

Multi-modal and frequency-weighted tensor nuclear norm for hyperspectral image denoising

Sheng Liu, Xiaozhen Xie, Wenfeng Kong, and Jifeng Ning

Abstract—Low-rankness is important in the hyperspectral image (HSI) denoising tasks. The tensor nuclear norm (TNN), defined based on the tensor singular value decomposition, is a state-of-the-art method to describe the low-rankness of HSI. However, TNN ignores some of the physical meanings of HSI in tackling the denoising tasks, leading to suboptimal denoising performance. In this paper, we propose the multi-modal and frequency-weighted tensor nuclear norm (MFWTNN) and the non-convex MFWTNN for HSI denoising tasks. Firstly, we investigate the physical meaning of frequency components and reconsider their weights to improve the low-rank representation ability of TNN. Meanwhile, we also consider the correlation among two spatial dimensions and the spectral dimension of HSI and combine the above improvements to TNN to propose MFWTNN. Secondly, we use non-convex functions to approximate the rank function of the frequency tensor and propose the NonMFWTNN to relax the MFWTNN better. Besides, we adaptively choose bigger weights for slices mainly containing noise information and smaller weights for slices containing profile information. Finally, we develop the efficient alternating direction method of multiplier (ADMM) based algorithm to solve the proposed models, and the effectiveness of our models are substantiated in simulated and real HSI datasets.

Index Terms—hyperspectral image, denoising, non-convex approximation, frequency components, multi-modal.

I. INTRODUCTION

HYPERSPECTRAL image (HSI) has been widely used in many fields due to its wealthy spectral information of real scenes, e.g., target detection [1], classification [2] and so on. Unfortunately, due to imaging conditions and the influence of weather, observed HSIs are often accompanied by the Gaussian noise, salt and pepper noise, and stripe noise. This not only reduces the quality of HSIs, but also hinders subsequent applications. Therefore, to improve the accuracy of subsequent applications, an important research topic is to recover the clear HSI from the observed HSI in the hyperspectral processing tasks [3, 4].

Since the spectral bands are the imaging results of the same scene under different wavelengths, they have global correlation or low-rankness. Based on this prior, HSI can be unfolded into matrices along the spectral mode and then processed by matrix restoration methods. The low-rank matrix recovery-based method restores the clean HSIs by minimizing the

ranks of unfolding matrices, which is the most popular way [3, 5–8]. Zhang et al.[6] propose to disassemble the noisy HSI into the clean HSI and noise information and use the matrix nuclear norm as the relaxation of the rank function. To improve the low-rank characterization performance of the model, some non-convex methods are introduced into the HSI denoising task. Xie et al. [9] use the $\log \text{sum}$ function, Chen et al.[10] use the γ function, Lou et al.[11] use the $L_1 - L_2$ function and Gao et al.[12] use the SCAD and MCP function as the non-convex approximation of the rank function. Low-rank matrix decomposition is also an effective way to solve the low-rank minimization problem [13, 14]. However, the unfolding operator destroys the high-dimensional structure of HSIs. Treating HSI as a third-order tensor can protect this structure, and tensor rank minimization is the most effective way to overcome the above shortcoming. Due to the non-unique definitions of the tensor rank, different tensor decompositions and the corresponding tensor ranks are proposed, e.g., the CANDECOMP/PARAFAC (CP) decomposition [15, 16], Tucker decomposition [17, 18] and tensor singular value decomposition (t-SVD) [19, 20]. CP rank is defined as the minimum of rank-1 tensors by CP decomposition. Whereas, it is difficult to estimate when HSI is unknown. By Tucker decomposition, a tensor can be decomposed into core tensor and modal matrixes. Then, Tucker rank is approximated by the sum of modal matrix nuclear norms (SNN) [18]. However, SNN has shortcomings in preserving low-rank structure, and its results are materially sub-optimal [21, 22].

Based on t-SVD, the tensor tubal rank has attracted extensive attention recently. Its convex relaxation is the tensor nuclear norm (TNN), which is defined as the sum of the matrix nuclear norms of each frontal slice in the Fourier transformed tensor [19, 21, 23]. The minimization of TNN can be quickly solved by convex optimization algorithms, and it is effective in characterizing the low-rankness of tensor. Therefore, it is widely used in HSI denoising task [24, 25]. In the definition process of TNN, the information in the image domain of HSI is transformed into frequency information in the Fourier domain. In the Fourier transformed tensor, the low-frequency slices mainly carry the profile information, while the high-frequency slices mainly carry detail and noise information. Besides, in each frontal slice of the Fourier transformed tensor, bigger singular values mainly contain information of clean data and smaller singular values mainly contain information of noise. Meanwhile, HSIs have correlations among different modes, but TNN lacks the flexibility to explore different correlations along with different modes [26]. Thus, it is advantageous to improve the flexibility of TNN when considering the prior information

This work was supported by the National Natural Science Foundation of China under Grant No. 61876153. (Corresponding author: Xiaozhen Xie.)

Sheng Liu, Xiaozhen Xie, and Wenfeng Kong are with College of Science, Northwest A&F University, Yangling 712100, China (e-mail: liusheng16@163.com; e-mail: xiexzh@nwsuaf.edu.cn; e-mail: kksama@nwsuaf.edu.cn).

Jifeng Ning is with the College of Information Engineering, Northwest A&F University, Yangling 712100, China (njf@nwsuaf.edu.cn).

about the physical meaning of frequency, correlation among different modes, and singular value.

In this paper, to take full advantage of the above prior information and improve the capability and flexibility of model, we propose the multi-modal and frequency-weighted tensor nuclear norm (MFWTNN) and non-convex MFWTNN. In the proposed MFWTNN, we explore the physical meaning of frequency to explore the low-rankness of HSI in the Fourier domain and give an adaptive calculation method of frequency weight. In addition, we consider the correlation among the spatial modes and the spectral mode. Furthermore, based on MFWTNN, we take into account the physical meaning of singular values inside frequency slices and propose the non-convex approximation of their nuclear norms. Finally, we apply them to the HSI denoising task. The main contributions of this paper are summarized as follows:

- We investigate the physical meaning of frequency components in the definition of TNN and reconsider the weight of frequency components to enhance the low-rank representation ability of TNN. Meanwhile, we consider the correlation among spectral and spatial modes of HSI and use their weighted sum to explore the low-rankness of HSI. Furthermore, we combine the advantages of multi modes with the improvement of TNN in the frequency domain and propose MFWTNN in HSI denoising.
- In each frontal slice of the Fourier transformed tensor, we choose the *log sum* function as a non-convex approximation to the rank function, which can provide a better approximation to rank function compared with nuclear norms in MFWTNN.
- According to information types in different frequency slices in the Fourier transformed tensor, we adaptively choose bigger weights for slices mainly containing noise information and smaller weights for slices containing profile information, which can depress noise more and simultaneously preserve the profile information of clean HSIs better.
- We develop the efficient alternating direction method of multiplier based algorithm to solve the proposed models, and obtain the best restorative performance both on the simulated and real HSI dataset in comparison to all competing HSI denoised methods.

II. RELATED WORK

A. Notations

In this section, we follow the work in [27, 28] and give some basic notations used about matrix and tensor. We use bold upper-case letter \mathbf{X} and calligraphic letter \mathcal{X} to denote matrix and tensor, respectively. For a 3rd-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its (i, j, k) -th component is denoted as $\mathcal{X}(i, j, k)$. For $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, their inner product is defined as $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} x_{ijk} y_{ijk}$. Then the Frobenius norm of a tensor \mathcal{X} is computed as $\|\mathcal{X}\|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$. The ℓ_1 -norm of tensor \mathcal{X} is calculated as $\|\mathcal{X}\|_1 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} |x_{ijk}|$. The k -th frontal slice of \mathcal{X} is represented as $\mathbf{X}^{(k)} = \mathcal{X}(:, :, k)$. For three-dimensional tensor \mathcal{X} , its fast Fourier transform is represented as $\bar{\mathcal{X}} = \text{fft}(\mathcal{X}, [], 3)$ and its

inverse operation is $\mathcal{X} = \text{ifft}(\bar{\mathcal{X}}, [], 3)$. The mode- p permutation of \mathcal{X} is defined as $\mathcal{X}_p = \text{permute}(\mathcal{X}, p)$, $p = 1, 2, 3$, where the m -th mode-3 slice of \mathcal{X}_p is the m -th mode- p slice of \mathcal{X} , i.e., $\mathcal{X}(i, j, k) = \mathcal{X}_1(j, k, i) = \mathcal{X}_2(k, i, j) = \mathcal{X}_3(i, j, k)$. Also, its inverse operation is $\mathcal{X} = \text{ipermute}(\mathcal{X}_p, p)$.

B. Problem Formulation

A clean HSI can be treated as a third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and is usually assumed to be low-rank. Corrupted by mixed noise, its observed version can be modeled as

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

where $\mathcal{Y}, \mathcal{S}, \mathcal{N} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$; \mathcal{S} denotes the sparse noise; \mathcal{N} denotes the Gaussian white noise.

HSI denoising aims to recover the clean HSI \mathcal{X} from the observed HSI \mathcal{Y} in (1). Under the framework of regularization theory, it can briefly be formulated as

$$\arg \min_{\mathcal{X}, \mathcal{S}, \mathcal{N}} \text{Rank}(\mathcal{X}) + \lambda \|\mathcal{S}\|_1 + \tau \|\mathcal{N}\|_F^2, \text{ s.t. } \mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N}, \quad (2)$$

where $\|\cdot\|_1$ describes the sparse noise; $\|\cdot\|_F$ describes the Gaussian noise; $\text{Rank}(\cdot)$ represents the rank of unknown ideal HSI; λ and τ are non-negative parameters.

C. Tensor Nuclear Norm

Since the tensor rank function is non-convex and discrete, its rank minimization problem is NP-hard [29]. To overcome computational difficulties, an usual approach is to approximate the non-convex rank function to the convex function and then substitute it into the minimization problem. Fazel [30] proves that the matrix nuclear norm is the convex envelope rank function on the unit sphere of the spectral norm. Thereafter, the nuclear norm is used as a convex replacement of the rank function. The rank minimization problem is transformed into nuclear norm minimization, and it is widely used in low-rank matrix recover based HSI denoising tasks [6].

