.734	31.358	<b>36.950</b>	35.146	32.843	35.344	<u>36.868</u>
		MSSIM	0.789	0.740	0.921	0.908	0.908	<u>0.943</u>	0.924	0.905	0.924	<b>0.954</b>
		MFSIM	0.855	0.812	0.957	0.938	0.940	<u>0.971</u>	0.960	0.942	0.960	<b>0.972</b>
		ERGAS	151.617	156.377	93.139	85.166	109.026	<b>54.374</b>	67.259	92.667	65.540	<u>55.726</u>
		MSAD	5.497	6.455	3.640	3.383	4.224	<u>2.626</u>	3.313	3.989	3.225	<b>2.566</b>
Case 5	Gaussian + impulse	MPSNR	31.652	22.038	32.885	33.359	32.834	25.692	33.013	32.671	<u>35.681</u>	<b>36.183</b>
		MSSIM	0.919	0.456	0.931	0.901	<u>0.946</u>	0.776	0.884	0.904	0.918	<b>0.946</b>
		MFSIM	0.949	0.751	0.960	0.935	<u>0.964</u>	0.911	0.943	0.942	0.955	<b>0.967</b>
		ERGAS	109.698	387.000	97.882	96.372	96.578	265.887	94.408	97.950	<u>93.572</u>	<b>76.849</b>
		MSAD	3.917	16.027	<u>3.644</u>	4.290	<b>3.517</b>	10.914	4.608	4.428	6.021	3.759
Case 6	Gaussian + impulse + deadline	MPSNR	29.723	21.998	32.787	33.245	32.106	25.647	32.874	33.693	<u>35.407</u>	<b>35.856</b>
		MSSIM	0.860	0.453	0.931	0.901	<u>0.936</u>	0.774	0.884	0.921	0.921	<b>0.950</b>
		MFSIM	0.909	0.750	0.960	0.934	<u>0.957</u>	0.909	0.943	0.951	0.957	<b>0.969</b>
		ERGAS	132.040	387.352	98.848	96.969	103.078	265.972	95.402	<u>88.388</u>	89.029	<b>67.595</b>
		MSAD	4.709	16.168	3.654	4.377	<u>3.708</u>	11.006	4.711	4.150	5.524	<b>3.127</b>
Case 7	Gaussian + impulse + stripe	MPSNR	29.732	23.215	32.786	33.237	32.120	25.644	32.808	32.627	<u>35.596</u>	<b>36.074</b>
		MSSIM	0.860	0.563	0.930	0.899	<u>0.936</u>	0.775	0.883	0.904	0.918	<b>0.951</b>
		MFSIM	0.908	0.756	0.960	0.934	0.957	0.910	0.943	0.941	0.955	<b>0.970</b>
		ERGAS	131.907	322.774	98.618	97.794	102.834	266.217	95.793	98.625	<u>94.856</u>	<b>68.253</b>
		MSAD	4.783	13.613	<u>3.720</u>	4.410	3.788	10.951	4.723	4.492	6.067	<b>3.076</b>
Case 8	Gaussian + impulse + deadline + stripe	MPSNR	29.676	23.194	32.685	33.155	32.030	25.606	32.670	32.536	<u>35.063</u>	<b>35.663</b>
		MSSIM	0.859	0.561	0.930	0.898	<u>0.936</u>	0.773	0.882	0.902	0.916	<b>0.949</b>
		MFSIM	0.908	0.755	0.959	0.931	<u>0.957</u>	0.908	0.942	0.940	0.953	<b>0.968</b>
		ERGAS	132.727	322.763	99.743	99.675	103.834	266.496	<u>96.967</u>	99.981	98.295	<b>69.804</b>
		MSAD	4.776	13.693	<u>3.735</u>	4.636	3.783	11.052	4.842	4.629	6.243	<b>3.263</b>



**FIGURE 12.** Denoising result of Bnad 80 of Washington DC Mall under the noise case 6. (a) Original image of band 6, (b) PSNR. Image denoising results of (c) BM3D, (d) LRTA, (e) LRTV, (f) NLM3D, (g) LRMR, (h) BM4D, (i) NAILRMA, (j) LRTDTV, (k) LLRSSTV, (l) NLRPnP.

shows the ones in USGS Indian Pines dataset. The best values of each index are labeled in bold. One can obviously see that our model significantly outperforms other comparison

methods with respect to almost the indices. Taking the MPSNR as an example, our model achieves nearly 2.5 dB improvement than the second-best results in USGS Indian

