Hyperspectral Image Denoising via L_0 Regularized Low-Rank Tucker Decomposition

Xin Tian , Kun Xie , *Member, IEEE*, and Hanling Zhang

Abstract—This article studies the mixed noise removal problem for hyperspectral images (HSIs), which often suffer from Gaussian noise and sparse noise. Conventional denoising models mainly employ the L_1-norm-based regularizers to remove sparse noise and ensure piecewise smoothness. However, the denoising performance is poor for highly structured images with severe noise since the L_1-norm overpenalizes large entries. To tackle this limitation, we propose a denoising model that combines tensor decomposition with two kinds of L_0-norm-based regularizers. First, we use low-rank Tucker decomposition with the Stiefel manifold to characterize the global correlation of HSIs. Then, we utilize the L_0-norm to leverage the intrinsic sparsity information of the corruption domain, thereby enhancing the effectiveness of sparse noise removal. Simultaneously, we introduce a weighted L_0-norm regularizer on the gradient of each pixel to promote the local spectral-spatial smoothness. To solve the proposed model, we design a hard-thresholding-based alternating direction method of multipliers algorithm. Instead of spending time to find a proper rank in advance, we adopt a rank-increasing scheme to dynamically adjust the tensor rank during the optimization procedure. In this way, our algorithm avoids the rank selection burden and improves computational efficiency. Finally, we test the proposed method on both synthetic and real datasets. Numerical results demonstrate its superiority, especially, the improvements of our method over the best-compared results up to 2.07 dB for mixed noise removal.

Index Terms—Low-rank, piecewise smooth, rank-increasing, sparse.

I. INTRODUCTION

DUE to the rapid advancement of digital imaging technology, hyperspectral images (HSIs) have become increasingly prevalent in real-life applications, such as medical diagnosis, military surveillance, food quality testing, etc.

However, HSIs are vulnerable to severe noise corruption, thus leading to a significant degradation in the image quality and limiting the precision of the subsequent works. Hence, separating the clean HSI from the raw observation corrupted by the noise is of acute and growing importance.

To date, various denoising techniques have been proposed for HSI restoration. These techniques can be primarily divided into two types, namely deep learning (DL)-based and model-based techniques. Existing DL-based methods [1], [2], [3], [4], [5], [6] have obtained excellent denoising results through learning non-linear mapping directly from the image data, but these methods lack interpretability, thus difficult to provide the reliability of performances. In contrast, the model-based methods are easily understandable. This kind of method mainly describes the prior knowledge of HSIs through mathematical formulas, so as to make use of the essential features. Considering this advantage, we mainly research the model-based method.

As the strong correlations and redundancy between the pixels in the spectral and spatial modes, HSIs typically exhibit a low-rank structure [7], [8]. To utilize the low-rank prior knowledge, tensor decomposition has become an effective tool in signal processing. Recently, several tensor decomposition methods have emerged for high-dimensional data analysis, such as CP decomposition (CPD) [9], Tucker decomposition (TKD) [10], tensor train decomposition (TTD) [11], and tensor ring decomposition (TRD) [12], etc. Among these types of decomposition, TKD can be seen as a multilinear extension of principal component analysis (PCA) [10], making it a valuable tool in various signal processing domains. Furthermore, the factors of the TKD are linked via a core tensor, which allows us to model more complex hidden data structures. To this end, we focus on the TKD in this article.

In real scenes, HSIs are inevitably affected by mixed noise, including Gaussian noise, impulse noise, deadlines, stripes, etc. Conventional TKD-based denoising methods mainly employ the L_1-norm minimization for non-Gaussian noise removal [13], and exhibit promising performance. Simultaneously, the visual image often exhibits local piecewise smoothness structure due to the presence of objects or edges. To concurrently preserve this edge information, an anisotropic spatial-spectral total variation (SSTV) regularizer, defined as the L_1-norm of gradient values, has been widely used for HSI restoration and shows reasonable performances [8], [14].

Although existing TKD-based methods combined with the L_1-norm-based regularizers have shown strong denoising ability [8], [15], [16], there still exist the following limitations:

1) *Fail to recover the image corrupted by severely sparse noise:* From a statistical perspective, solutions derived through L_1-norm minimization deviate significantly from both the prior and data acquisition models [17]. Therefore, L_1-norm minimization may not provide the optimal estimation performance for sparse noise removal, particularly when the sparse noise dominates. To overcome the shortcoming of L_1-norm, some nonconvex sparse regularizations such as minimax concave penalty (MCP) [18], [19], smoothly clipped absolute deviation (SCAD) [20], ϵ-penalty [21], and L_0-norm [22] have been proposed for the sparse noise removal, but these methods cannot take advantage of global correlations of HSIs as much as TKD-based methods.

2) *Fail to protect the intrinsic smoothness structure of the HSIs:* For a natural image, the large gradient magnitudes are closely associated with essential edges. However, the L_1-norm-based SSTV regularizer shrinks all gradient values equally, which makes those large gradient magnitudes overpenalized, thus leading to spatial oversmoothing. Although some improved TV regularizers [23], [24], [25] have been proposed, there is still room for improvement in the performance of these methods.

3) *Suffer from high sensitivity of rank selection:* For the TKD-based models, the Tucker rank should be predefined in advance. As is well known, determining the tensor rank is an NP-hard problem, and inappropriate ranks may result in unsatisfactory recovery results. Generally, we try to find a proper rank by tuning it as a hyperparameter manually, then fix this rank in the denoising algorithm, referred to as the rank-fixed scheme. However, we need to repeatedly execute the denoising algorithm multiple times to find a proper rank, which is time-consuming, and impractical in practice.

To overcome the above drawbacks, we propose a mixed noise removal model through the low-rank TKD with two kinds of L_0-norm-based regularizers. We summarize the main contributions as follows:

1) Under the low-rank TKD framework, we employ the L_0-norm for sparse noise removal, which alleviates the biased estimators caused by the L_1-norm.

2) We introduce the L_0-norm to directly control the number of nonzero gradients, which helps the low-rank Tucker model to capture the local spatial-spectral smoothness of the nature HSIs.

3) We adopt a rank-increasing scheme to dynamically adjust the tensor rank for the proposed TKD-based denoising algorithm, which avoids the rank selection burden, and reduces the computation time.

The rest of this article is organized as follows. Section II introduces basic notations and related works. Section III presents the proposed model with the corresponding optimization algorithm. Section IV demonstrates the results of extensive experiments and provides the ablation experiments of the proposed model. Finally, Section V concludes this article.