Furthermore, HSI can be regarded as a third-order tensor. Then, based on t-SVD [19], Lu et al. [21] propose the tensor nuclear norm (TNN) as the relaxation function of the tensor rank function. TNN is defined as

$$\|\mathcal{X}\|_{TNN} := \frac{1}{n_3} \sum_{k=1}^{n_3} \|\bar{\mathbf{X}}^{(k)}\|_*, \quad (3)$$

where $\bar{\mathbf{X}}^{(k)}$ is the k th frontal slice of $\bar{\mathcal{X}}$. Due to the property of *fft* [21, 31], we have

$$\begin{cases} \bar{\mathbf{X}}^{(1)} \in \mathbb{R}^{n_1 \times n_2} \\ \text{conj}(\bar{\mathbf{X}}^{(i)}) = \bar{\mathbf{X}}^{(n_3-i+2)}, i = 2, \dots, \lfloor \frac{n_3+1}{2} \rfloor \end{cases} \quad (4)$$

Based on (4), frequency component is defined as

$$\{\bar{\mathcal{X}}_{i-1}\} = \begin{cases} \bar{\mathbf{X}}^{(1)} & \text{if } i = 1 \\ \bar{\mathbf{X}}^{(i)} + \bar{\mathbf{X}}^{(n_3-i+2)} & \text{if } i = 2, \dots, \lceil \frac{n_3+1}{2} \rceil \end{cases} \quad (5)$$

Fig. 1 shows the relationship between the frontal slice and the frequency component.

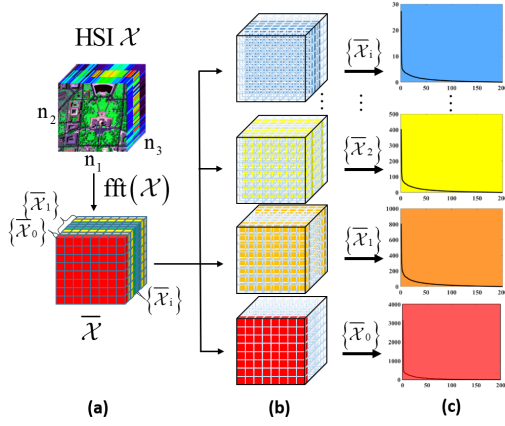


Fig. 1. Illustration of the frequency component. (a) $\bar{\mathcal{X}}$ is the Fourier transform of HSI \mathcal{X} , (b) Frequency component of \mathcal{X} , (c) The singular value curve of the frequency component $\{\bar{\mathcal{X}}_i\}$ of $\bar{\mathcal{X}}$.

D. Multi-modal Tensor Nuclear Norm

Due to TNN can be effective in expressing the low-rank characteristics, it has received extensive attention in the HSI denoising task. However, it focuses on the low-rankness in the spectral dimension. For HSI, its low-rankness not only exists in the spectral dimension, but also exists in the spatial dimensions [28, 32]. To connect the correlation along different dimensions, Liu et al. [18] use modal matrices of the tensor in different directions to explore low-rank structures, and then use SNN as the convex relaxation of the rank function. SNN is defined as

$$\|\mathcal{X}\|_{SNN} := \sum_{p=1}^3 \alpha_p \|\mathbf{X}_p\|_*, \quad (6)$$

where \mathbf{X}_p is the modal matrix that unfold \mathcal{X} along the direction of the mode- p , α_p is weighting parameter of modes and $\sum_{p=1}^3 \alpha_p = 1$. However, this way will inevitably destroy the internal structure of HSI. Zheng et al. [26] propose the weighted sum of TNN (WSTNN). They consider the low-rankness of the three modes, which are two spatial modes and one spectral mode. WSTNN is defined as

$$\|\mathcal{X}\|_{WSTNN} := \sum_{p=1}^3 \alpha_p \|\mathcal{X}_p\|_{TNN}, \quad (7)$$

where \mathcal{X}_p is the mode- p permutation of \mathcal{X} .

III. MULTI-MODAL AND FREQUENCY-WEIGHTED TENSOR NUCLEAR NORM AND ITS NON-CONVEX APPROXIMATION.

A. Multi-modal and Frequency-weighted Tensor Nuclear Norm

In the definition of TNN (3), fft plays a core important role in the process of transforming t-products into matrix multiplication in the Fourier domain. It makes TNN to have fast calculation capabilities. However, when Fourier matrix \mathbf{F}_n can be treated as a transform matrix, $\bar{\mathcal{X}}$ is the feature tensor of \mathcal{X} . Combining with the physical meaning of fft , this feature information is frequency information. As shown in Fig. 1, for \mathcal{X} , its profile information is contained in the low-frequency

frontal slices, while its detailed and texture information is contained in the high-frequency ones. Specifically, for the clean Washington DC Mall dataset, sparse noises are added to form noise data, and then the zero frequency component and non-zero frequency component of the noise data are separated. Fig. 2 shows the grayscale image of the 60th band of all data. The above results show, when using TNN to explore low-

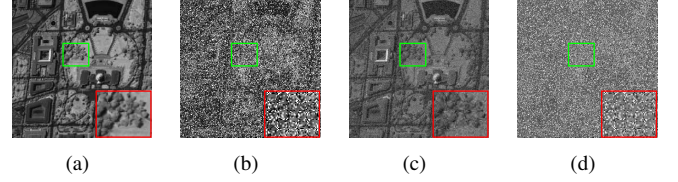


Fig. 2. Recovery of band 60 of Washington DC Mall dataset. (a) clear data \mathcal{X} . (b) noisy data \mathcal{Y} . (c) zero frequency component of \mathcal{Y} . (d) non-zero frequency component of \mathcal{Y} .

rankness of tensors, if it can reduce the shrinkage of the low-frequency matrix, and at the same time increase the penalty of the high-frequency matrix, which will show more excellent denoising performance. Therefore, in the TNN definition process of the tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, according to the physical meaning of the low and high frequency components. The frequency-weighted tensor nuclear norm (FWTNN), which is the weighted sum of the nuclear norm of the frequency slice matrix, is defined as

$$\|\mathcal{X}\|_{FW*} := \sum_{k=1}^{n_3} w_k (\bar{\mathbf{X}}^{(k)}) \|\bar{\mathbf{X}}^{(k)}\|_*, \quad (8)$$

where $w_k, k = 1, 2, \dots, n_3$ are frequency domain weight parameters.

In the definition of FWTNN (8), fft is directional. Generally speaking, this direction is the spectral direction. However, the low-rank information of HSI not only exists in the spectral dimension, but also the spatial dimensions. Besides, in the HSI denoising task, the HSI band may be polluted by stripe noise in the same direction. In a low-rank model that only considers the spectral dimension, these stripe noises will be regarded as part of the low-rank image, which will cause the denoised result to still contain some stripe noises. Furthermore, due to Gaussian noise is independent and identically distributed, this model cannot completely remove Gaussian noise. Therefore, it is necessary to consider the low-rankness of different dimensions for HSI denoising task. In the frequency domain, for the same HSI, the information expressed by \mathcal{X}_p permuted in different directions is even more different. It will be more in line with the physical properties of tensors in the frequency domain and help improve the accuracy of the denoised results. Based on the above analysis, we propose the multi-modal and frequency-weighted tensor nuclear norm (MFWTNN) as follows:

$$\|\mathcal{X}\|_{MFW*} := \sum_{p=1}^3 \alpha_p \|\mathcal{X}_p\|_{FW*} = \sum_{p=1}^3 \sum_{k=1}^{n_3} \alpha_p w_k^p \|\bar{\mathbf{X}}_p^{(k)}\|_*, \quad (9)$$

where $\bar{\mathcal{X}}_p = \text{fft}(\mathcal{X}_p, [], 3)$; $\bar{\mathbf{X}}_p^{(k)}$ is the k -th frontal slice of $\bar{\mathcal{X}}_p$ and its assigned weight is w_k^p ; $\alpha_p > 0$ and $\sum_{p=1}^3 \alpha_p = 1$.

B. Non-convex Multi-modal and Frequency-weighted Tensor Nuclear Norm

Although FWTNN improves TNN in the frequency direction, it inherits the shortcomings of TNN in the singular value direction. In the definition of FWTNN (8), the frequency nuclear norm is the ℓ_1 norm of singular values, and its approximation accuracy for the rank function is limited. In the same frequency slice, its singular values are treated equally. However, the major information of HSIs, such as smooth zones and profile, is contained in the larger singular value; the noise information of HSIs is contained in the smaller singular value [9–11, 33]. To combine both the frequency and the singular value prior information, we use the log norm to perform non-convex relaxation of the nuclear norm in the frequency slice to more accurately describe the tensor rank function[9]. Thus, we propose the non-convex multi-modal and frequency-weighted tensor nuclear norm (NonMFWTNN) to more accurately describe the low-rank structure of HSI \mathcal{X} . It is defined as follows:

$$\begin{aligned} \|\mathcal{X}\|_{MFW*,Log} &:= \sum_{p=1}^3 \alpha_p \|\mathcal{X}_p\|_{FW*,Log} \\ &= \sum_{p=1}^3 \sum_{k=1}^{n_3} \alpha_p w_k^p \log \left(\|\bar{\mathbf{X}}_p^{(k)}\|_* \right), \end{aligned} \quad (10)$$

where $\log \left(\|\bar{\mathbf{X}}_p^{(k)}\|_* \right) = \sum_{i=1}^{n_{12}} \left(\log \left(\sigma_i \left(\bar{\mathbf{X}}_p^{(k)} \right) + \varepsilon \right) \right)$, $n_{12} = \min(n_1, n_2)$. Fig. 3 shows the mechanism of NonMFWTNN.