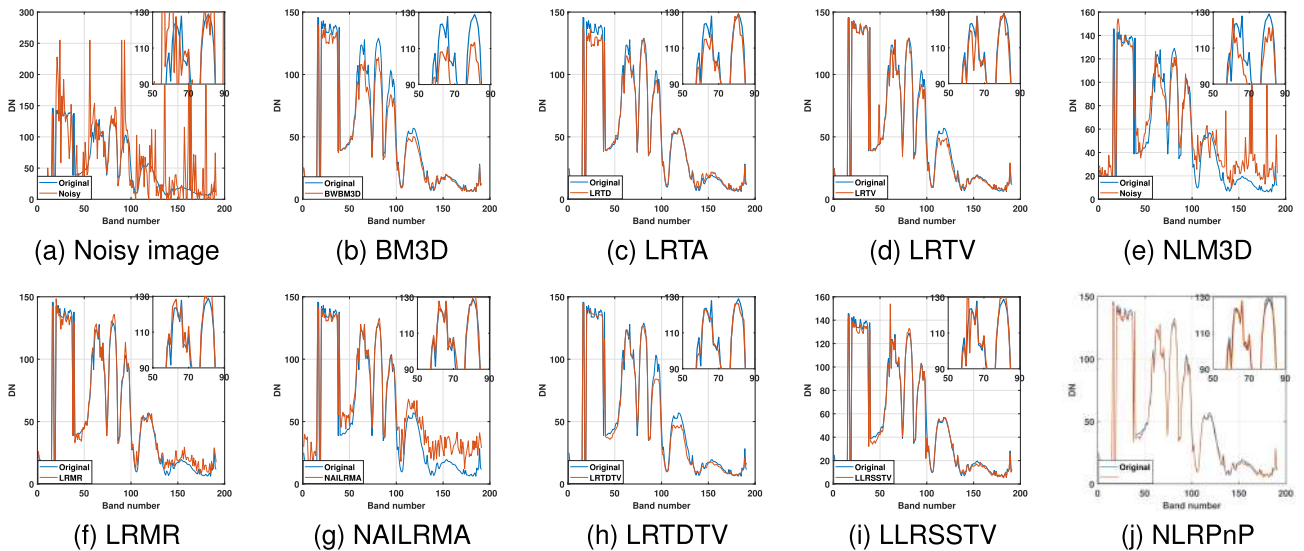


FIGURE 13. The reflectance of special pixel of Washington DC Mall in (220,190) under noise case 6.

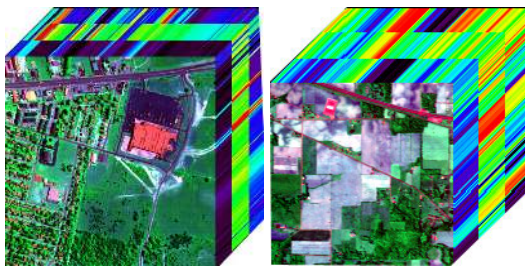


FIGURE 14. Dataset used in the real data experiment (a) HYDICE urban dataset used in the simulated experiment (R: 20, G: 90, B: 180). (b) AVIRIS Indian Pines dataset (R: 1, G: 103, B: 220).

Pines dataset, and 1.5 dB improvement in Washington DC Mall dataset. Furthermore, Fig. 10 lists the PSNR, SSIM index for different denoised bands in Washington DC Mall

dataset, and Fig. 5 lists the ones in USGS Indian Pines dataset. It can be seen that our model has higher PSNR, SSIM values than other methods in most denoised bands.

In noise case 6, Fig. 13 shows the spectral curves of all denoising methods at pixel (220,190) in Washington DC Mall dataset. In noise Case 2, Fig. 9 shows the spectral curves of all denoising methods at pixel (40,10) in USGS Indian Pines dataset. It is easy to see that spectral curves of noisy HSIs fluctuate violently. After denoising by various methods, the fluctuation amplitude of spectral curves is depressed. Compared with all the contrast methods, the spectral curves in our denoised HSIs have less spectral distortions. It means that our model achieves better denoised results in the removal of mixed noise.