TABLE I
BASIC NOTATIONS AND THEIR DESCRIPTIONS

Notation	Descriptions
$\mathcal{X}, X, x$	Tensor, matrix, vector
$\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$	An Nth-order tensor
$\mathcal{X}_{i,j,k}$	Element (i, j, k) of a third-order tensor $\mathcal{X}$
$X_{i,j}$	Element (i, j) of a matrix X
x_i	The ith entry of a vector x
$< \mathcal{X}, \mathcal{Y} >$	Inner product of two tensors $\mathcal{X}$ and $\mathcal{Y}$
$\|\mathcal{X}\|_F = \sqrt{< \mathcal{X}, \mathcal{X} >}$	The Frobenius norm of a tensor $\mathcal{X}$
$\text{unfold}_n(\mathcal{X}) := X_{(n)}$	The mode-n unfolding of a tensor $\mathcal{X}$
$\text{fold}_n(X_{(n)}) := \mathcal{X}$	Refolding of the matrix into a tensor
$\times_n$	The mode-n matrix product of a tensor

II. PRELIMINARIES AND RELATED WORKS

A. Preliminaries

Tensor is an extension of vectors and matrices in high-dimensional space, its order is also referred to as the number of dimensions or modes. In this article, we adopt the notations in [9]. Some basic notations are presented in Table I.

Definition 1 (see [9] (n-Rank)): The n-rank of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ corresponds to the column rank of the mode-n unfolding $X_{(n)}$, which is denoted $\text{rank}_n(\mathcal{X})$. We say $\mathcal{X}$ is (approximately) low-rank if $X_{(n)}$ is (approximately) low-rank for all n.

Definition 2 (see [9] (TKD for the third-order tensor)): Suppose that $\mathcal{X} = (\mathcal{X}_{ijk}) \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. Let $U = (U_{ip}) \in \mathbb{R}^{I_1 \times P}$, $V = (V_{jq}) \in \mathbb{R}^{I_2 \times Q}$, $W = (W_{kl}) \in \mathbb{R}^{I_3 \times L}$. $\mathcal{C} = (\mathcal{C}_{pql}) \in \mathbb{R}^{P \times Q \times L}$. Here, I_1, I_2, I_3, P, Q, L are positive integers. If

$$\mathcal{X}_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{l=1}^{L} U_{ip} V_{jq} W_{kl} \mathcal{C}_{pql} \tag{1}$$

for $i \in [I_1], j \in [I_2], k \in [I_3]$, then $\mathcal{X}$ has a TKD $\mathcal{X} = \mathcal{C} \times_1 U \times_2 V \times_3 W$. The tensor $\mathcal{C}$ is called Tucker core, and the matrices U, V, W are called factor matrices of TKD. For a given Nth-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, we can easily find an exact TKD of rank $r = (r_1, r_2, \ldots, r_N)$, where $r_n = \text{rank}_n(\mathcal{X}), n = 1, \ldots, N$.

B. Related Works

In recent years, the tensor robust principal component analysis (TRPCA) framework [26], [27], [28] gained popularity due to its effectiveness in handling signal recovery. Within this framework, the degraded HSI $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is divided into two components, a low-rank component representing the clean HSI $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and a sparse corruption component $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. TRPCA framework can be mathematically formulated as follows:

$$\min_{\mathcal{X}, \mathcal{S}} \ M_1(\mathcal{X}) + \lambda M_2(\mathcal{S})$$

$$\text{s.t. } \|\mathcal{Y} - \mathcal{X} - \mathcal{S}\| \leqslant \xi \tag{2}$$

where $M_1(\mathcal{X})$ and $M_2(\mathcal{S})$ represent the prior knowledge of the noiseless HSI $\mathcal{X}$ and sparse noise $\mathcal{S}$, respectively. λ is a

positive regularization parameter, ξ is related to the variance of the nonsparse noise density.

According to the choice of low-rank regularizer M_1, framework (2) can be categorized into two types: 1) tensor decomposition-based methods, and 2) rank minimization-based methods. To avoid the high computational cost associated with large-scale singular value decomposition (SVD), tensor decomposition has been widely used in recent works, showing a strong ability to explore global information in visual data recovery [29], [30].

The TRPCA framework with low-rank tensor decomposition has achieved satisfactory results for noise removal. To further improve the denoising performance, L_1-norm and TV regularizer have been extensively incorporated into this framework. For instance, Wang et al. [8] embedded the L_1-norm and SSTV regularizer into the low-rank tensor decomposition for mixed noise removal; To exploit the shared group sparsity pattern of the HSIs, Chen et al. [15] proposed a weighted group sparsity-regularized low-rank tensor decomposition method by combining the L_1-norm and $L_{2,1}$-norm-based TV regularizer. Furthermore, Xu et al. [31] validated that the gradients of an HSI are heavy-tailed, and then proposed a hyper-Laplacian TV to replace the SSTV and $L_{2,1}$-norm-based TV. Mathematically speaking, both the non-Gaussian noise removal and the local smoothness prior are essentially sparse optimization problems in different domains, the convex L_1-norm or nonconvex regularizer is not the optimal choice to depict the sparsity [32].

Different from the aforementioned methods, in this article, we consider: 1) how to character the intrinsic sparsity of the corruption domain, 2) how to enhance the sparsity of the gradient image domain, thus promoting the local smoothness of HSIs, and 3) how to alleviate the burden of rank selection problem for low-rank TKD-based approaches.

III. HSI DENOISING VIA L_0 REGULARIZERED LOW-RANK TKD

In this section, we first introduce a TV-based denoising framework. Second, we formulate our model which joins low-rank TKD and two kinds of L_0-based regularizers. Third, to efficiently solve the proposed model, we develop a hard-thresholding-based alternating direction method of multipliers (ADMM) algorithm with the rank-increasing scheme. Meanwhile, we give the computational complexity analysis.