C. Frequency-weighted parameter

In the definition of MFWTNN and NonMFWTNN, the value of the frequency-weighted parameter w_k^p is important. If $w_k^p = 1$, then MFWTNN degenerates into the WSTNN. The ideal frequency-weighted parameter w_k^p can take a smaller value in the low-frequency slice carrying the main information of HSI, then the penalty of the low-frequency slice nuclear norm can be reduced. When taking a larger value in the high-frequency slice containing noise, the penalty of the high-frequency slice nuclear norm can be increased, to fully remove the noise in the HSI. It can be seen from Fig. 4(a) that the nuclear norm is opposite to the distribution characteristics of the frequency-weighted parameter w_k^p in the frequency domain. However, the calculation of the nuclear norm brings high cost, which requires singular value decomposition. In a finite dimensional space, any two norms are equivalent. This shows that when transforming from the nuclear norm to the Frobenius norm, they are equal under the normalization condition. Thus, the reciprocal of the Frobenius norm of each frequency slice matrix can be used as the frequency-weighted parameter w_k^p .

$$w_k^p \left(\bar{\mathbf{X}}_p^{(k)} \right) = \frac{1}{\log \left(\|\bar{\mathbf{X}}_p^{(k)}\|_F^2 \right) + \delta}, \quad (11)$$

where $\delta = 10^{-6}$. The purpose of using log function is to shrink the scale of the value without changing the trend of

the Frobenius norm to prevent the extreme influence of the maximum value.

The above method of calculating weights is feasible when the clean HSI is known. However, for HSI denoising tasks, we often only have noisy HSI, and clean HSI is unknown. It can be seen from Fig. 4(b) that the weight calculated by the denoised HSI is closer to the result obtained by the clean HSI than the noisy HSI. Further, as shown in Fig. 3(e), as the number of iterations n continues to increase, the weight calculated by the denoised HSI will get closer to the result of clean HSI. Based on this, we propose an iterative calculation method to update the frequency-weighted parameter w_k^p . Assuming that the n th iteration gets $\bar{\mathbf{X}}_p^{(k,n)}$, then the corresponding $w_{k,(n+1)}^p$ of the $(n+1)$ th iteration is

$$w_{k,(n+1)}^p = C_1 h_{k,n}^p + C_2, \quad (12)$$

where $h_{k,n}^p = 1/(\log \left(\|\bar{\mathbf{X}}_p^{(k,n)}\|_F^2 \right) + \delta)$. C_1 is the scaling factor after frequency normalization; C_2 is a smaller constant.

IV. HSI DENOISING VIA MFWTNN AND NONMFWTNN MINIMIZATION

A. Proposed model

MFWTNN uses frequency components, modal information, and NonMFWTNN uses the physical meaning of singular value distribution on the basis of MFWTNN. They can provide a better approximation to the tensor rank. Then we use MFWTNN and NonMFWTNN to replace the regularization term Rank in (2) and propose the HSI denoising model as follows:

$$\begin{aligned} \arg \min_{\mathcal{X}, \mathcal{S}, \mathcal{N}} \sum_{p=1}^3 \alpha_p \|\mathcal{Z}_p\|_{FW*} + \lambda \|\mathcal{S}\|_1 + \tau \|\mathcal{N}\|_F^2, \\ s.t. \mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N}. \end{aligned} \quad (13)$$

$$\begin{aligned} \arg \min_{\mathcal{X}, \mathcal{S}, \mathcal{N}} \sum_{p=1}^3 \alpha_p \|\mathcal{Z}_p\|_{FW*,Log} + \lambda \|\mathcal{S}\|_1 + \tau \|\mathcal{N}\|_F^2, \\ s.t. \mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N}. \end{aligned} \quad (14)$$

Introducing auxiliary variables, model (13) and (14) are equivalent to

$$\begin{aligned} \arg \min_{\mathcal{X}, \mathcal{S}, \mathcal{N}} \sum_{p=1}^3 \alpha_p \|\mathcal{Z}_p\|_{FW*} + \lambda \|\mathcal{S}\|_1 + \tau \|\mathcal{N}\|_F^2, \\ s.t. \mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N}, \mathcal{Z}_p = \mathcal{X}_p, p = 1, 2, 3. \end{aligned} \quad (15)$$

$$\begin{aligned} \arg \min_{\mathcal{X}, \mathcal{S}, \mathcal{N}} \sum_{p=1}^3 \alpha_p \|\mathcal{Z}_p\|_{FW*,Log} + \lambda \|\mathcal{S}\|_1 + \tau \|\mathcal{N}\|_F^2, \\ s.t. \mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N}, \mathcal{Z}_p = \mathcal{X}_p, p = 1, 2, 3. \end{aligned} \quad (16)$$

By augmented Lagrangian multiplier method, the Lagrangian function of model (15) can be written as

$$\begin{aligned} L_{\mu_p, \beta}(\mathcal{X}, \mathcal{Z}_p, \mathcal{N}, \mathcal{S}, \Gamma_p, \Lambda) &= \lambda \|\mathcal{S}\|_1 + \tau \|\mathcal{N}\|_F^2 \\ &+ \langle \mathcal{Y} - (\mathcal{X} + \mathcal{S} + \mathcal{N}), \Lambda \rangle + \frac{\beta}{2} \|\mathcal{Y} - (\mathcal{X} + \mathcal{S} + \mathcal{N})\|_F^2 \\ &+ \sum_{p=1}^3 \left\{ \alpha_p \|\mathcal{X}_p\|_{FW*} + \langle \mathcal{X}_p - \mathcal{Z}_p, \Gamma_p \rangle + \frac{\mu_p}{2} \|\mathcal{X}_p - \mathcal{Z}_p\|_F^2 \right\}, \end{aligned} \quad (17)$$

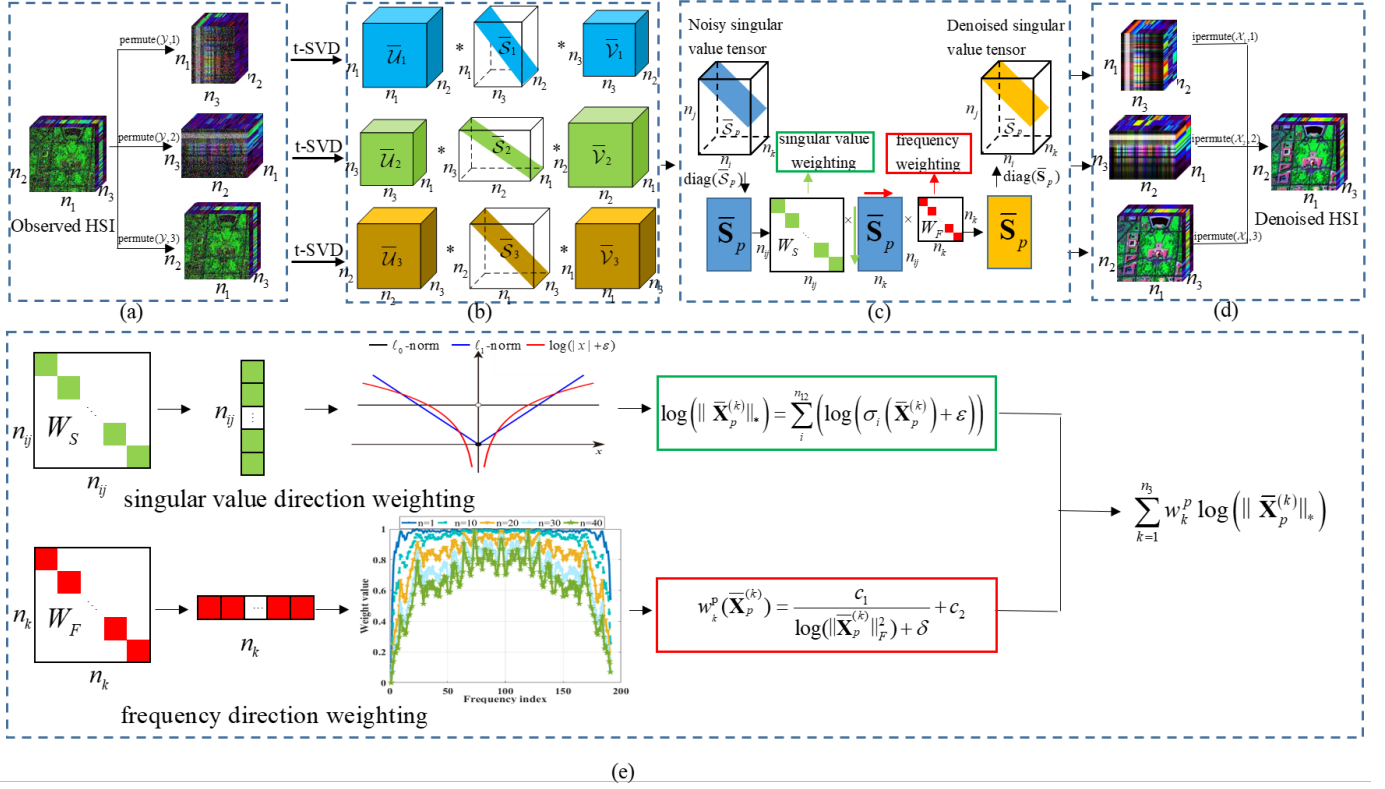


Fig. 3. Illustration of Non-MFWTNN. (a) Multi-modal permutations and Fourier transforms. (b) t-SVD of \mathcal{X}_p , $p=1,2,3$. (c) Shrinking the singular value tensor \bar{S}_p , where the size of \mathcal{X}_p is $n_i \times n_j \times n_k$, $n_{ij}=\min(n_i, n_j)$. The horizontal direction is the frequency weighting direction, and the vertical direction is the singular value weighting direction. (d) Synthesis from multi-modal denoised results. (e) The top is the singular value weighting, which is the weight of \bar{S}_p in the vertical direction. And the bottom is the frequency weighting, which is the weight of \bar{S}_p in the horizontal direction.

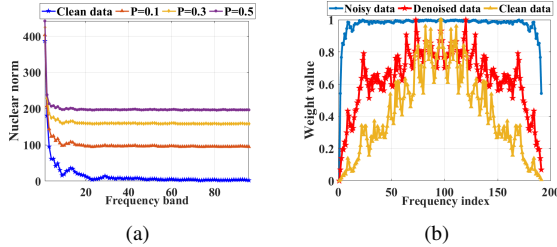


Fig. 4. (a) The distribution of the Frobenius norm of each frequency slice of Pavia data under different sparse noise intensities. (b) Frequency weights obtained from different data.

where Λ and Γ_p are the Lagrangian multipliers; β and μ_p are the Lagrange penalty parameters. Compared with MFWTNN, the model based on NonMFWTNN only differs in the steps of solving \mathcal{Z}_p . See the solution of \mathcal{Z}_p for details.