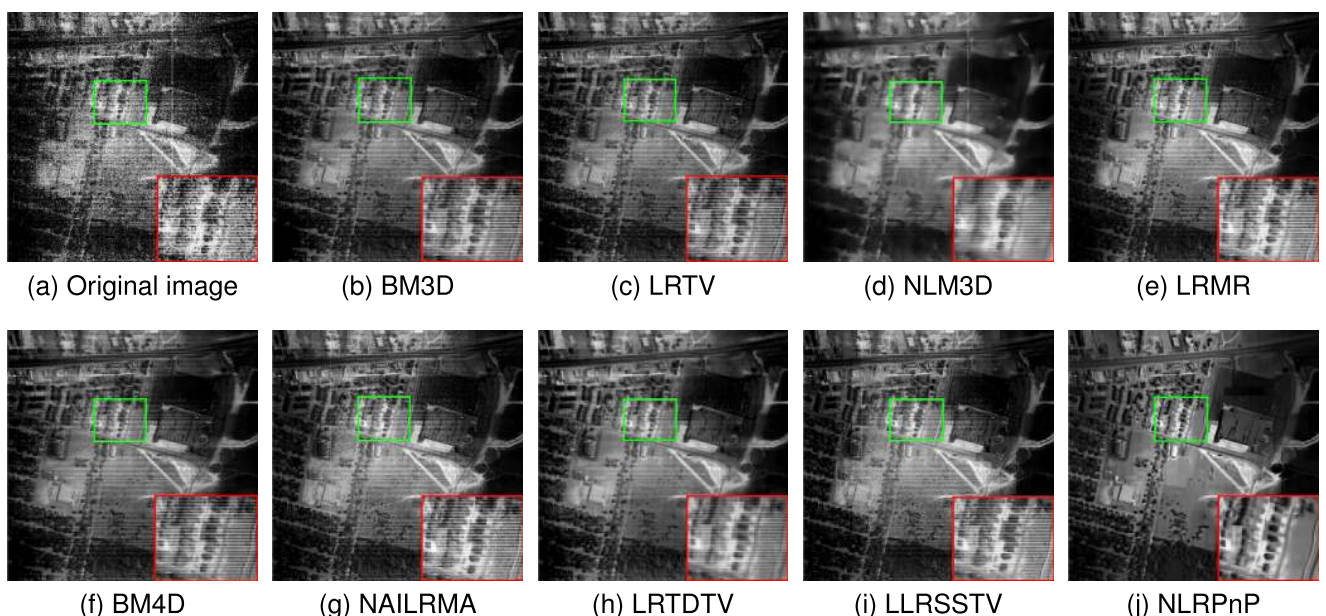
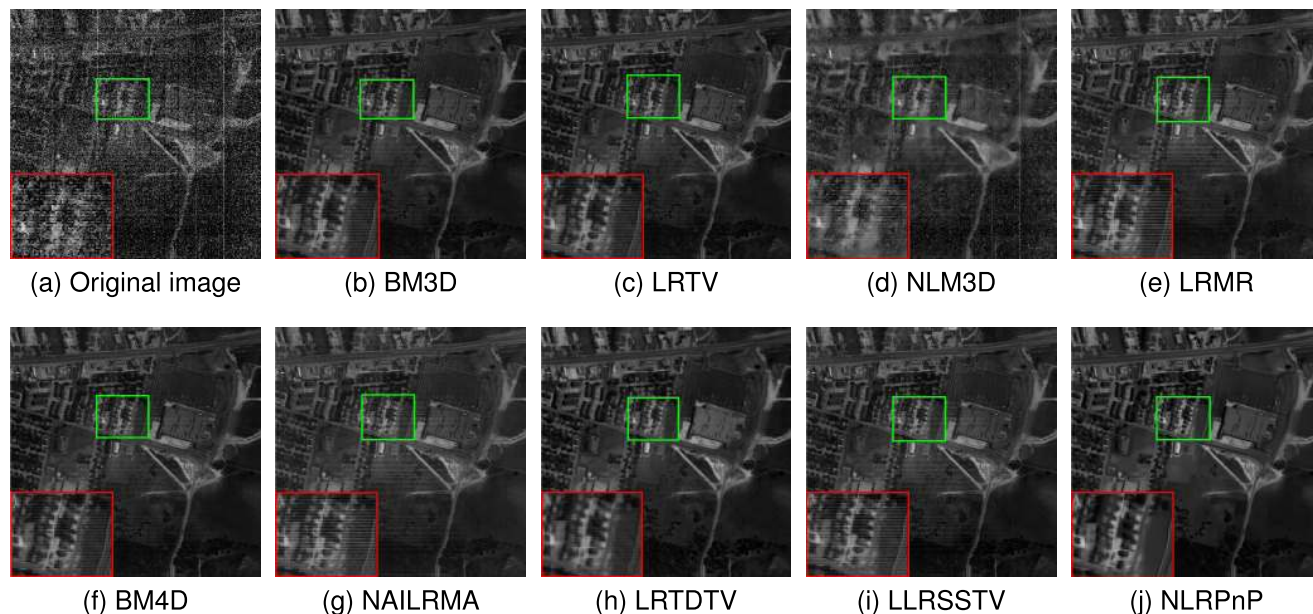
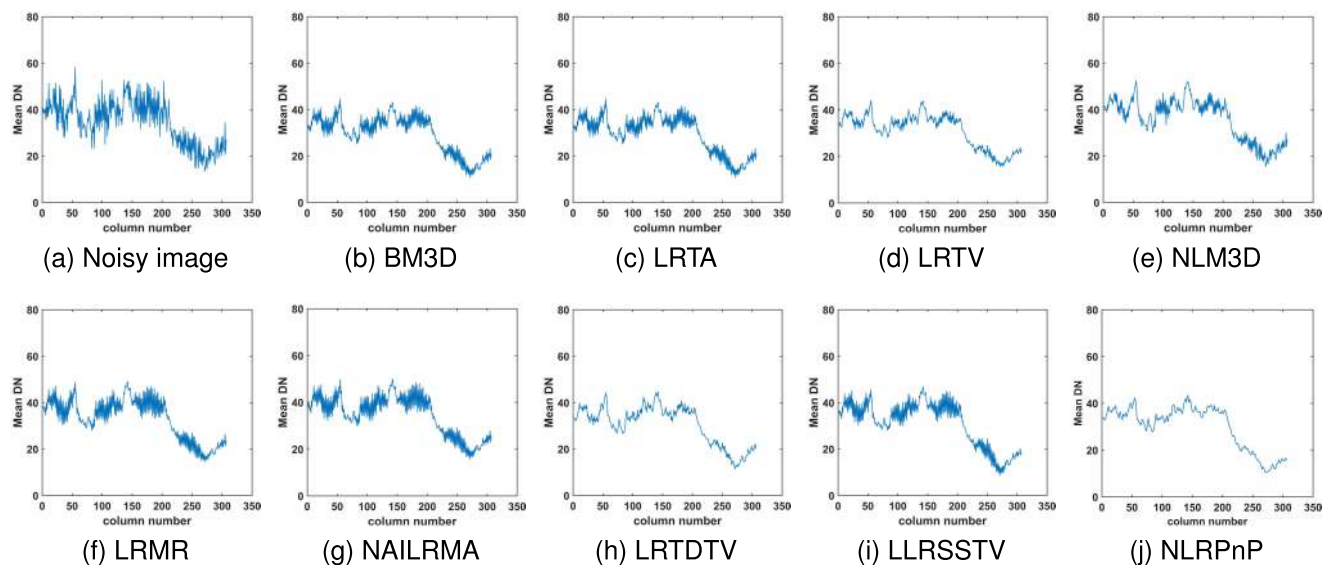


FIGURE 15. Band 139 of the Urban dataset before and after denoising via the different methods. (a) Original image of band 150. Image denoising results of (b) BM3D, (c) LRTV, (d) NLM3D, (e) LRMR, (f) BM4D, (g) NAILRMA, (h) LRTDTV, (i) LLRSSTV, (j) NLRPNP.