A. Total Variation-Based HSI Restoration Framework

As introduced before, the combination of low-rank tensor decomposition with TV regularization can achieve satisfactory restoration results for HSIs by simultaneously considering the global spatial-spectral correlation and local piece-wise smoothness. We formulate the TV-based HSI restoration framework as follows:

$$\min_{\mathcal{X},\mathcal{S}} M_1(\mathcal{X}) + \lambda M_2(\mathcal{S}) + \tau \parallel \bigtriangledown\mathcal{X} \parallel_{p,1}$$

$$\text{s.t.} \parallel \mathcal{Y} - \mathcal{X} - \mathcal{S} \parallel \leqslant \xi \qquad (3)$$

where λ and τ are the regularization parameters, p is a positive parameter, $\bigtriangledown = (\bigtriangledown_h, \bigtriangledown_v, \bigtriangledown_s)$ is the gradient operator along two spatial dimensions and one spectral dimension of an HSI, and $\parallel \bigtriangledown\mathcal{X} \parallel_{p,1}$ represents the TV regularizer used to explore the piecewise smooth structure of HSIs. Denote $\bigtriangledown\mathcal{X}(i,j,k)$ as the gradient operator at (i,j,k) of $\mathcal{X}$, then its components are defined as follows:

$$\begin{cases} \bigtriangledown_h \mathcal{X}(i,j,k) = \mathcal{X}_{i+1,j,k} - \mathcal{X}_{i,j,k} \\ \bigtriangledown_v \mathcal{X}(i,j,k) = \mathcal{X}_{i,j+1,k} - \mathcal{X}_{i,j,k} \\ \bigtriangledown_s \mathcal{X}(i,j,k) = \mathcal{X}_{i,j,k+1} - \mathcal{X}_{i,j,k}. \end{cases}$$

Next, we show how to derive formulations for the $M_1(\cdot)$, $M_2(\cdot)$, and $\parallel \bigtriangledown\mathcal{X} \parallel_{p,1}$.

1) The Selection of M_1: M_1 is the regularizer to reveal the low-rankness of clean HSI $\mathcal{X}$. We empirically interpret this property via showing the singular value curves of the unfolding matrices for an HSI. In Fig. 1, we can apparently see only a small fraction of singular values are greater than zero in all modes, which indicates the high-dimensional HSI essentially lies at a low-dimensional space.

Inspired by this property, we adopt the well-known TKD in *Definition 2* to model the global spatial—spectral correlations of an HSI. Meanwhile, we also consider the orthogonality of the factor matrices. Specifically, a rank-(r_1, r_2, r_3) TKD for a given third-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is expressed as

$$M_1 : \mathcal{X} = \mathcal{C} \times_1 \mathrm{U}_1 \times_2 \mathrm{U}_2 \times_3 \mathrm{U}_3, \mathrm{U}_i \in St(I_i, r_i), i = 1, 2, 3.$$

where $\mathcal{C}$ is the core tensor, $\mathrm{U}_i s$ are factor matrices, $St(I_i, r_i) = \{\mathrm{U}_i \in \mathbb{R}^{I_i \times r_i} : \mathrm{U}_i^T \mathrm{U}_i = I_{r_i}\}$ denotes the Stiefel manifold, i.e., the set of columnwise orthonormal matrices of size $I_i \times r_i$. On the one hand, the unconstrained TKD is in general not unique, imposing the orthogonality constraint on the factor matrices can assist in guaranteeing the uniqueness of the Tucker representation. On the other hand, the Tucker model with factor orthogonality constraint can be regarded as an extension of the SVD model, we can employ algorithms like HOSVD or HOOI to find solutions with high accuracy, thereby preserving the key features of HSIs [33].

2) Selection of M_2: The regularizer M_2 is used for sparse noise removal, it is essentially a combinatorial optimization problem with L_0 minimization, i.e., $M_2(\mathcal{S})$ can be modeled as

$$M_2(\mathcal{S}) := \|\mathcal{S}\|_0 \qquad (4)$$

where $\| \cdot \|_0$ represents the number of nonzero elements. However, L_0-norm is nonsmooth, which leads to an NP-hard problem. To overcome this problem, previous denoising works always used the convex L_1-norm as the relaxation to restrict this sparsity, i.e.,

$$M_2(\mathcal{S}) := \|\mathcal{S}\|_1 \qquad (5)$$

where $\| \cdot \|_1$ is defined as the sum of absolute values of all elements.

Actually, if the degraded data $\mathcal{Y}$ is not severely corrupted by sparse noise, the sparse fitting term should be less or null penalized, while enforcing those data that are more likely to be severely damaged. However, L_1-norm adds the absolute values of all elements together, which indicates that larger elements
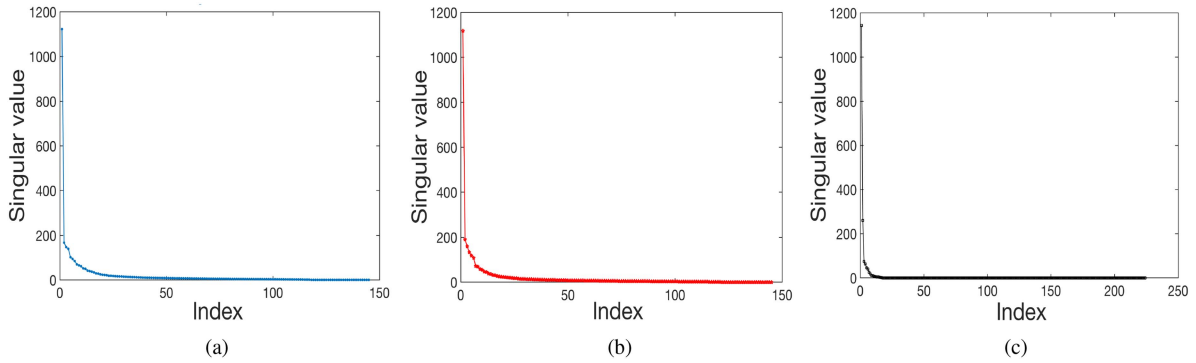
Fig. 1.   Singular values of the unfolding matrices. (a) Mode-1 unfolding. (b) Mode-2 unfolding. (c) Mode-3 unfolding.
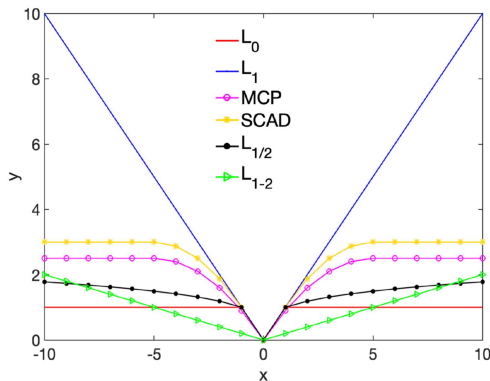

Fig. 2.   Illustration of convex and nonconvex functions.

are penalized more heavily than smaller ones, thus resulting in a biased estimator [32].

In order to obtain more desirable sparsity solutions, some nonconvex data fitting terms were discussed to approximate the L_0-norm. In Fig. 2, we plot the L_0-norm, L_1-norm, L_{1-2}-norm, and some nonconvex functions, including $L_{1/2}$-norm, MCP [18], [19], and SCAD [20] penalized likelihood functions for scalars. It is easy to see that the L_1-norm, L_{1-2}-norm, as well as the nonconvex functions, deviate significantly from 1 when the elements are greater than 1, thus can not provide the accurate sparse measurement.