B. Algorithmic optimization

ADMM is an effective framework to solve these type of minimization problems [34]. When other variables are fixed at the n -th iteration, each variable in the Lagrangian function can be updated by solving its corresponding subproblem respectively at the $(n+1)$ -th iteration.

For \mathcal{Z}_p , $p = 1, 2, 3$, their corresponding subproblems can

be written as

$$\arg \min_{\mathcal{Z}_p} \alpha_p \|\mathcal{Z}_p\|_{FW*} + \frac{\mu_p}{2} \left\| \mathcal{Z}_p - \left(\mathcal{X}_p^n + \frac{\Gamma_p^n}{\mu_p} \right) \right\|_F^2. \quad (18)$$

The closed-form solution of (18) obtained from theorem 1 are as follows:

$$\mathcal{Z}_p^{n+1} = \mathcal{FW}^{w(\mathcal{X}_p^n), \frac{\alpha_p}{\mu_p}} \left(\mathcal{X}_p^n + \frac{\Gamma_p^n}{\mu_p} \right). \quad (19)$$

Similarly, when $\text{Rank}(\mathcal{X})$ is replaced by $\sum_{p=1}^3 \alpha_p \|\mathcal{X}_p\|_{FW*, \text{Log}}$ in (16), the sub-problem of \mathcal{Z}_p is rewritten as

$$\arg \min_{\mathcal{Z}_p} \alpha_p \|\mathcal{Z}_p\|_{FW*, \text{Log}} + \frac{\mu_p}{2} \left\| \mathcal{Z}_p - \left(\mathcal{X}_p^n + \frac{\Gamma_p^n}{\mu_p} \right) \right\|_F^2. \quad (20)$$

The closed-form solution of (20) obtained from theorem 2 are as follows:

$$\mathcal{Z}_p^{n+1} = \mathcal{DW}^{w(\mathcal{X}_p^n), \frac{\alpha_p}{\mu_p}} \left(\mathcal{X}_p^n + \frac{\Gamma_p^n}{\mu_p} \right). \quad (21)$$

For \mathcal{X} , its corresponding subproblem can be reformulated as

$$\begin{aligned} \mathcal{X}^{n+1} = \arg \min_{\mathcal{X}} & \sum_{p=1}^3 \frac{\mu_p}{2} \left\| \mathcal{X} - \mathcal{Z}_p^{n+1} + \frac{\Gamma_p^n}{\mu_p} \right\|_F^2 \\ & + \frac{\beta}{2} \left\| \mathcal{Y} - (\mathcal{X} + \mathcal{S}^n + \mathcal{N}^n) + \frac{\Lambda^n}{\beta} \right\|_F^2. \end{aligned} \quad (22)$$

\mathcal{X} can be updated as follows:

$$\mathcal{X}^{n+1} = \frac{\sum_{p=1}^3 \mu_p \left(\mathcal{Z}_p^{n+1} - \frac{\Gamma_p^n}{\mu_p} \right) + \beta (\mathcal{Y} - \mathcal{S}^n - \mathcal{N}^n + \frac{\Lambda^n}{\beta})}{1 + \beta}. \quad (23)$$

For \mathcal{S} , its corresponding subproblem can be reformulated as

$$\arg \min_{\mathcal{S}} \lambda \|\mathcal{S}\|_1 + \frac{\beta}{2} \|\mathcal{Y} - (\mathcal{X}^{n+1} + \mathcal{S} + \mathcal{N}^n) + \frac{\Lambda^n}{\beta}\|_F^2. \quad (24)$$

It can be solved by the soft-thresholding operator [35] as:

$$\mathcal{S}^{n+1} = \text{shrink} \left(\mathcal{Y} - \mathcal{X}^{n+1} - \mathcal{N}^{n+1} + \frac{\Lambda^n}{\beta}, \frac{\lambda}{\beta} \right). \quad (25)$$

For \mathcal{N} , its corresponding subproblem can be reformulated as

$$\arg \min_{\mathcal{N}} \tau \|\mathcal{N}\|_F^2 + \frac{\beta}{2} \|\mathcal{Y} - (\mathcal{X}^{n+1} + \mathcal{S}^{n+1} + \mathcal{N}) + \frac{\Lambda^n}{\beta}\|_F^2. \quad (26)$$

\mathcal{N} can be updated as follows :

$$\mathcal{N}^{n+1} = \frac{\beta (\mathcal{Y} - \mathcal{X}^{n+1} - \mathcal{S}^{n+1} + \frac{\Lambda^n}{\beta})}{2\tau + \beta}. \quad (27)$$

For multipliers Γ_p and Λ , they can be updated as follows:

$$\begin{cases} \Gamma_p^{n+1} = \Gamma_p^n + \mu_p (\mathcal{Z}_p^{n+1} - \mathcal{X}^{n+1}), p = 1, 2, 3 \\ \Lambda^{n+1} = \Lambda^n + \beta (\mathcal{Y} - \mathcal{X}^{n+1} - \mathcal{S}^{n+1} - \mathcal{N}^{n+1}). \end{cases} \quad (28)$$

C. Time Complexity Analysis

The HSI denoising models based on MFWTNN and NonMFWTNN can be solved by Algorithm 1. Further, we discuss the time complexity of this algorithm. The cost of computing \mathcal{X} , \mathcal{S} , and \mathcal{N} are all $O(n_1 n_2 n_3)$. Calculating w_k has a complexity of $O(\sum_{i=1}^3 \min(n_i, n_{i+1}) n_{i+2})$. Computing $\mathcal{Z}_p, p = 1, 2, 3$, in both MFWTNN based solver and NonMFWTNN based solver, has a complexity of $O(n_1 n_2 n_3 \log(n_1 n_2 n_3) + \sum_{i=1}^3 \max(n_i, n_{i+1}) \min^2(n_i, n_{i+1}) n_{i+2})$, where $n_4 = n_1, n_5 = n_2$. Then, by calculating the complexity of the above variables, the total complexity of the proposed algorithm can be obtained as $O(n_1 n_2 n_3 \log(n_1 n_2 n_3) + \sum_{i=1}^3 \max(n_i, n_{i+1}) \min^2(n_i, n_{i+1}) n_{i+2} + \sum_{i=1}^3 \min(n_i, n_{i+1}) n_{i+2} + 3n_1 n_2 n_3)$.

Algorithm 1 HSI Restoration via the MFWTNN and NonMFWTNN minimization

Input: The observed tensor \mathcal{Y} ; weight parameters c_1, c_2, ε ; regularization parameters λ, τ ; and stopping criterion ϵ .

Output: Denoised image \mathcal{X} .

- 1: Initialize: $\mathcal{Y} = \mathcal{X} = \mathcal{S} = \mathcal{N} = \mathcal{Z}_p$; $\Gamma_p = \Lambda = 0$; $\mu_p = \beta = 10^{-3}$; $p = 1, 2, 3$; $\mu_{max} = 10^{10}$; $\rho = 1.2$ and $n = 1$.
 - 2: Repeat until convergence:
 - 3: Update $\mathcal{X}, \mathcal{S}, \mathcal{N}, \mathcal{Z}_p, \Lambda, \beta, \mu_p, w_k, \Gamma_p$ via
 - step 1: Update \mathcal{Z}_p by (19) or (21)
 - step 2: Update \mathcal{X} by (23)
 - step 3: Update \mathcal{S} by (25)
 - step 4: Update \mathcal{N} by (27)
 - step 5: Update Γ_p, Λ by (28)
 - step 6: Update $\mu_p = \rho \mu_p, \beta = \rho \beta, w_k$ by (12)
 - 4: Check the convergence condition.
-

V. EXPERIMENT RESULTS AND DISCUSSION

In this section, we conduct experiments on simulated and real-world HSIs to substantiate the effectiveness of the proposed MFWTNN and NonMFWTNN model for HSI denoising. For evaluating the effectiveness of the models, our models are compared with four state-of-the-art HSI denoised methods, i.e., BM4D [36], LRMR [6], LRTDTV [32], 3DLogTNN [28]. Due to the BM4D method that is only suitable for removing Gaussian noise, we first use the RPCA [21] model to remove the sparse noise, and then use it to obtain the denoised results. Before denoising HSIs, all pixels in the HSI band are linearly mapped to [0,1]. In all experiments, the hyper-parameters in these comparative methods are manually adjusted according to the suggestions in relevant papers. In Section V-C, the parameters are discussed in detail in our proposed models. We use visual comparison and quantitative comparison to comprehensively evaluate the performance of different denoising methods. In quantitative comparisons, we use five quantitative picture quality indices (PQIs), i.e., PSNR [37], SSIM [38], FSIM [39], ERGAS [40] and SAM [41], to assess the denoised results. The better denoised results correspond to higher values in MPSNR, MSSIM, MFSIM, and lower values in EGRAS, MSAM.

A. Simulated HSI Data Experiments

In this section, we select two HSI datasets for the simulated experiments to show the robustness of our methods. One is filmed by the reflection optical system imaging spectrometer (ROSIS-03) from Pavia City Center dataset¹. The size is 1096×1096 , with a total of 102 bands. The reason we chose a sub-block with a size of $200 \times 200 \times 80$ as a simulation dataset is because some bands are seriously polluted by noise and have lost the meaning of reference. The other one is collected by a HSI acquisition sensor (HYDICE) from Washington DC Mall dataset². The size is 1208×307 , with a total of 191 bands. We choose a sub-block with a size of $256 \times 256 \times 191$ as a simulation dataset.

The observed HSIs are degraded by the hybrid noises, i.e., Gaussian noise, sparse noise, and stripe noise. Thus, to better simulate the degradation mechanism of HSI, we add various intensities of hybrid noises to the clean HSI. In the simulation experiments, the hybrid noises with eight different intensity levels are added to simulation dataset band by band. Let G and P denote the variances of Gaussian white noise and percentages of impulse noise, respectively. In noise cases 1-5, the same intensity noise is added to all the bands. In noise case 1, $G=0.1$ and $P=0.2$; In noise case 2, $G=0.1$ and $P=0.3$; In noise case 3, $G=0.1$ and $P=0.4$; In noise case 4, $G=0.15$ and $P=0.2$; In noise case 5, $G=0.2$ and $P=0.2$; In noise cases 6-8, the noise intensities are different for different bands. In noise case 6, $G=0.1$ and P is randomly selected from 0.2 to 0.4. In noise case 7, G is randomly selected from 0.1 to 0.3 and $P=0.2$; In noise case 8, G is randomly selected from 0.1 to 0.3 and P is randomly selected from 0.1 to 0.3. In addition, stripe

¹<http://www.ehu.es/ccwintco/index.php/>

²<http://lesun.weebly.com/hyperspectral-data-set.html>

TABLE I
QUANTITATIVE COMPARISON AND TIME OF EIGHT DENOISING MODELS ON SIMULATED DATASET UNDER EIGHT NOISE CASES.