**FIGURE 16.** Band 207 of the Urban dataset before and after denoising via the different methods. (a) Original image of band 150. Image denoising results of (b) BM3D, (c) LRTV, (d) NLM3D, (e) LRMR, (f) BM4D, (g) NAILRMA, (h) LRTDTV, (i) LLRSSTV, (j) NLRPnP.



**FIGURE 17.** Spectral signatures curve of band 207 estimated by different methods.

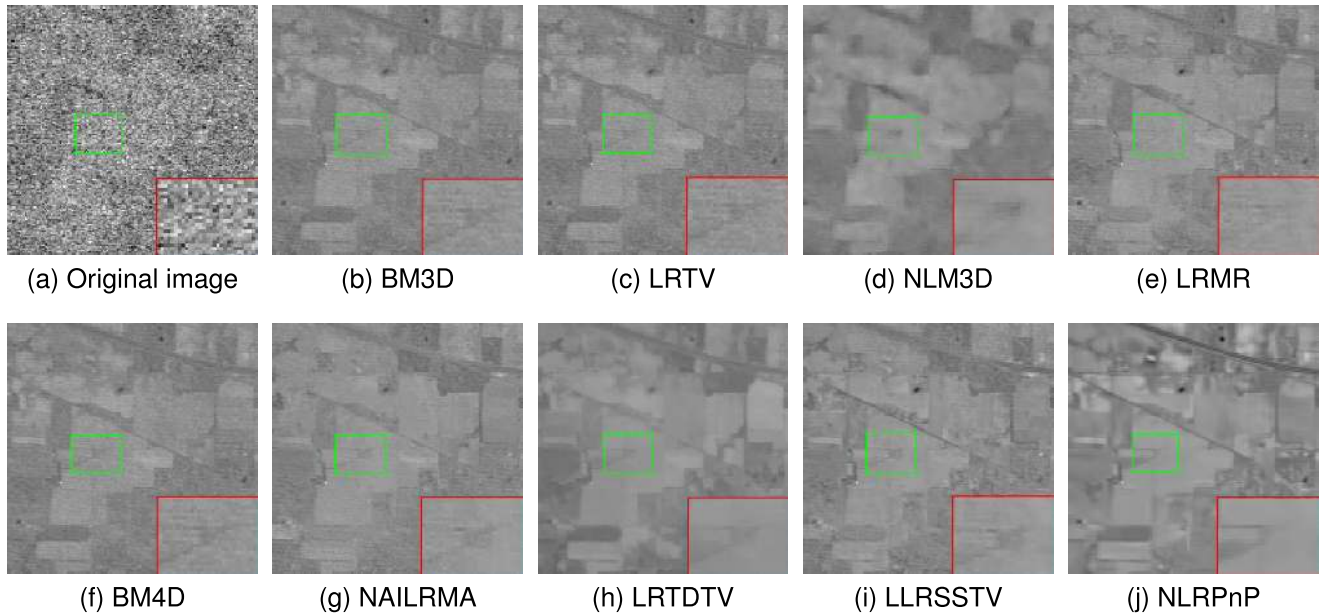
**B. REAL DATA EXPERIMENTS**

In this section, two real datasets are selected to design and perform experiments, i.e., the HYDICE Urban dataset [44] and the AVIRIS Indian Pines dataset [39], which are shown in Fig. 14.

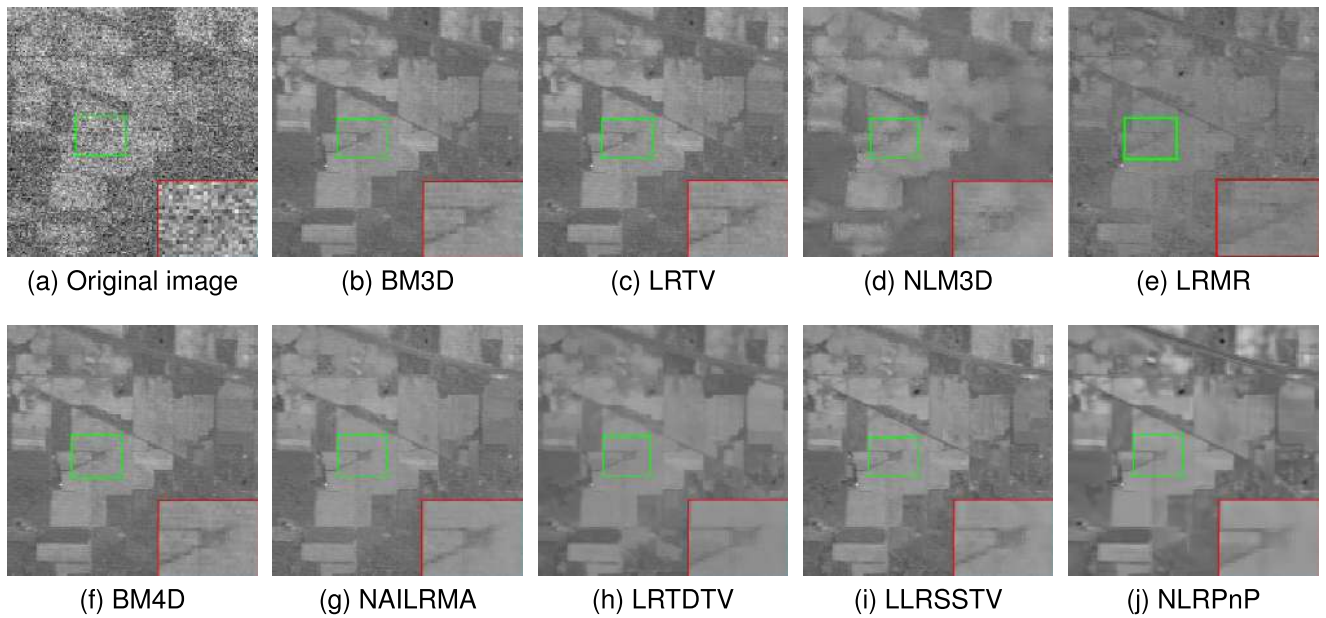
1) HYDICE URBAN DATASET

Fig. 15 and Fig. 16 show the 139th and 207th bands before denoising and after denoising by various methods, respectively. It is easy to see that LRMR, NAILRMA and NLM3D do not effectively remove the stripes. In addition, there is no clear distinction in the denoised HSIs by the three methods.