To this end, we employ the L_0-norm directly as the sparsity measure for sparse noise removal, which makes our model utilize the intrinsic sparsity information.

3) Selection of $\| \bigtriangledown \mathcal{X} \|_{p,1}$: Due to the edge of the object, HSIs always exhibit piecewise smoothness [23], [34]. To verify this structure, we plot the distribution histogram of gradient values along with different modes of an HSI in Fig. 3. The histogram shows that most gradient values between adjacent pixels in both spatial and spectral domains are close to zero, which demonstrates the piecewise smoothness of the HSI. To characterize this prior information, the TV regularizer has been widely used in the computer vision field, which can be formulated as

$$\| \bigtriangledown \mathcal{X} \|_{p,1} := \sum_{i,j,k} (w_1 |\bigtriangledown_h \mathcal{X}(i,j,k)|^p + w_2 |\bigtriangledown_v \mathcal{X}(i,j,k)|^p$$

$$+ w_3 |\bigtriangledown_s \mathcal{X}(i,j,k)|^p)^{1/p}$$

where $w_i (i = 1, 2, 3)$ are the weights to control the smoothness strength.

When $p = 1$(or $p = 2$), $\| \bigtriangledown \mathcal{X} \|_{p,1}$ represents the well-known anisotropic (or isotropic) spatial-spectral TV (SSTV) [8], [19], [25], i.e.,

$$\| \bigtriangledown \mathcal{X} \|_{\text{SSTV}}^{\text{ani}} := \sum_{i,j,k} w_1 |\bigtriangledown_h \mathcal{X}(i,j,k)| + w_2 |\bigtriangledown_v \mathcal{X}(i,j,k)|$$

$$+ w_3 |\bigtriangledown_s \mathcal{X}(i,j,k)|$$

$$\| \bigtriangledown \mathcal{X} \|_{\text{SSTV}}^{\text{iso}} := \sum_{i,j,k} (w_1 |\bigtriangledown_h \mathcal{X}(i,j,k)|^2 + w_2 |\bigtriangledown_v \mathcal{X}(i,j,k)|^2$$

$$+ w_3 |\bigtriangledown_s \mathcal{X}(i,j,k)|^2)^{1/2}.$$

Xu et al. [31] pointed out that the gradient distributions of HSIs are heavy-tailed, and the anisotropic (isotropic) SSTV can not precisely describe this feature. To this end, they proposed a hyper-Laplacian spectral-spatial total variation (HTV)

$$\| \bigtriangledown \mathcal{X} \|_{\text{HTV}} := \sum_{i,j,k} (w_1 |\bigtriangledown_h \mathcal{X}(i,j,k)|^{1/2}$$

$$+ w_2 |\bigtriangledown_v \mathcal{X}(i,j,k)|^{1/2}$$

$$+ w_3 |\bigtriangledown_s \mathcal{X}(i,j,k)|^{1/2})^{1/2}$$

which is also called as $L_{1/2}$-norm based SSTV.

Subsequently, Zeng et al. developed a 3-D L_{1-2} SSTV [24] to globally represent the smoothness prior of HSIs

$$\| \bigtriangledown \mathcal{X} \|_{L_{1-2}\text{SSTV}} := \| \bigtriangledown \mathcal{X} \|_{\text{SSTV}}^{\text{ani}} - \alpha \| \bigtriangledown \mathcal{X} \|_{\text{SSTV}}^{\text{iso}}$$

where $\alpha \in [0, 1]$ is a regularization parameter. To further explore the shared sparse pattern of a difference image, Chen et al. [15] introduced a weighted $L_{2,1}$-norm based group sparsity TV

$$\| \bigtriangledown \mathcal{X} \|_{\text{GroupTV}} := \sum_{i,j} w_1 \| \bigtriangledown_h \mathcal{X}(i,j,:) \|_2$$

$$+ w_2 \| \bigtriangledown_v \mathcal{X}(i,j,:) \|_2.$$

Actually, the design of $\| \bigtriangledown \mathcal{X} \|_{p,1}$ can be interpreted as how to characterize the sparsity of the gradient image domain. Although the above $L_{1/2}$-norm, L_{1-2}-norm, and $L_{2,1}$-norm based SSTV
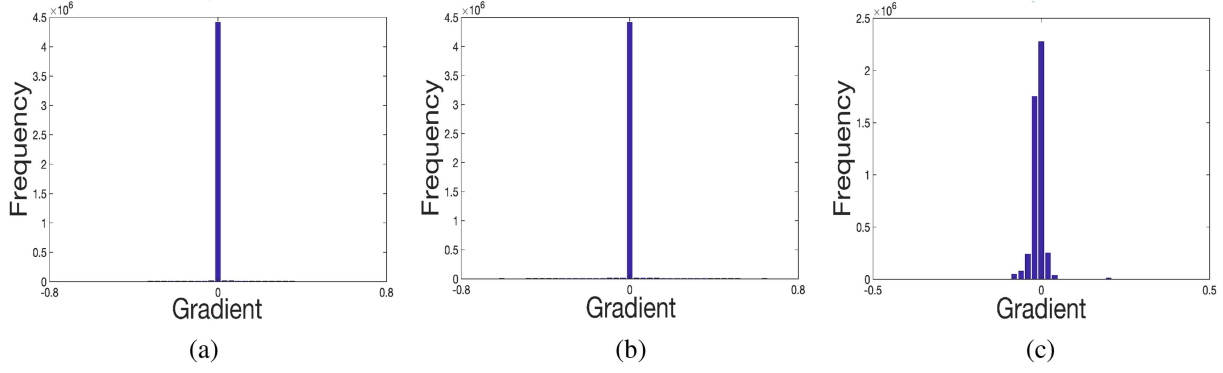
Fig. 3. Histogram of gradient values along each direction. (a) Horizontal. (b) Vertical. (c) Spectral.

regularizers can overcome the shortcoming of the anisotropic (isotropic) SSTV to some extent, they can not suppress artifacts effectively when dealing with the images with complexity structure. Considering the superiority of L_0-norm in promoting the piecewise smoothness [23], [29], [35], [36], we employ the following weighted spectral—spatial L_0 gradient regularization to describe the local smoothness of HSIs

$$\| \bigtriangledown \mathcal{X}\|_0 := \sum_{i,j,k} w_1 | \bigtriangledown_h \mathcal{X}(i,j,k)|_0 + w_2 | \bigtriangledown_v \mathcal{X}(i,j,k)|_0$$
$$+ w_3 | \bigtriangledown_s \mathcal{X}(i,j,k)|_0.$$

The superiority of the weighted spectral—spatial L_0 gradient regularization will be shown via extensive experiments in Section IV.