Case	Level	Index	Pavia City Center						Washington DC Mall							
			Noise	BM4D	LRMR	LRTDTV	3DLogTNN	MFWTNN	NonMFWTNN	Noise	BM4D	LRMR	LRTDTV	3DLogTNN	MFWTNN	NonMFWTNN
Case 1	G=0.1	MPSNR	11.122	29.701	31.259	32.297	33.487	32.470	34.398	11.024	31.209	31.856	33.243	35.518	34.296	36.344
		MSSIM	0.105	0.920	0.905	0.914	0.942	0.928	0.945	0.077	0.889	0.848	0.900	0.932	0.923	0.942
		MFSIM	0.510	0.949	0.946	0.942	0.964	0.955	0.968	0.438	0.919	0.929	0.931	0.961	0.950	0.966
	P=0.2	MERGA	1013.518	119.608	99.715	87.316	77.865	89.194	69.458	1333.908	115.882	113.823	96.698	75.616	86.176	67.825
		MSAM time/s	45.712	5.840	6.824	4.930	4.394	5.656	4.339	43.237	4.519	4.937	4.175	3.148	3.579	2.864
Case 2	G=0.1	MPSNR	-	115.944	116.111	128.850	86.658	66.917	96.791	-	544.679	424.176	545.315	404.145	330.736	543.209
		MPSNR	9.541	28.864	30.173	31.188	32.823	31.600	33.420	9.420	30.585	30.595	32.208	35.030	33.329	35.746
		MSSIM	0.065	0.910	0.881	0.897	0.933	0.915	0.935	0.049	0.878	0.814	0.881	0.927	0.908	0.935
	P=0.3	MFSIM	0.452	0.942	0.935	0.931	0.958	0.947	0.961	0.384	0.912	0.913	0.919	0.957	0.941	0.962
		MERGA	1216.767	132.124	112.427	99.239	83.831	97.753	77.512	1611.157	124.189	130.309	110.034	80.302	96.378	73.121
Case 3	G=0.1	MPSNR	47.676	6.236	7.250	5.349	4.570	6.065	4.528	47.557	5.675	5.667	4.718	3.305	3.987	3.052
		MFSIM	-	116.229	109.304	125.285	89.352	67.251	98.926	-	554.675	382.550	539.389	407.267	330.584	554.722
		MPSNR	8.384	27.745	28.937	29.741	31.339	30.570	31.879	8.253	29.630	29.149	30.686	33.579	32.342	34.402
	P=0.4	MSSIM	0.043	0.892	0.848	0.873	0.915	0.898	0.913	0.034	0.859	0.768	0.850	0.910	0.887	0.922
		MFSIM	0.411	0.933	0.920	0.917	0.945	0.936	0.948	0.346	0.899	0.890	0.901	0.940	0.928	0.947
Case 4	G=0.15	MERGA	1390.482	150.540	128.938	117.629	99.511	109.129	92.185	1846.435	138.002	151.548	133.569	94.251	108.312	85.311
		MSAM time/s	48.580	6.745	7.740	6.007	5.042	6.506	4.925	50.441	5.262	6.635	6.007	3.761	4.478	3.490
		MPSNR	-	111.924	113.685	132.354	94.827	72.480	107.781	-	528.381	447.568	584.706	424.376	370.543	601.149
	P=0.2	MPSNR	10.716	27.420	29.013	30.111	31.133	30.451	32.021	10.635	28.928	29.286	31.359	33.121	32.220	34.568
		MSSIM	0.088	0.879	0.849	0.867	0.905	0.891	0.913	0.066	0.834	0.774	0.856	0.896	0.881	0.916
Case 5	G=0.2	MFSIM	0.483	0.925	0.921	0.913	0.942	0.933	0.948	0.411	0.879	0.893	0.902	0.941	0.923	0.952
		MERGA	1060.262	154.021	127.846	112.153	101.226	110.568	91.095	1379.815	148.879	150.799	120.696	102.607	109.969	83.629
		MSAM time/s	46.411	6.671	7.690	5.848	5.227	6.562	5.050	44.251	5.774	6.561	5.241	4.167	4.636	3.518
	P=0.2	MPSNR	-	112.103	112.627	131.750	94.155	72.962	104.037	-	529.783	446.593	585.548	445.453	366.438	600.614
		MPSNR	10.265	25.619	27.321	28.605	29.134	29.088	30.520	10.188	27.245	27.427	29.970	31.223	30.919	33.034
Case 6	G=0.1	MSSIM	0.074	0.835	0.791	0.821	0.860	0.856	0.885	0.056	0.783	0.705	0.814	0.859	0.844	0.891
		MFSIM	0.461	0.900	0.898	0.886	0.917	0.913	0.931	0.390	0.841	0.861	0.875	0.922	0.900	0.933
		MERGA	1115.059	188.524	154.623	133.338	126.652	128.254	108.312	1437.539	179.235	183.763	141.567	132.159	127.459	100.032
	P=0.2	MSAM time/s	46.978	7.170	8.385	6.685	5.981	7.143	5.580	45.387	6.845	8.049	6.214	5.291	5.466	4.167
		MPSNR	-	115.494	117.568	126.082	88.231	67.559	97.910	-	547.776	427.316	540.747	399.630	340.013	546.378
Case 7	G=0.1	MPSNR	9.515	28.761	30.090	31.062	32.569	31.512	33.405	9.512	30.555	30.572	32.063	34.525	33.337	35.728
		MSSIM	0.066	0.908	0.879	0.895	0.927	0.913	0.937	0.052	0.878	0.812	0.881	0.909	0.908	0.937
		MFSIM	0.452	0.942	0.933	0.931	0.956	0.946	0.960	0.388	0.911	0.912	0.920	0.950	0.941	0.960
	P=(0.2,0.4)	MERGA	1235.513	133.652	113.657	100.960	86.176	98.851	78.252	1605.940	124.905	131.211	112.085	84.541	96.388	73.122
		MSAM time/s	47.824	6.281	7.294	5.501	4.991	6.148	4.589	47.511	4.814	5.708	4.944	3.907	3.991	3.045
Case 8	G=(0.1,0.3)	MPSNR	-	115.407	116.126	128.118	87.868	67.084	97.160	-	543.748	427.928	541.382	411.018	331.695	543.483
		MPSNR	10.193	25.589	27.437	28.773	29.211	29.218	30.800	10.175	27.445	27.667	30.182	31.356	31.197	32.843
		MSSIM	0.074	0.833	0.798	0.826	0.868	0.861	0.893	0.057	0.795	0.714	0.827	0.869	0.853	0.895
	P=0.2	MFSIM	0.461	0.898	0.900	0.890	0.920	0.916	0.934	0.393	0.848	0.865	0.883	0.924	0.906	0.937
		MERGA	1131.564	189.307	153.217	131.728	126.702	126.994	105.541	1444.165	176.071	182.967	142.743	128.476	124.285	104.171
Case 9	G=(0.1,0.3)	MSAM time/s	47.045	7.138	8.518	6.650	6.194	6.885	5.672	45.531	6.700	8.063	6.347	5.226	5.369	4.389
		MPSNR	-	116.565	116.969	125.336	87.113	67.129	96.770	-	543.508	429.160	534.911	394.217	336.045	540.726
		MPSNR	9.052	24.653	26.402	27.922	28.711	28.489	29.332	8.937	26.614	26.396	29.159	29.804	30.031	30.303
	P=(0.2,0.4)	MSSIM	0.053	0.812	0.759	0.803	0.852	0.839	0.865	0.040	0.771	0.662	0.802	0.836	0.818	0.844
		MFSIM	0.425	0.887	0.884	0.876	0.913	0.903	0.916	0.357	0.830	0.841	0.870	0.914	0.887	0.909
Case 10	stipes	MERGA	1299.108	210.596	172.408	145.080	134.585	136.996	125.036	1684.836	193.299	209.482	165.661	162.363	144.022	145.351
		MSAM time/s	48.246	7.394	9.122	7.027	7.233	7.689	7.526	48.823	7.322	9.412	7.297	6.768	6.360	6.325
			-	116.280	113.436	125.442	88.043	67.376	97.515	-	543.885	424.387	532.568	407.518	341.152	548.572

noises are added to the 54–64 bands of the Pavia City Center and the 70–100 bands of the Washington DC Mall. we show the evaluation results of all denoising models including visual and quantitative quality on Pavia City Center and Washington DC Mall datasets.

Table I reports the denoised results and CPU running time by all the compared methods under the condition of the eight noise cases on simulated datasets. When compared to the other methods, our proposed methods obtain the optimal PQIs among all the denoising models in most noise cases, indicating the advantage of the proposed methods in HSI denoising. Although LRTDTV also considers the low-rankness between different modes, it is affected by the matrix unfold operation which destroys the connection between the modes. Therefore, LRMR model obtains the lower PQIs. BM4D and LRTDTV use non-local similarity information and spatial smoothing information respectively. These make the denoised result excessively smooth and lack texture details. Although 3DlogTNN considers the correlation between different modes, it is still not accurate enough to represent the low-rankness of HSI. MFWTNN uses the physical meaning of singular values and frequency components to improve 3DTNN respectively. But it is not the best description of the low-rankness of HSI. NonMFWTNN inherits the advantages of 3DLogTNN and MFWTNN, and at the same time improves the ability to express low-rankness of HSI in terms of frequency components and singular values. Although the time complexity of our models is higher, our models have obtained better denoised results.