This is mainly because the stripes and deadlines exist at the same position in band 104 to band 110 and band 199 to band 210. That is, when the model performing low-rank and sparse decomposition, stripes are more likely to be considered as low-rank content and are mistaken for being part of a clean image. The BM3D and LLRSSTV remove some of the noise partly. Although LRTDTV and LRTV show better denoising performance, when they smooth the noise, they also remove the details and texture information at the same time. Our method can simultaneously remove complex mixed noise and preserve spatial texture information compared to other contrast methods.



**FIGURE 18.** Band 108 of the Indian Pines dataset before and after denoising via the different methods. (a) Original image of band 108. Image denoising results of (b) BM3D, (c) LRTV, (d) NLM3D, (e) LRMR, (f) BM4D, (g) NAILRMA, (h) LRTDTV, (i) LLRSSTV, (j) NLRPNP.



**FIGURE 19.** Band 219 of the Indian Pines dataset before and after denoising via the different methods. (a) Original image of band 219. Image denoising results of (b) BM3D, (c) LRTV, (d) NLM3D, (e) LRMR, (f) BM4D, (g) NAILRMA, (h) LRTDTV, (i) LLRSSTV, (j) NLRPNP.

In addition to the above qualitative visual evaluation, the denoising HSIs are further evaluated by the quantitative mean profile. The smaller the fluctuation of the mean profile is, the higher the image quality is. The horizontal average profile of the band 207 before and after denoising is shown in Fig. 17. As shown in Fig. 17(a), due to the existence of mixed noise such as Gaussian noise, stripes and deadlines, the mean profile curves of the noisy image appear to fluctuate rapidly. After denoising, we can see that the mean profile curve of our method is the most stable and its fluctuation is

the smallest. This fact is also consistent with the visual result shown in Fig. 16.

## 2) AVIRIS INDIAN PINES DATASET

This dataset is acquired by the NASA airborne visible/infrared imaging spectrometer (AVIRIS) instrument over the Indian Pines in Northwestern Indiana in 1992, and it has  $145 \times 145$  pixels and 220 bands. Fig. 14 shows the panorama of Indian Pines. It can be seen that it is mainly polluted by

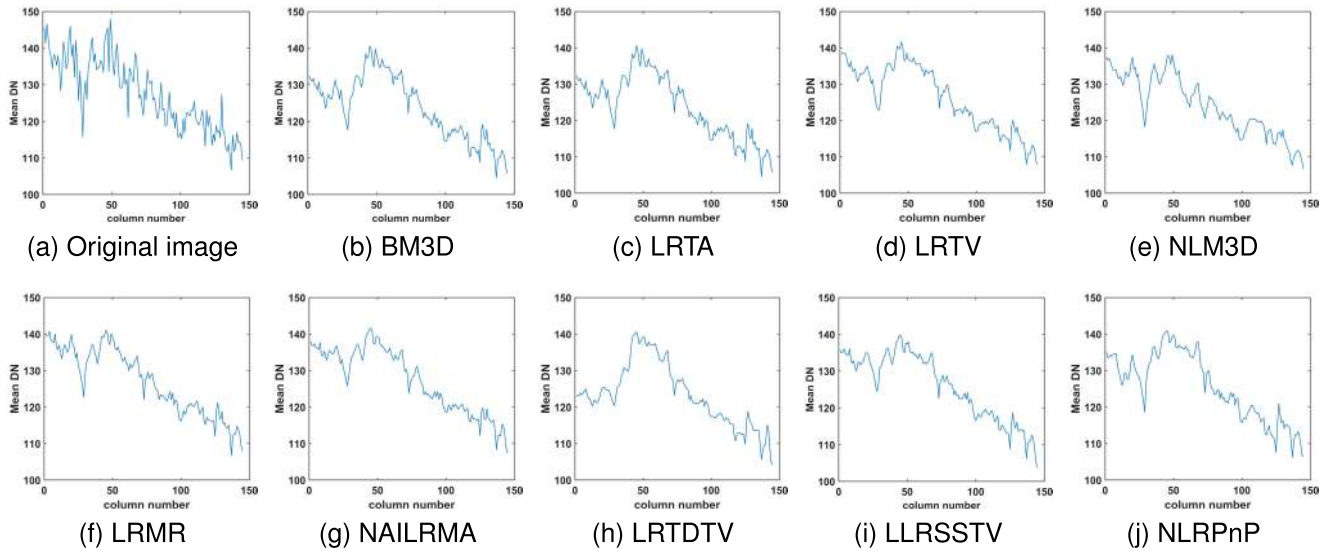


FIGURE 20. Spectral signatures curve of band 219 with Indian pines estimated by different methods.

the atmosphere, water absorption, stripes and other unknown noise.