B. Proposed Model

In our work, we aim to exploit the intrinsic sparsity of corruption and gradient image domains simultaneously. Based on the above discussion, we propose a denoising model via L_0 regularized low-rank Tucker decomposition (L_0RLRTD) for HSI mainly with Gaussian noise and sparse noise

$$\underset{\mathcal{X},\mathcal{S},\mathcal{N}}{\text{argmin}} \ \tau \| \bigtriangledown \mathcal{X}\|_0 + \lambda \|\mathcal{S}\|_0 + \mu \| \mathcal{N}\|_F^2$$

$$\text{s.t.} \ \mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N}$$
$$\mathcal{X} = \mathcal{C} \times_1 \text{U}_1 \times_2 \text{U}_2 \times_3 \text{U}_3$$
$$\text{U}_i \in St(I_i, r_i), i = 1, 2, 3. \quad (6)$$

where τ, λ, and μ are the regularization parameters to control the tradeoff between the sparsity and smoothness, $\mathcal{N}$ represents the Gaussian noise. It is obvious that model (6) is nonsmooth and nonconvex because of the TKD and two L_0-norm-based regularizers. Compared to the methods using L_1-norm-based regularizers, the proposed L_0RLRTD has the following distinctive advantages:

1) In the corruption domain, the L_0-norm regularizer $\|\mathcal{S}\|_0$ could characterize the intrinsic quality of the sparse noise $\mathcal{S}$ such as impulse noise, stripes, and deadlines, thus it can facilitate exact recovery of sparse signal with higher probability.

2) In the gradient image domain, minimizing the L_0-norm-based gradient operator $\| \bigtriangledown \mathcal{X}\|_0$ not only protects the piecewise smoothness, but also maintains sharp edges, thus more useful edge information can be better preserved.

Next, we will introduce the solving process of the proposed L_0RLRTD (6) in detail.

C. Optimization Procedure

In this section, we introduce how to solve the proposed L_0RLRTD model (6) by the ADMM [37] algorithm with a hard-thresholding operator (short for ADMM-HT).

1) Hard-Thresholding Operator: Before solving the problem (6), we introduce the hard-thresholding operator to tackle the following nonsmooth nonconvex problem

$$\min_{x \in \mathbb{R}^n} \ f(x) = \eta \| x \|_0 + \frac{1}{2} \| x - u \|_F^2 \quad (7)$$

where $\eta > 0$ represents a regularization parameter, $x = [x_1, x_2, \ldots, x_n]$ is a variable, $u = [u_1, u_2, \ldots, u_n]$ is a given vector, and $f(x)$ denotes the objective function. Next, we show that the solution of (7) can be given by a simple explicit formula.

It is important to note that the minimization functional is separable and, therefore, we can minimize with respect to x_i separately. There are two possible cases for each x_i.

Case 1: $x_i \neq 0$. The objective functional value of (7) related to the x_i is $f(x_i) = \eta + \frac{1}{2}(x_i - u_i)^2$, and the minimum value is η at $x_i = u_i$.

Case 2: $x_i = 0$. The objective functional $f(x_i) = \frac{1}{2}u_i^2$.

Then, we can easily derive the solution of problem (7), and denoted as Hard-Thresholding operator [29], [38], [39]

$$\text{prox}_{\eta\|\cdot\|_0}(u) = \left\{ x \in \mathbb{R}^n | x_i = \left\{ \begin{array}{ll} u_i & \text{if } |u_i| > \sqrt{2\eta}, \\ \{0, u_i\} & \text{if } |u_i| = \sqrt{2\eta}, \\ 0 & \text{otherwise}, \end{array} \right. \right\}.$$

2) Optimization Procedure: First, we introduce several auxiliary tensors and derive the subsequent equivalent formulation

$$\underset{\mathcal{C}, \text{U}_i, \mathcal{X}, \mathcal{F}, \mathcal{Z}, \mathcal{S}, \mathcal{N}}{\text{argmin}} \ \tau \| \mathcal{F} \|_0 + \lambda \|\mathcal{S}\|_0 + \mu \| \mathcal{N}\|_F^2$$

$$\text{s.t.} \ \mathcal{Y} = \mathcal{X} + \mathcal{S} + \mathcal{N}$$

$$\mathcal{X} = \mathcal{Z}, \bigtriangledown \mathcal{Z} = \mathcal{F}$$
$$\mathcal{X} = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3$$
$$U_i \in St(I_i, r_i), i = 1, 2, 3. \tag{8}$$

The augmented Lagrangian function related to (8) is

$$\mathcal{L}_\beta(\mathcal{X}, \mathcal{F}, \mathcal{Z}, \mathcal{S}, \mathcal{N}, \Gamma_1, \Gamma_2, \Gamma_3) = \tau \parallel \mathcal{F} \parallel_0 + \lambda \|\mathcal{S}\|_0$$
$$+ \mu \parallel \mathcal{N} \parallel_F^2 + \langle \Gamma_1, \mathcal{Y} - \mathcal{X} - \mathcal{S} - \mathcal{N} \rangle + \langle \Gamma_2, \mathcal{X} - \mathcal{Z} \rangle$$
$$+ \langle \Gamma_3, \bigtriangledown \mathcal{Z} - \mathcal{F} \rangle + \frac{\beta}{2}(\parallel \mathcal{Y} - \mathcal{X} - \mathcal{S} - \mathcal{N} \parallel_F^2$$
$$+ \parallel \mathcal{X} - \mathcal{Z} \parallel_F^2 + \parallel \bigtriangledown \mathcal{Z} - \mathcal{F} \parallel_F^2). \tag{9}$$

under the constraints $\mathcal{X} = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3$, $U_i \in St(I_i, r_i)$. $\Gamma_i \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ are the Lagrange multipliers, $\beta > 0$ is the penalty parameter. By applying the ADMM, all subproblems of the proposed algorithm can be very efficiently solved.