In Fig. 5, we show the denoised results of the Pavia City Center and Washington DC Mall dataset in terms of PSNR and SSIM under all noisy cases experiments. As this figure shows, our methods obtain the optimal PSNR and SSIM values in most bands. In Fig. 6, we show the 50th band of the denoised results of all comparison methods in case 2 on Pavia City Center dataset. Also in Fig. 7, we show the 57th band of the denoised results in case 7 on Washington DC Mall dataset. As these grayscale images show, LRMR can not completely remove some high-intensity noises and retain more noise. For BM4D and LRTDTV, although they can remove more noise, they also lose more details. This makes the denoised result too smooth. Compared with them, our proposed models can remove more noise while retaining more details. The spectral signatures are important indicators for HSI. To further compare the denoised image quality, we show the original and denoised spectral signature curves. Therefore, we show the spectral curve of the pixel (170, 99) under case 1 and pixel (234, 171) under case 6 in Fig. 8. It can be seen that spectral signature curves are relative to ERGAS values in table I, and our models can obtain the optimal value.

B. Real HSI Data Experiments

In this subsection, we chose the real HSI dataset in this experiment. It is AVIRIS Indian Pines dataset³. It is filmed by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS). The reason we chose a sub-block with a size of $145 \times 145 \times 220$

³<https://engineering.purdue.edu/>

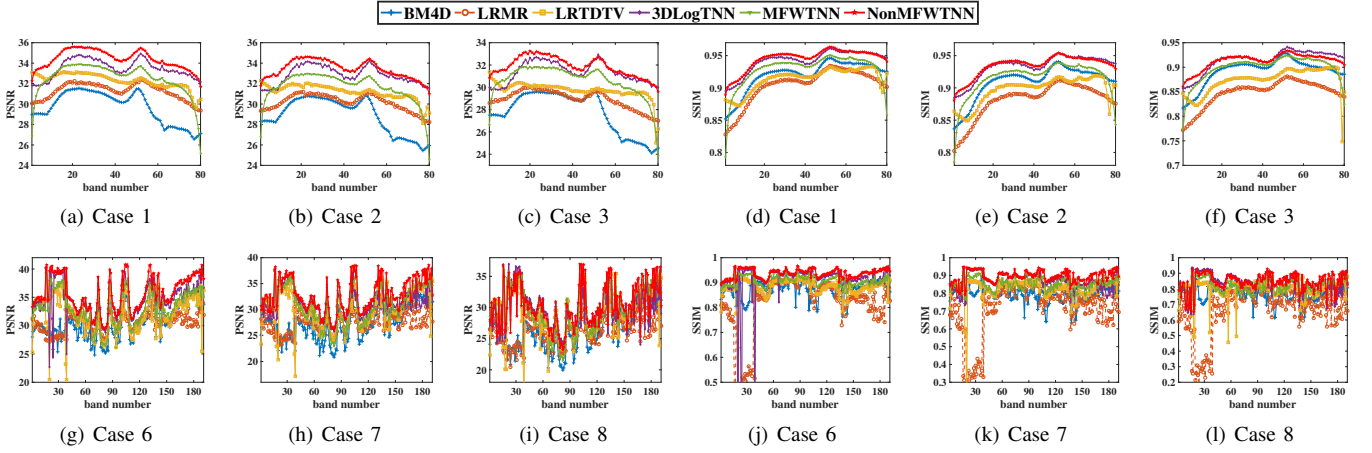


Fig. 5. The PSNR and SSIM of all denoising models for each band under different noise cases in simulated dataset. The top is Pavia City Center dataset, the bottom is Washington DC Mall dataset.

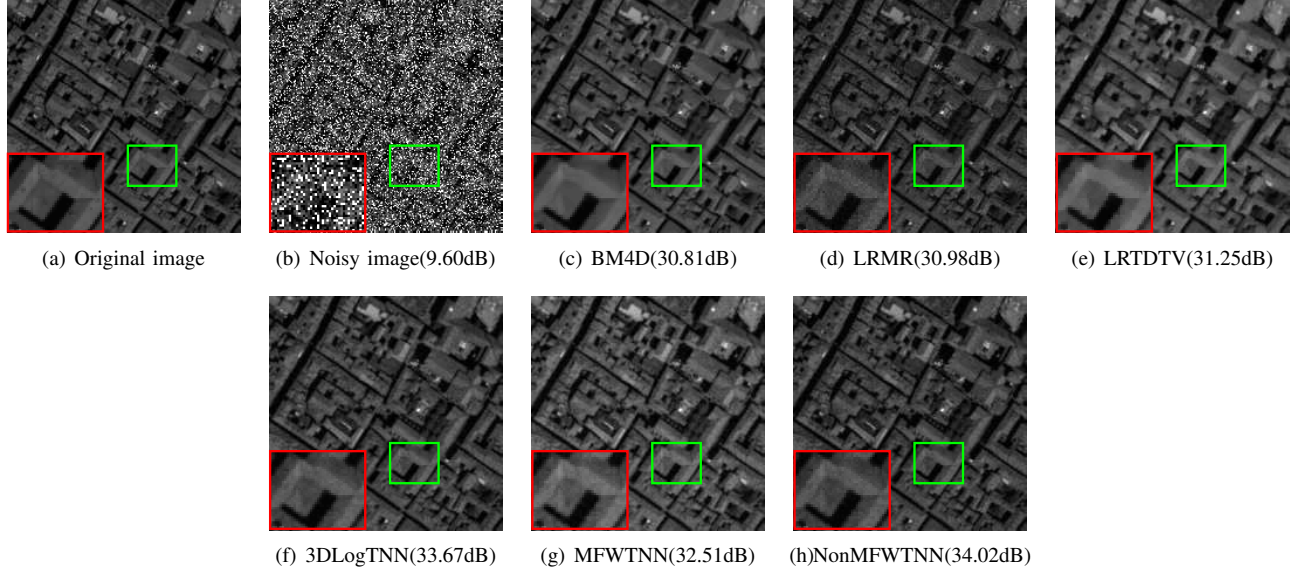


Fig. 6. The 50th band of the denoising results of the Pavia City Center dataset under noise Case 2.

as a real dataset is because some bands are seriously polluted by Gaussian white noise and impulse noise and have lost the meaning of reference. We evaluate our denoising model by visual evaluation and vertical average profile. In Fig. 9, we show the 105th band of the observed and denoised results in Indian dataset. It can be seen that it is seriously polluted by Gaussian and sparse noise, and the original ground features information cannot be distinguished. Although the competitive denoising methods restore the feature information of HSI, the denoised HSIs still lose detailed information or residual noise, resulting in missing or blurred texture details of HSI. Our models not only restore more feature information of HSI but also remove more local noise than the competitive methods. Therefore, these denoised HSIs have clearer texture information. In Fig. 10, we show the vertical mean profiles of the 50th band of Indian dataset. Due to the influence of hybrid noises, the vertical mean profiles of the observed HSI show rapid fluctuations. This means that the smaller fluctuations in the vertical mean profiles, the higher quality of the denoised

HSI. Obviously, the mean profile curves obtained by our model are the most stable and its fluctuation is the smallest. This result is relative to the facts reflected in Fig. 9.

C. Discussion

1) Parameter Analysis: In the NonMFWTNN model, five parameters need to be discussed, i.e., the modal weight parameter α , the regularization parameters λ , τ and frequency weight parameter C_1 , and C_2 . All experiments are done under Case 2.

The modal weight α is to control the ratio between different modes. The low-rank characteristics of the spatial dimensions are similar. To reduce the parameters, we set the weights of the two spatial dimensions to equal weights. Therefore, it is set to $\alpha = [1, 1, \alpha_3]/(1 + 1 + \alpha_3)$, which makes $\sum_{i=1}^3 \alpha_i = 1$. We study the impact of α_3 which is selected from 0.1 to 1.0 on the robustness of the model, and its optimal value is obtained at 0.2. It can be observed from

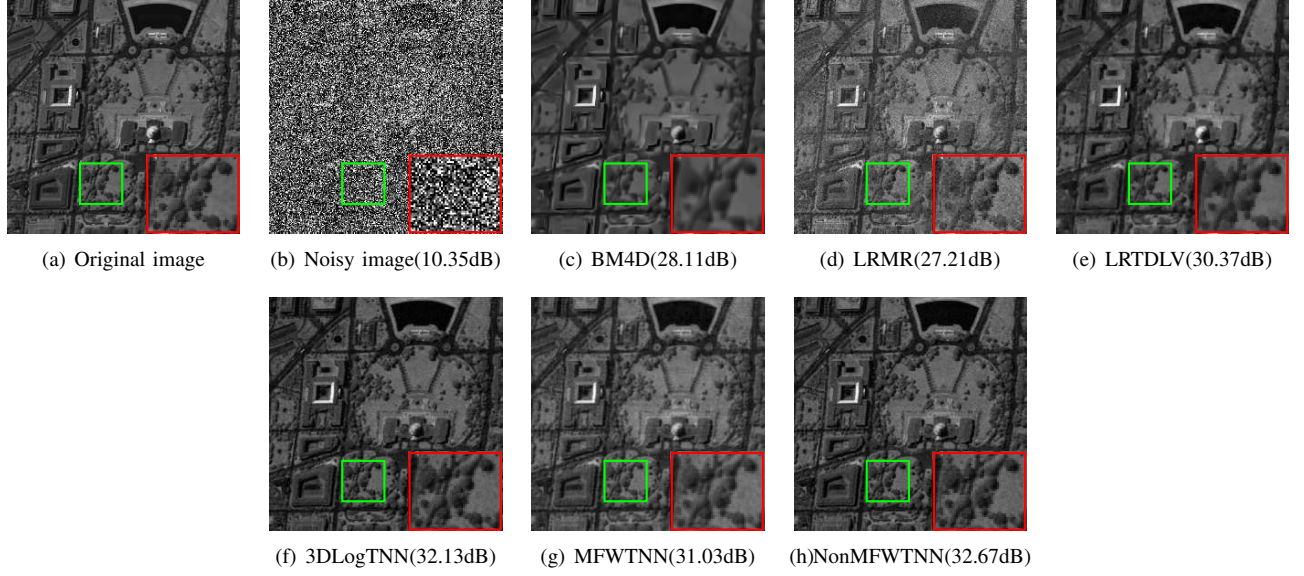


Fig. 7. The 57th band of the denoising results of the Washington DC Mall dataset under noise Case 7.