Fig. 18 and Fig. 19 show the denoised images of the band 108 and 219 by different methods, respectively. Fig. 18(a) and Fig. 19(a) show that the image is completely covered by noise and the details are completely invisible. After denoising, it is easy to observe that LRMR, NAILRMR, and BM3D lose some texture details and distort the denoised images when the intensity of the noise is large. Although LRTDTV and LRSSTV can effectively remove noise, the details of the red boxes in the denoised images are severely degraded. On the contrary, our model can maintain more complete texture information when depressing the mixed noise. Similar to the experimental analysis of the HYDICE Urban dataset, the vertical mean profile of the band 108 before and after denoising by various models is shown in Fig. 20. It can be clearly observed again that our model has the smoothest mean profile curve. This is also consistent with the visual results of Fig. 18 and Fig. 19.

### C. DISCUSSION

#### 1) SENSITIVITY ANALYSIS OF PARAMETER $\lambda$

From our denoising model (9), it is easy to see that  $\lambda$  is an important parameter to balance the influence of sparse noise term and the rest of the regular terms. In RPCA model [12], the sparsity regularization parameter is set to  $\lambda = 1/\sqrt{mn}$ . In our model, there is the new non-local penalty, and the low-rank model is performed on patches, therefore, it is different from RPCA. Hence, we set  $\lambda$  to  $C/\sqrt{mn}$ , where  $C$  is a adjusted parameters. In Washington DC Mall dataset, Fig. 23 shows MPSNR and MSSIM values of our model when  $C$  changes in the set  $\{1, 10, 20, 30, 40, 50, 60, 65, 70, 75, 77, 79, 81, 85, 90, 95, 100\}$ . It can be easily seen from the figure that the results of our model are relatively stable in

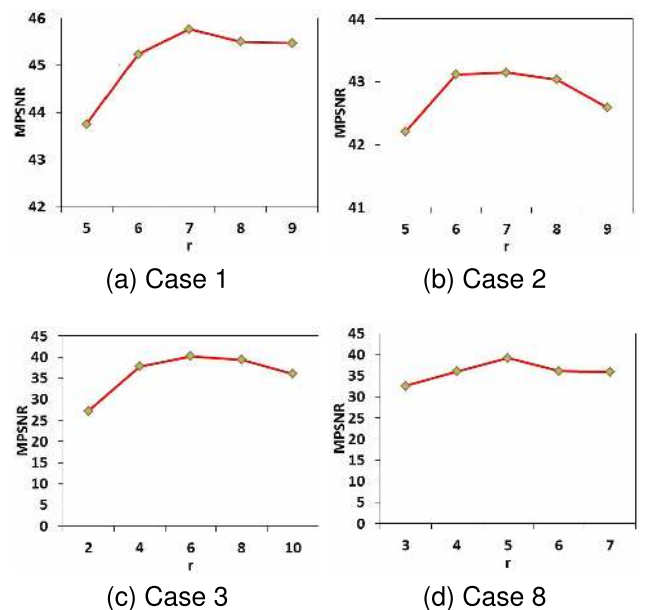


FIGURE 21. Sensitivity analysis of the rank constraint on the Indian Pines image.

terms of MPSNR and MSSIM values. Based on the above discussion, we recommend setting the  $\lambda$  to 0.3 in all simulated data experiments.

#### 2) SENSITIVITY ANALYSIS OF PARAMETER $\beta$

In our model (9),  $\beta$  also is an important parameter used to balance the influence of non-local term and the rest of the regular terms. In Washington DC Mall dataset, Fig. 24 shows MPSNR and MSSIM values of our model when  $\beta$  changes in the set  $\{0.00001, 0.00005, 0.00009, 0.0002, 0.0003, 0.0005, 0.001, 0.002, 0.005, 0.007, 0.02, 0.05, 0.1\}$ . we set  $\beta$  to the



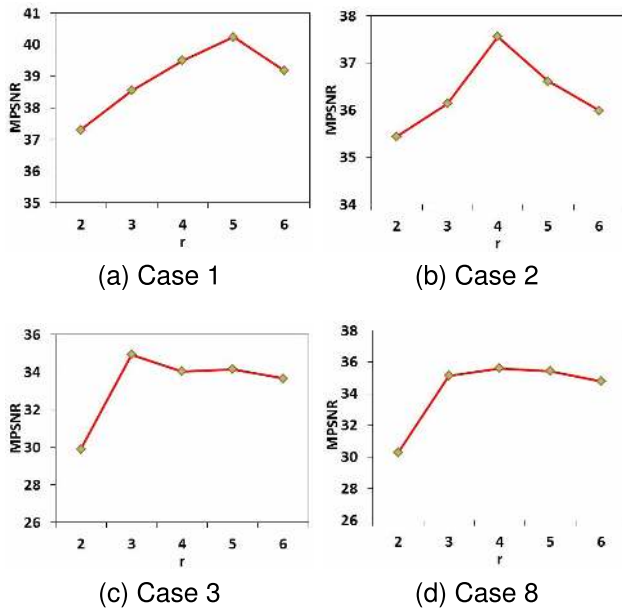


FIGURE 22. Sensitivity analysis of the rank constraint on the Washington DC image.

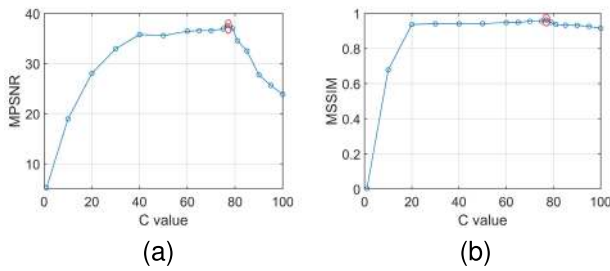


FIGURE 23. Sensitivity analysis of the C value. (a) Change in the MPSNR value, (b) Change in the MSSIM value.