Given $\mathcal{Z}^0, \mathcal{S}^0, \mathcal{N}^0$, and $(\Gamma_1^0, \Gamma_2^0, \Gamma_3^0)$, the ADMM-HT iterates as

$$\begin{cases} (\mathcal{X}^{k+1}, \mathcal{F}^{k+1}) & (10) \\ = \underset{\mathcal{X}, \mathcal{F}}{\operatorname{argmin}} \{\mathcal{L}_\beta(\mathcal{X}, \mathcal{F}, \mathcal{Z}^k, \mathcal{S}^k, \mathcal{N}^k, \Gamma_1^k, \Gamma_2^k, \Gamma_3^k) | \mathcal{X} \in \mathcal{T}\} \\ (\mathcal{Z}^{k+1}, \mathcal{S}^{k+1}) & (11) \\ = \underset{\mathcal{Z}, \mathcal{S}}{\operatorname{argmin}} \{\mathcal{L}_\beta(\mathcal{X}^{k+1}, \mathcal{F}^{k+1}, \mathcal{Z}, \mathcal{S}, \mathcal{N}^k, \Gamma_1^k, \Gamma_2^k, \Gamma_3^k)\} \\ \mathcal{N}^{k+1} & (12) \\ = \underset{\mathcal{N}}{\operatorname{argmin}} \{\mathcal{L}_\beta(\mathcal{X}^{k+1}, \mathcal{F}^{k+1}, \mathcal{Z}^{k+1}, \mathcal{S}^{k+1}, \mathcal{N}, \Gamma_1^k, \Gamma_2^k, \Gamma_3^k)\} \\ \begin{pmatrix} \Gamma_1^{k+1} \\ \Gamma_2^{k+1} \\ \Gamma_3^{k+1} \end{pmatrix} = \begin{pmatrix} \Gamma_1^k + \beta(\mathcal{Y} - \mathcal{X}^{k+1} - \mathcal{S}^{k+1} - \mathcal{N}^{k+1}) \\ \Gamma_2^k + \beta(\mathcal{X}^{k+1} - \mathcal{Z}^{k+1}) \\ \Gamma_3^k + \beta(\bigtriangledown Z^{k+1} \mathcal{F}^{k+1}) \end{pmatrix} & (13) \end{cases}$$

where $\mathcal{T} = \{\mathcal{X} | \mathcal{X} = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3, U_i \in St(I_i, r_i), i = 1, 2, 3\}$. Before ending this section, we further discuss the specific solution of the subproblems in (10)–(12) as follows.

1) *Update* $(\mathcal{C}, U_i, \mathcal{X}, \mathcal{F})$: Extracting all terms containing $\mathcal{X}$ from the augmented Lagrangian function (9), we can obtain

$$\underset{\mathcal{X}, \mathcal{C}, U_i}{\operatorname{argmin}} \langle \Gamma_1^k, \mathcal{Y} - \mathcal{X} - \mathcal{S}^k - \mathcal{N}^k \rangle + \langle \Gamma_2^k, \mathcal{X} - \mathcal{Z}^k \rangle$$
$$+ \frac{\beta}{2}(\parallel \mathcal{Y} - \mathcal{X} - \mathcal{S}^k - \mathcal{N}^k \parallel_F^2 + \parallel \mathcal{X} - \mathcal{Z}^k \parallel_F^2)$$
$$\text{s.t.} \quad \mathcal{X} = \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3, U_i \in St(I_i, r_i).$$

Then ignoring the constant term, we need to solve the following equivalent problem:

$$\underset{\mathcal{C}, U_i, U_i \in St(I_i, r_i)}{\operatorname{argmin}} \beta \left\| \mathcal{C} \times_1 U_1 \times_2 U_2 \times_3 U_3 \right.$$
$$\left. - \frac{1}{2}\left(\mathcal{Y} - \mathcal{S}^k - \mathcal{N}^k + \mathcal{Z}^k + \frac{\Gamma_1^k}{\beta} - \frac{\Gamma_2^k}{\beta}\right) \right\|_F^2.$$

We can obtain $\mathcal{C}^{k+1}$ and U_i^{k+1} by the high-order orthogonal iteration (HOOI) algorithm, then update $\mathcal{X}$ as follows:

$$\mathcal{X}^{k+1} = \mathcal{C}^{k+1} \times_1 U_1^{k+1} \times_2 U_2^{k+1} \times_3 U_3^{k+1}. \tag{14}$$

The augmented Lagrangian functions w.r.t. $\mathcal{F}$ is given by

$$\underset{\mathcal{F}}{\operatorname{argmin}} \ \tau \|\mathcal{F}\|_0 + \frac{\beta}{2} \left\| \bigtriangledown \mathcal{Z}^k - \mathcal{F} + \frac{\Gamma_3^k}{\beta} \right\|_F^2$$

we can update $\mathcal{F}$ as

$$\mathcal{F}^{k+1} = \operatorname{prox}_{\frac{\tau}{\beta} \|\cdot\|_0} \left(\bigtriangledown \mathcal{Z}^k + \frac{\Gamma_3^k}{\beta} \right). \tag{15}$$

2) *Update* $(\mathcal{Z}, \mathcal{S})$: Extracting all terms containing $\mathcal{Z}$ from (9), we can obtain the subproblem w.r.t. $\mathcal{Z}$

$$\underset{\mathcal{Z}}{\operatorname{argmin}} \ \langle \Gamma_2^k, \mathcal{X}^{k+1} - \mathcal{Z} \rangle + \langle \Gamma_3^k, \bigtriangledown \mathcal{Z} - \mathcal{F}^{k+1} \rangle$$
$$+ \frac{\beta}{2}(\parallel \mathcal{X}^{k+1} - \mathcal{Z} \parallel_F^2 + \parallel \bigtriangledown \mathcal{Z} - \mathcal{F}^{k+1} \parallel_F^2).$$

Via the optimality condition, $\mathcal{Z}$ can be obtained by solving the following linear system:

$$(\beta I + \beta \bigtriangledown^* \bigtriangledown)\mathcal{Z} = \beta \mathcal{X}^{k+1} + \beta \bigtriangledown^* (\mathcal{F}^{k+1}) + \Gamma_2^k - \bigtriangledown^* \Gamma_3^k. \tag{16}$$

To improve the computational efficiency, we utilize conjugate gradient (CG) method to solve the large-scale problem (16).