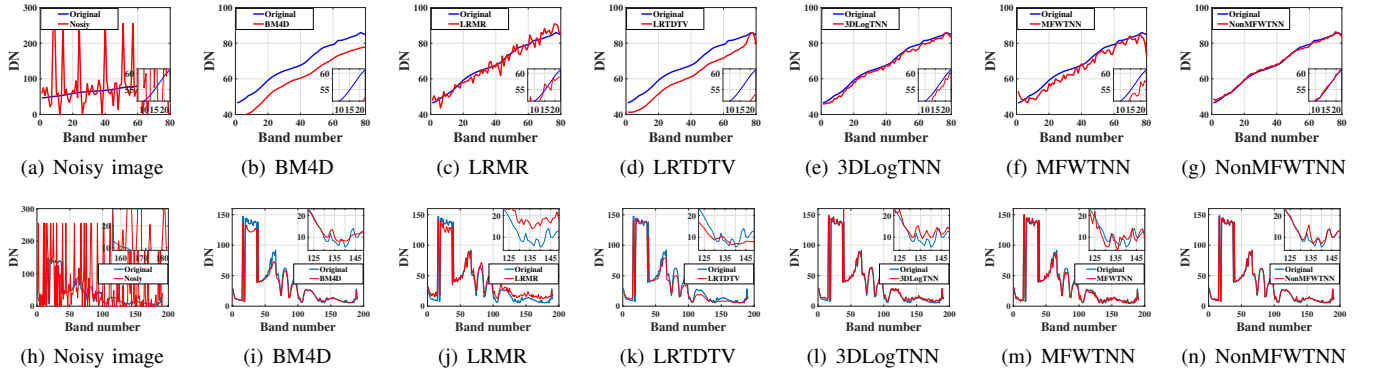


Fig. 8. The top is spectral signatures curve of Pavia City Center dataset in (170,99) under noise case 1, the bottom is spectral signatures curve of Washington DC Mall in (234,171) under noise case 6

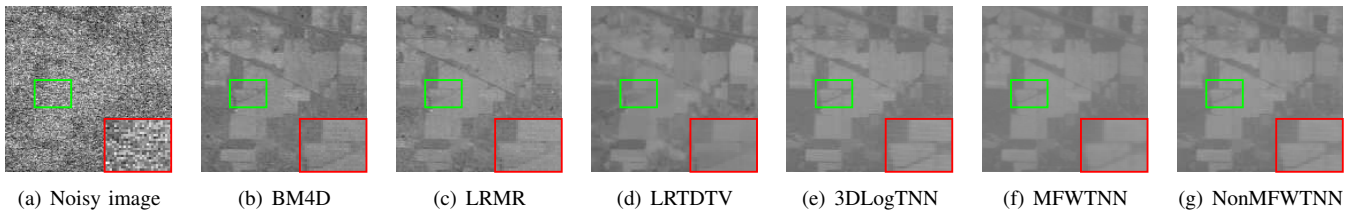


Fig. 9. The 105th band of all denoising results for the Indian dataset.

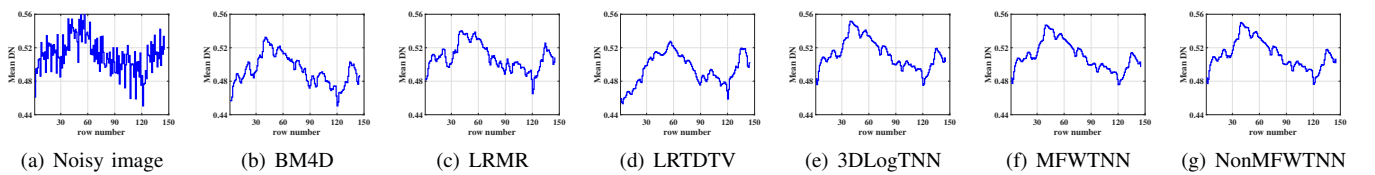


Fig. 10. The vertical mean profiles of band 50 in Indian dataset.

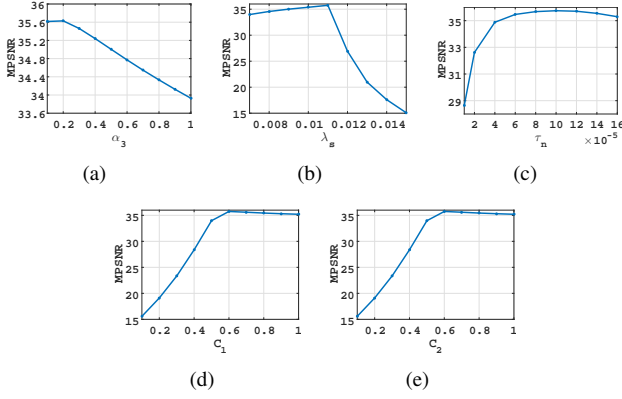


Fig. 11. MPSNR values with respect to different values of parameters.

Fig. 11(a) that when α_3 takes a smaller value, the PSNR value of the denoised result is better. This is because the low-rank characteristic of the spectral mode is stronger than the spatial mode. The regularization parameter τ controls the weight of the Gaussian noise term. Its value is related to the intensity of Gaussian noise. Therefore, it is set to $\tau = \tau_n/\sigma$, where σ is Gaussian noise intensity. We study the impact of τ_n which is selected from 10^{-5} to 1.6×10^{-4} on the robustness of the model, and its optimal value is obtained at 10^{-4} . It can be observed from Fig. 11(c) that when $\tau \leq 0.0001$, the MPSNR of the denoised result is low and fluctuates greatly. The main reason is that when τ is small, the ability to remove Gaussian noise is weak, and there is a lot of remaining Gaussian noise. The regularization parameter λ controls the weight of the sparse noise term and is defined as $\lambda = \lambda_s \left(\frac{\alpha_1}{\sqrt{\max(n_2, n_3)n_1}} + \frac{\alpha_2}{\sqrt{\max(n_3, n_1)n_2}} + \frac{\alpha_3}{\sqrt{\max(n_1, n_2)n_3}} \right)$. As observed from Fig. 11(b), our model can obtain remarkable results when λ_s which is selected from 0.006 to 0.016. Its optimal value is obtained at 0.011. Similar to τ , when λ is small, the noise removal effect is not good. The frequency weight parameter C_1 reflects the scaling ratio of the weight. Fig. 11(d) shows the PSNR values when C_1 is selected from 0.1 to 1.0. It can be seen from these figures that when C_1 is small, the frequency weight becomes smaller, and the noise is more difficult to remove; when C_1 is large, the frequency weight becomes larger, and the detailed information of tensor is removed. Its optimal value is obtained at 0.6. The frequency weight parameter C_2 reflects the initial weight of the iteration. Fig. 11(e) shows the PSNR values when C_2 is selected from 0.1 to 1.0. Its optimal value is obtained at 0.6. It shrinks equally of all frequency components. It inherits the advantages of TNN and ensures that noise can be removed correctly.

2) Convergence Analysis: Fig.12 shows the MPSNR value, the MSSIM value, and the Error value of the proposed model according to the iterations. As shown in Fig.12, the number of iterations increases, the value gradually stabilizes by the proposed method, which justifies the numerical convergence of the proposed method.

VI. CONCLUSION

In this paper, we proposed the multi-modal and frequency-weighted tensor nuclear norm (MFWTNN) and Non-convex

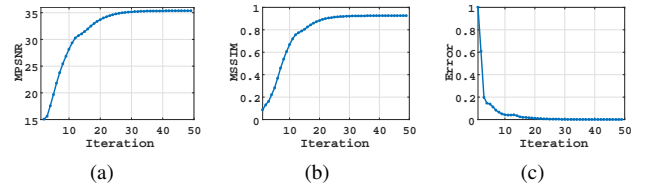


Fig. 12. MPSNR, SSIM and Error values with respect to the number of iterations.

MFWTNN to approximate the rank function and applied them to the HSI denoising task. The MFWTNN not only considered the correlation between two spatial modes and one spectral mode but also considered the frequency meaning of HSI in the Fourier domain. And we gave the frequency weight adaptive calculation way. Since the calculation of weights depends on iteratively restored and observed HSI, our models are more robust to different degrees of noise. It can better characterize the low-rankness of HSI. Besides, based on the MFWTNN model, we also considered the physical meaning of the internal singular values of the frequency slice and proposed its non-convex approximation. They powerfully improved capability and flexibility for describing low-rankness in HSIs. The experiments conducted with both simulated and real HSI datasets showed that our proposed models based HSI denoising model were competitive methods to remove the hybrid noises.

REFERENCES

- [1] C. Zhao, C. Li, S. Feng, N. Su, and W. Li, "A spectral-spatial anomaly target detection method based on fractional fourier transform and saliency weighted collaborative representation for hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5982–5997, 2020.
- [2] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.
- [3] H. Zeng, X. Xie, H. Cui, H. Yin, and J. Ning, "Hyperspectral image restoration via global l_{1-2} spatial-spectral total variation regularized local low-rank tensor recovery," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.
- [4] W. He, H. Zhang, H. Shen, and L. Zhang, "Hyperspectral image denoising using local low-rank matrix recovery and global spatial-spectral total variation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 713–729, 2018.
- [5] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 138–142, 2008.
- [6] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, 2013.