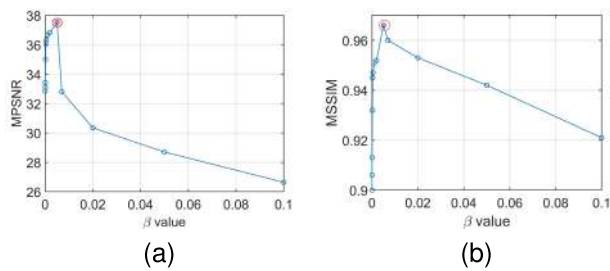


FIGURE 24. Sensitivity analysis of the  $\beta$  value. (a) Change in the MPSNR value, (b) Change in the MSSIM value.

value 0.005 corresponding to the peak values of MPSNR and MSSIM.

### 3) RANK VALUE $r$ SENSITIVITY ANALYSIS

In our model (9), the value of rank  $r$  is related to the intensity of noise and the inherent structure of the image. In order to verify the influence of the value of rank  $r$  on the final results, we analyze the performance of our model in Washington DC Mall dataset and USGS Indian Pines dataset. As shown

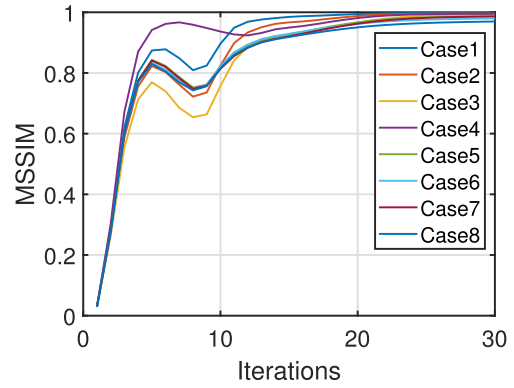


FIGURE 25. MSSIM values versus the iteration number of NLRPNP with the USGS Indian Pines dataset in the simulated experiments.

in Fig. 21 and Fig. 22, as the  $r$  value increases, the MPSNR of our model first increases to a peak value, and then decreases. Therefore, we set  $r$  to the value corresponding to the peak value of MPSNR.

### 4) CONVERGENCE ANALYSIS OF THE MODEL

Fig. 25 shows the MSSIM value of our NLRPNP method according to the iterations. After few iterations, the MPSNR value and the MSSIM value tend to be stable, which reflects the convergence of our algorithm.

### 5) COMPUTATIONAL TIME COMPARISON

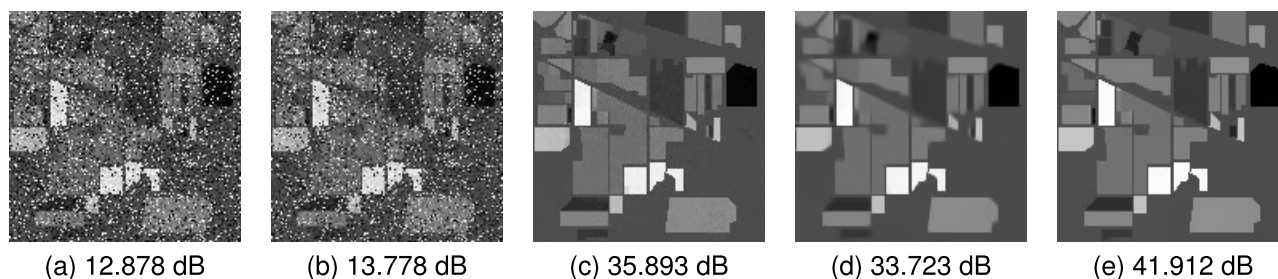
The running time is an effective way to measure the efficiency of a denoising method. For each iteration of the proposed method, the computational burden consists of two parts, i.e., local low-rank and sparse matrix decomposition and non-local regularized image reconstruction. In order to speed up the operation efficiency of the algorithm and enhance the practicability of the proposed model, the algorithm is accelerated by parallel computing of the SVD of all the patches and the BM3D of all the bands, which occupies the most time in each iteration. When the algorithm runs on MATLAB 2018b, the CPU of the computer is Inter core i7@2.20GHz and the memory is 64GB. For the Washington DC Mall dataset with size  $256 \times 256 \times 191$  and the USGS Indian Pines dataset with size  $145 \times 145 \times 224$ , we have averaged the program running time of 8 groups of experiments under the conditions of noise case 1-8. The timing function built into MATLAB is used for timing. The average running time of each model is shown in Table 3.

### 6) NOT A SIMPLE COMBINATION

The NLRPNP is not just simple combination of non-local method, e.g., BM3D, and LLR but is a systematical integration of BM3D with LLR in the Plug-and-Play framework. In Fig. 26, we demonstrate the results on USGS Indian Pines recovered by BM3D and LLR, directly performing BM3D after LLR, and NLRPNP with the noise case 3. The BM3D gets the worst performance, because it is designed remove the

**TABLE 3.** Running times (in seconds) of the different methods in the simulated data experiments.

HSI Data	BM3D	NLM3D	LRTA	LRTV	BM4D	NAILRMA	LRMR	LRTDTV	LLRSTV	NLRPnP
Washington DC Mall	89.4	520.3	30.5	99.2	456.6	58.7	154.9	185.2	318.2	172.6
USGS Indian Pines	33.2	206.0	21.6	36.5	147.4	20.2	52.8	89.5	118.6	60.2

**FIGURE 26.** Band 204 of the Indian Pines dataset before and after denoising via the different methods. (a) Noisy image, image denoising results of (b) BM3D, (c) LLR, (d) LLR+BM3D, (e) NLRPnP.