Similarly, the $\mathcal{S}$-update is given by

$$\underset{\mathcal{S}}{\operatorname{argmin}} \ \lambda \|\mathcal{S}\|_0 + \frac{\beta}{2} \left\| \mathcal{Y} - \mathcal{S} - \mathcal{X}^{k+1} - \mathcal{N}^k + \frac{\Gamma_1^k}{\beta} \right\|_F^2. \tag{17}$$

Using the hard-thresholding operator, the problem (17) can be tackled by

$$\mathcal{S}^{k+1} = \operatorname{prox}_{\frac{\lambda}{\beta} \|\cdot\|_0} \left(\mathcal{Y} - \mathcal{X}^{k+1} - \mathcal{N}^k + \frac{\Gamma_1^k}{\beta} \right) \tag{18}$$

3) *Update* $\mathcal{N}$: By keeping the other variables fixed, $\mathcal{N}$-update can be formulated as follows:

$$\underset{\mathcal{N}}{\operatorname{argmin}} \ \mu \parallel \mathcal{N} \parallel_F^2 + \frac{\beta}{2} \left\| \mathcal{Y} - \mathcal{X}^{k+1} - \mathcal{S}^{k+1} - \mathcal{N} + \frac{\Gamma_1^k}{\beta} \right\|_F^2.$$

Setting the derivative of the above formula with respect to variation $\mathcal{N}$ and making it equal to zero, we obtain the closed-form solution as

$$\mathcal{N}^{k+1} = \frac{1}{2\mu + \beta}(\Gamma_1^k + \beta \mathcal{Y} - \beta \mathcal{X}^{k+1} - \beta \mathcal{S}^{k+1}). \tag{19}$$

D. Rank-Increasing Scheme

Notably, the HOOI algorithm used to solve (14) needs to know the rank, which is often unknown for an incomplete tensor. Usually, we manually tune the tensor rank as a hyper-parameter, but this tactic is not appropriate for the TKD-based approaches, because: 1) the Tucker rank of an Nth-order tensor is a vector with N components, and each component may differ. As the tensor dimension order increases, there are numerous possibilities for the combination of rank components, making it infeasible to determine the proper rank manually, thus cannot guarantee the recovery accuracy; 2) the computational complexity of the HOOI algorithm is exponentially related to the size of the tensor rank [40], which means that the denoising algorithm is time-consuming to deal with the tensor with a higher rank.

To overcome these challenges, we adopt a rank-increasing scheme to dynamically refine the tensor rank $r = (r_1, r_2, r_3)$ at each iteration step. This scheme starts with an underestimated rank $r^0 = (r_1^0, r_2^0, r_3^0)$, then the corresponding rank r_i increases to $\min\{r_i + \triangle r, r_i^{\max}\}$ at each iteration if

$$\left| 1 - \frac{\|\mathcal{X}^{k+1}\|_F}{\|\mathcal{X}^k\|_F} \right| \geqslant 1e - 5 \tag{20}$$

where $r^{\max} = (r_1^{\max}, r_2^{\max}, r_3^{\max})$ is a given upper bound of the Tucker rank, $\triangle r$ represents the step size, which is a positive integer. The criterion (20) ensures that when the recovery results of two successive steps change relatively small, i.e., the algorithm reaches a local or global optimum, then the rank remains unchanged at the current iteration.

In the subsequent numerical experiments, we set $r_i^0 = \text{round}(20\% I_i)$, and $r_i^{\max} = \text{round}(90\% I_i)$ for $i = 1, 2$, where round (a) means to select the nearest integer of a. Since the rank along spectral dimension of an HSI is much smaller than its dimension [7], we use the HSI subspace estimation method (e.g., HySime [41]) to estimate the rank for spectral mode.

Compared with the rank-fixed scheme, the advantages of the rank-increasing scheme are twofold: 1) it avoids the rank selection burden, that is, we need not to tune the rank as a parameter by running the denoising algorithm several times, which greatly reduces the workload; 2) even though we can find the proper rank in advance, the computational efficiency of the denoising algorithm with a fixed rank is lower than the rank-increasing scheme. Because the rank-increasing scheme starts from a relatively smaller rank, thus reduces the computation time for the corresponding tensor decomposition. We will validate these advantages in Section IV-D.

E. Algorithm and Computational Complexity Analysis

Summarizing the aforementioned discussion, we display the specific algorithm in Algorithm 1.

Except for $\mathcal{X}$ and $\mathcal{Z}$ solvers need via iterative algorithms, the other subproblems in Algorithm 1 can be solved through simple algebraic operations of addition and subtraction, which makes the process relatively fast. To simplify the analysis, we let $\hat{r} = r_1 = r_2 = r_3$, the specific analysis is as follows:

1) As for $\mathcal{X}, \mathcal{C}, \mathbf{U}_i$, the total computational complexity of HOOI is $O((I_1 I_2 I_3 \hat{r} + (I_1 + I_2 + I_3)\hat{r}^4 + \hat{r}^6)t)$[40], where t is the number of iterations in HOOI.

2) As for $\mathcal{Z}$, the cost is $O(I_1 I_2 I_3 h)$ [29], where h is the number of iterations in CG.

Thus, the overall computational complexity of each iteration in the proposed Algorithm 1 is bounded by $O((I_1 I_2 I_3 \hat{r} + (I_1 + I_2 + I_3)\hat{r}^4 + \hat{r}^6)t + I_1 I_2 I_3 h)$.

IV. NUMERICAL RESULTS

In this section, we evaluate the performance of the proposed L_0RLRTD model on both synthetic and real datasets. Specifically, the original synthetic HSI data can be viewed as clean data, and we manually add different types of noise to simulate real noise. Unlike the synthetic datasets, the noise type and intensity of the real HSIs are unknown, as a result, recovering clean HSIs

---

Algorithm 1: ADMM-HT With the Rank-Increasing Scheme.

Input: The observed noisy image $\mathcal{Y}$, the regularized parameters τ, λ, μ, the weight w_i, the initial rank $r^0 = (r_1^0, r_2^0, r_3^0)$, the maximum rank $r^{\max} = (r_1^{\max}, r_2^{\max}, r_3^{\max})$, the step size $\triangle r$, and the penalty parameter β

Output: The restored image $\mathcal{X}$

1: **Initialization:** $\mathcal{X}^0 = \mathcal{F}^0 = \mathcal{Z}^0 = \mathcal{S}^0 = \mathcal{N}^0 = \text{zeros(size}(\mathcal{Y})), \Gamma_1^0 = \Gamma_2^0 = \Gamma_3^0 = \text{zeros(size}(\mathcal{Y})), k = 0$

2: **Repeat**

3:     • Update $\mathcal{X}^{k+1}, \mathcal{F}^{k+1}, \mathcal{Z}^{k+1}, \mathcal{S}^{k+1}, \mathcal{N}^{k+1}$ via (14)–(19)

4:     • Update the multipliers $\Gamma_1^{k+1}, \Gamma_2^{k+1}, \Gamma_3^{k+1}$ via (13)

5:     • Update the rank $r^{k+1} = (r_1^{k+1}, r_2^{k+1}, r_3^{k+1})$ via the rank-increasing scheme in Section III-D

6:     • $k := k + 1$

7: **Until** $\frac{\|\mathcal{X}^{k+1} - \mathcal{X}^k\|_F}{\|\mathcal{X}^k\|_F} < 1e - 5$ or $k > 200$

---

from such data is an extremely challenging task since there is no ground truth data as a reference.