- [7] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 178–188, 2015.
- [8] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 730–742, 2018.
- [9] Q. Xie, Q. Zhao, D. Meng, Z. Xu, S. Gu, W. Zuo, and L. Zhang, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1692–1700.
- [10] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, and G. He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5366–5380, 2017.
- [11] Y. Lou and M. Yan, "Fast l1–l2 minimization via a proximal operator," *Journal of Scientific Computing*, vol. 74, no. 2, pp. 767–785, 2018.
- [12] C. Gao, N. Wang, Q. Yu, and Z. Zhang, "A feasible nonconvex relaxation approach to feature selection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 25, no. 1, 2011.
- [13] Y. Chen, J. Li, and Y. Zhou, "Hyperspectral image denoising by total variation-regularized bilinear factorization," *Signal Processing*, vol. 174, p. 107645, 2020.
- [14] L. Yang, L. Xu, J. Peng, Y. Song, A. Wong, and D. A. Clausi, "Nonlocal band-weighted iterative spectral mixture model for hyperspectral imagery denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5588–5601, 2020.
- [15] R. A. Harshman and M. E. Lundy, "Parafac: Parallel factor analysis," *Computational Statistics & Data Analysis*, vol. 18, no. 1, pp. 39–72, 1994.
- [16] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the parafac model and statistical performance analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3717–3724, 2012.
- [17] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [18] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2012.
- [19] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 1, pp. 148–172, 2013.
- [20] Z. Zhang and S. Aeron, "Exact tensor completion using t-svd," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1511–1526, 2016.
- [21] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 925–938, 2019.
- [22] C. Mu, B. Huang, J. Wright, and D. Goldfarb, "Square deal: Lower bounds and improved relaxations for tensor recovery," in *International Conference on Machine Learning*. PMLR, 2014, pp. 73–81.
- [23] H. Zeng, Y. Chen, X. Xie, and J. Ning, "Enhanced non-convex low-rank approximation of tensor multi-modes for tensor completion," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 164–177, 2021.
- [24] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 10, pp. 4589–4604, 2017.
- [25] 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 Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 6196–6213, 2018.
- [26] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-Y. Ji, and T.-H. Ma, "Tensor n-tubal rank and its convex relaxation for low-rank tensor recovery," *Information Sciences*, vol. 532, pp. 170–189, 2020.
- [27] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [28] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-H. Ma, and T.-Y. Ji, "Mixed noise removal in hyperspectral image via low-fibered-rank regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 734–749, 2019.
- [29] C. J. Hillar and L.-H. Lim, "Most tensor problems are np-hard," *Journal of the ACM (JACM)*, vol. 60, no. 6, pp. 1–39, 2013.
- [30] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, PhD thesis, Stanford University, 2002.
- [31] O. Rojo and H. Rojo, "Some results on symmetric circulant matrices and on symmetric centrosymmetric matrices," *Linear algebra and its applications*, vol. 392, pp. 211–233, 2004.
- [32] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1227–1243, 2017.
- [33] H. Zeng, X. Xie, and J. Ning, "Hyperspectral image denoising via global spatial-spectral total variation regularized nonconvex local low-rank tensor approximation," *Signal Processing*, vol. 178, p. 107805, 2021.
- [34] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.
- [35] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-

- rank matrices,” *arXiv preprint arXiv:1009.5055*, 2010.
- [36] M. Maggioni and A. Foi, “Nonlocal transform-domain denoising of volumetric data with groupwise adaptive variance estimation,” in *Computational Imaging X*, vol. 8296. International Society for Optics and Photonics, 2012, p. 82960O.
- [37] Q. Huynh-Thu and M. Ghanbari, “Scope of validity of psnr in image/video quality assessment,” *Electronics Letters*, vol. 44, no. 13, pp. 800–801, 2008.
- [38] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [39] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: A feature similarity index for image quality assessment,” *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [40] L. Wald, *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002.
- [41] F. Kruse, A. Lefkoff, and J. Dietz, “Expert system-based mineral mapping in northern death valley, california/nevada, using the airborne visible/infrared imaging spectrometer (aviris),” *Remote Sensing of Environment*, vol. 44, no. 2-3, pp. 309–336, 1993.
- [42] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

APPENDIX A

APPENDIX

In this section, we give detailed steps to solve the two theorems of \mathcal{Z}_p .

Lemma 1. [42] Let $\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2}$ be a given matrix, then the singular value decomposition (SVD) of a matrix \mathbf{Y} is defined as follows:

$$\mathbf{Y} = \mathbf{U} \Sigma \mathbf{V}^H, \Sigma = \text{diag}(\{\sigma_i\}). \quad (29)$$

The singular value shrinkage operator then obeys

$$\mathcal{D}_\tau(\mathbf{Y}) = \arg \min_{\text{rank}(\mathbf{X})} \tau \|\mathbf{X}\|_* + \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2, \quad (30)$$

where

$$\mathcal{D}_\tau(\mathbf{Y}) = \mathbf{U} \mathcal{D}_\delta(\Sigma) \mathbf{V}^H, \mathcal{D}_\tau(\Sigma) = \text{diag}\{\max((\sigma_i - \tau), 0)\}. \quad (31)$$

Theorem 1. For any positive threshold $\tau > 0$, $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, and frequency weighting w , the following problem:

$$\arg \min_{\mathcal{X}} \tau \|\mathcal{X}\|_{FW*} + \frac{1}{2} \|\mathcal{X} - \mathcal{Y}\|_F^2 \quad (32)$$

is given by the frequency-weighted tensor singular value thresholding

$$\mathcal{X}^* = \mathcal{F}W^{w,\tau}(\mathcal{Y}) = \mathcal{U} * \mathcal{S}^{w,\tau} * \mathcal{V}^H, \quad (33)$$

where $\mathcal{S}^{w,\tau} = \text{ifft}(\bar{\mathcal{S}}^{w,\tau}) = \text{ifft}(\max(\bar{\mathcal{S}} - w\tau), 0)$.

Proof. According to the properties of the tensor in the Fourier domain, the problem (32) is equivalent to

$$\begin{aligned} & \arg \min \tau \|\mathcal{X}\|_{FW*} + \frac{1}{2} \|\mathcal{X} - \mathcal{Y}\|_F^2 \\ \Leftrightarrow & \arg \min \tau \frac{1}{n_3} \sum_{k=1}^{n_3} w_k \|\bar{\mathcal{X}}^{(k)}\|_* + \frac{1}{2n_3} \|\bar{\mathcal{X}} - \bar{\mathcal{Y}}\|_F^2 \\ \Leftrightarrow & \arg \min \frac{1}{n_3} \sum_{k=1}^{n_3} \tau w_k \|\bar{\mathcal{X}}^{(k)}\|_* + \frac{1}{2} \|\bar{\mathcal{X}}^{(k)} - \bar{\mathcal{Y}}^{(k)}\|_F^2 \end{aligned} \quad (34)$$

Therefore, the problem in (32) can be divided into n_3 subproblems about problem (30). From Lemma 1, for k subproblems, its solution is

$$\bar{\mathcal{X}}^{(k)} = \mathcal{D}_{w_k \tau}(\bar{\mathcal{Y}}^{(k)}) = \mathbf{U}^{(k)} \max\left(\left(\mathbf{S}^{(k)} - w_k \tau\right), 0\right) \mathbf{V}^{(k)H}. \quad (35)$$

□

Lemma 2. [9] Let $0 < \tau$ and $0 < \varepsilon < \min\left(\sqrt{\tau}, \frac{\tau}{y}\right)$, the following problem:

$$\min_x f(x) = \tau \log(|x| + \varepsilon) + \frac{1}{2}(x - y)^2 \quad (36)$$

has a local minimal

$$\mathcal{D}_{\tau,\varepsilon}(y) = \begin{cases} 0 & \text{if } c_2 \leq 0 \\ \text{sign}(y) \left(\frac{c_1 + \sqrt{c_2}}{2}\right) & \text{if } c_2 > 0 \end{cases} \quad (37)$$

where $c_1 = |y| - \varepsilon$ and $c_2 = (c_1)^2 - 4(\tau - \varepsilon|y|)$.

Theorem 2. For any positive threshold $\tau > 0$, $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, and frequency weighting w , parameter $\varepsilon > 0$, the following problem:

$$\arg \min_{\mathcal{X}} \tau \|\mathcal{X}\|_{FW*, \text{Log}} + \frac{1}{2} \|\mathcal{X} - \mathcal{Y}\|_F^2 \quad (38)$$

is given by the double-weighted tensor singular value thresholding

$$\mathcal{X}^* = \mathcal{D}W^{w,\tau,\varepsilon}(\mathcal{Y}) = \mathcal{U} * \mathcal{S}^{w,\tau,\varepsilon} * \mathcal{V}^H, \quad (39)$$

where $\mathcal{Y} = \mathcal{U} * \mathcal{S} * \mathcal{V}^H$, $\mathcal{S}^{w,\tau,\varepsilon} = \text{ifft}\left(\left(\bar{\mathcal{S}}\right)^{w,\tau,\varepsilon}\right)$ and

$$\left(\bar{\mathcal{S}}\right)^{w,\tau,\varepsilon}(i, j, s) = \begin{cases} 0, & \text{if } c_2 \leq 0 \\ \text{sign}\left(\bar{\mathcal{S}}(i, j, s)\right) \left(\frac{c_1 + \sqrt{c_2}}{2}\right), & \text{if } c_2 > 0 \end{cases} \quad (40)$$

where $\bar{\mathcal{S}} = \text{fft}(\mathcal{S})$, $c_1 = |\bar{\mathcal{S}}(i, j, s)| - \varepsilon$, and $c_2 = c_1^2 - 4(w_k \tau - \varepsilon |\bar{\mathcal{S}}(i, j, s)|)$

Proof. Similar to the proof of Theorem 1, the problem (38) is equivalent to

$$\begin{aligned} & \arg \min \tau \|\mathcal{X}\|_{FW*, \text{Log}} + \frac{1}{2} \|\mathcal{X} - \mathcal{Y}\|_F^2 \\ \Leftrightarrow & \arg \min \tau \frac{1}{n_3} \sum_{k=1}^{n_3} w_k \log\left(\|\bar{\mathcal{X}}^{(k)}\|_*\right) + \frac{1}{2n_3} \|\bar{\mathcal{X}} - \bar{\mathcal{Y}}\|_F^2 \\ \Leftrightarrow & \arg \min \frac{1}{n_3} \sum_{k=1}^{n_3} \tau w_k \sum_{i=1}^{\min(n_1, n_2)} \log\left(\sigma_i\left(\bar{\mathcal{X}}^{(k)}\right) + \varepsilon\right) \\ & + \frac{1}{2} \|\bar{\mathcal{X}}^{(k)} - \bar{\mathcal{Y}}^{(k)}\|_F^2. \end{aligned} \quad (41)$$

Therefore, the problem in (38) can be divided into n_3 subproblems about problem (36). From Lemma 2, for k subproblems, its solution is

$$\bar{\mathbf{X}}^{(k)} = \mathcal{D}_{w_k \tau, \varepsilon}(\bar{\mathbf{Y}}^{(k)}) = \begin{cases} 0 & \text{if } c_2 \leq 0 \\ \text{sign}(\bar{\mathbf{Y}}^{(k)}) \left(\frac{c_1 + \sqrt{c_2}}{2} \right) & \text{if } c_2 > 0 \end{cases} .$$

(42)

□