Gaussian noise. And the result by directly performing BM3D after LLR is slightly worse than that by LLR. This direct combination-processing does not work. Meanwhile, we can see that the PSNR value of the result by NLRPnP is nearly 6 dB higher than that of simple combination of BM3D and LLR.

## VI. CONCLUSION

In this paper, based on the newly emerged PnP framework, we have proposed a novel NLRPnP method for HSI denoising. The HSI is first divided into local overlapping patches. Then, we adopt the patch-based low-rank matrix approximation to guarantee the local low-rankness while plugging in non-local based denoisers to promote the non-local self-similarity. Furthermore, to consider the inner geometry or structure of mixed noises, we integrate the  $L_1$  norm regularization to the denoising framework, to detect the local sparse noise, including stripes, impulse noise, and dead pixels. Simulated and real HSI experiment results confirmed that compared with competitive methods, the proposed model has certain advantages in preserving the abundant details and the indices for quantitative assessment.

## REFERENCES

- [1] A. F. H. Goetz, "Three decades of hyperspectral remote sensing of the Earth: A personal view," *Remote Sens. Environ.*, vol. 113, pp. S5–S16, Sep. 2009.
- [2] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [3] 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.
- [4] Z. Zhang, E. Pasolli, M. M. Crawford, and J. C. Tilton, "An active learning framework for hyperspectral image classification using hierarchical segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 640–654, Feb. 2016.
- [5] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.
- [6] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
- [7] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.
- [8] 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.
- [9] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of MR images with spatially varying noise levels," *J. Magn. Reson. Imag.*, vol. 31, no. 1, pp. 192–203, Jan. 2010.
- [10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [11] 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, 1999.
- [12] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, May 2011.
- [13] 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.
- [14] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.
- [15] Q. Zhao, D. Meng, X. Kong, Q. Xie, W. Cao, Y. Wang, and Z. Xu, "A novel sparsity measure for tensor recovery," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 271–279.
- [16] Y. Chang, L. Yan, H. Fang, and C. Luo, "Anisotropic spectral-spatial total variation model for multispectral remote sensing image destriping," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1852–1866, Jun. 2015.
- [17] H. Fan, J. Li, Q. Yuan, X. Liu, and M. Ng, "Hyperspectral image denoising with bilinear low rank matrix factorization," *Signal Process.*, vol. 163, pp. 132–152, Oct. 2019.
- [18] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.
- [19] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.
- [20] Z. Zhang and S. Aeron, "Exact tensor completion using t-SVD," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1511–1526, Mar. 2017.
- [21] 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.*, to be published.

- [22] 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.
- [23] 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.
- [24] 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.
- [25] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.
- [26] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5344–5353.
- [27] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.
- [28] S. V. Venkatakrisnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 945–948.
- [29] T. Tirmir and R. Giryes, "Image restoration by iterative denoising and backward projections," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1220–1234, Mar. 2019.
- [30] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising prior driven deep neural network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2305–2318, Oct. 2019.
- [31] 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.
- [32] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2005, pp. 60–65.
- [33] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Image restoration and reconstruction using variable splitting and class-adapted image priors," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3518–3522.
- [34] Y. Dar, A. M. Bruckstein, M. Elad, and R. Giryes, "Postprocessing of compressed images via sequential denoising," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3044–3058, Jul. 2016.
- [35] U. S. Kamilov, H. Mansour, and B. Wohlberg, "A plug-and-play priors approach for solving nonlinear imaging inverse problems," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1872–1876, Dec. 2017.
- [36] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
- [37] 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.
- [38] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.
- [39] Accessed: Apr. 4, 2018. [Online]. Available: <https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html/>
- [40] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, 2008.
- [41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [42] L. Wald, "Data fusion: Definitions and architectures: Fusion of images of different spatial resolutions," Presses des MINES, Tech. Rep., 2002.
- [43] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [44] Accessed: Apr. 4, 2018. [Online]. Available: <http://www.tec.army.mil/hypercube/>



**HUIJIN ZENG** was born in Jiangxi, China, in 1996. He received the B.S. degree in mathematics and statistics from the Henan University of Science and Technology, Luoyang, Henan, China, in 2018. He is currently pursuing the master's degree with Northwest A&F University, Yangling, China. His research interests include remote sensing data analysis and statistical signal processing.



**XIAOZHEN XIE** received the B.S. degree in mathematics from the Department of Applied Mathematics, Northwest A&F University, Yangling, China, in 2007, and the M.S. and Ph.D. degrees in applied mathematics from the School of Mathematics and Systems Science, Beihang University, Beijing, China, in 2010 and 2013, respectively. He is currently an Associate Professor with the College of Science, Northwest A&F University. His research interests include image processing and hyperspectral data analysis.



**WENFENG KONG** received the B.S. degree in information and computing science from Northwest A&F University, Yangling, China, in 2019, where she is currently pursuing the master's degree. Her research interest includes hyperspectral image processing.



**SHUANG CUI** is currently pursuing the bachelor's degree in computer and science with Northwest A&F University, Yangling, China. Her research interests include image identification and deep similarity learning.



**JIFENG NING** was born in Shaanxi, China, in 1975. He received the master's degree in engineering from Northwest A&F University, Shaanxi, in 2002, and the Ph.D. degree in information and communication engineering from Xidian University, Shaanxi, in 2009. He is currently a Professor with the College of Information Engineering, Northwest A&F University. His current research interests include visual tracking and remote sensing image analysis by computer vision and machine learning.

...