A. Experiment Setup

1) Comparison Methods: As a comparison, we select nine representative denoising methods in the experiments, including a fast and flexible denoising convolutional neural network (FFDNet) [1], a bandwise asymmetric Laplacian noise-based matrix factorization (BALMF) model [42], a three-directional log-based tensor nuclear norm (3DLogTNN) method [30], an improving sparse noise removal via L_0-Norm optimization (L_0-PLRTV) [22], a TV regularized low-rank tensor decomposition (LRTDTV) method [8], an extended model of LRTDTV with a weighted group sparsity regularizer (LRTDGS) [15], a low-rank tensor factorization with hyper-Laplacian total variation prior (LTHTV) [31], a L_{1-2}SSTV regularized local low-rank tensor recovery model (TLR-L_{1-2}SSTV) [24], and a L_0 gradient regularized low-rank tensor factorization (LRTFL_0) method [29]. We choose the best parameters for each algorithm based on the tuning suggestions given in each paper to achieve optimal performance.

2) Parameter Setting: In Algorithm 1, some parameters including τ, λ, μ, w_i, and $\triangle r$ need to be carefully identified. Typically, we tune one of parameters at a time while the other parameters are fixed, and repeat this process until finding a set of desirable parameters. To assess the impact of these parameters under different noise cases, we choose the synthetic Washington DC (WDC) Mall dataset as our experimental object and mean peak signal-to-noise ratio (MPSNR) as the evaluation metric (note that the noise details of four cases can be found in Section IV-B.).

1) The regularization parameters τ, λ, and μ: In Fig. 4(a)–(c), we show the MPSNR curves of four cases with respect to the τ, λ, and μ selected different values. The trends of
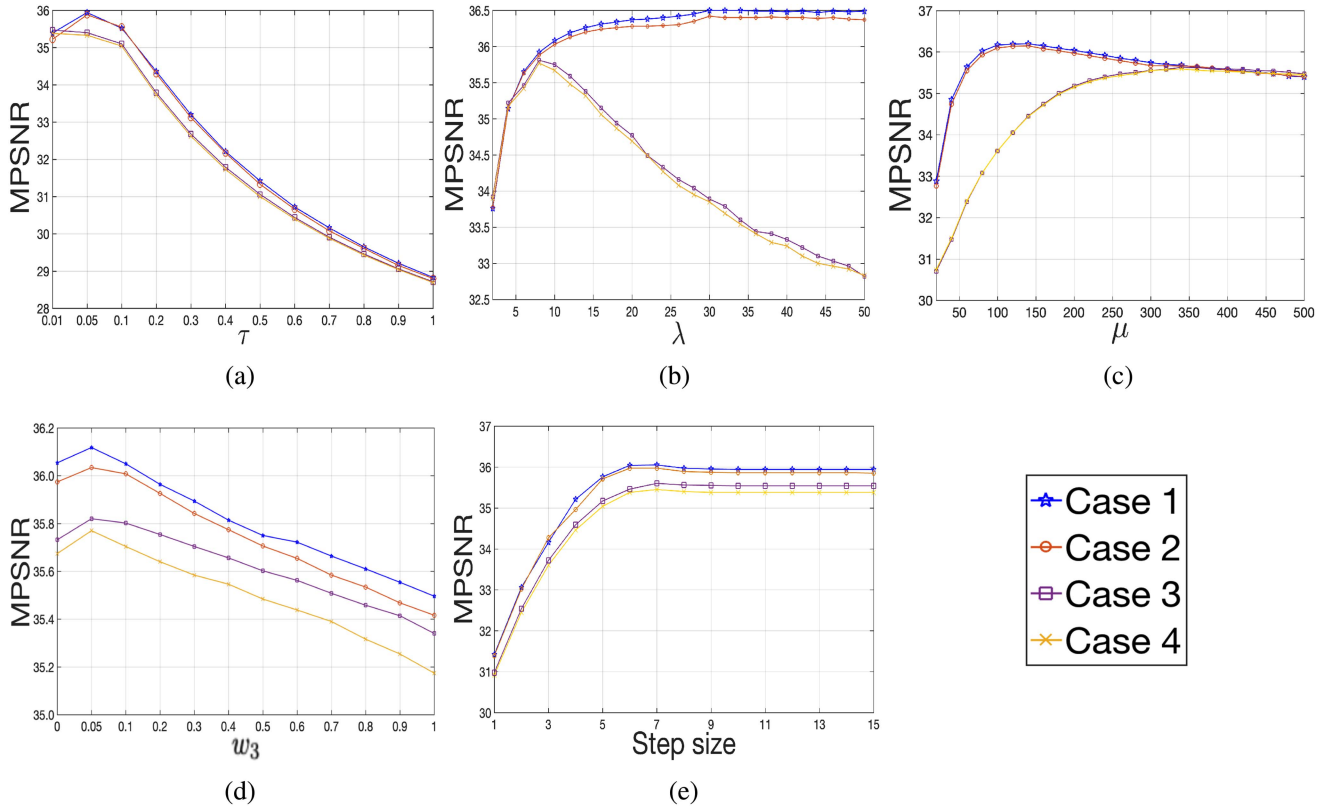
Fig. 4. Sensitivity of parameters: (a) τ, (b) λ, (c) μ, (d) w_3, (e) $\triangle r$.

these curves generally exhibit a monotonically increasing pattern initially, reaching a peak, and then either declining or stabilizing. To balance the performance and robustness, we set $\tau = 0.05$, $\lambda \in \{8, 30\}$, and $\mu = 300$ in the subsequent experiments.

2) The weight of spectral dimensional w_3: The relationship of MPSNR with w_3 is shown in Fig. 4(d). We can see that the best results are achieved with $w_3 = 0.05$ in all the experiments. Based on these results, we set $w_3 = 0.05$ in all experiments.

3) The step size $\triangle r$: The change curves of MPSNR value with different step sizes $\triangle r$ are shown in Fig. 4(e). As we can see, the peak values of all experiments are taken at $\triangle r = 6$, and then the MPSNR curves tend to stabilize when $\triangle r > 6$. Thus, we set $\triangle r = 6$ in all experiments.

Moreover, the selection of r^0 and $r^{\max}$ have been suggested in the Section III-D; the penalty parameter β is initialized to a small value 0.01, then increase it by a constant ratio $\rho \in [1.1, 1.5]$. It is worth noting that the optimal parameters for each dataset may be different, we try best to fix them as possible.

B. Experiments on Synthetic Data

To simulate noise HSI data, we introduce four types of noise to the original synthetic HSI data. In this section, we show the good performance of the proposed L_0RLRTD model on the three synthetic HSI datasets. These datasets include

1) *Indian Pines (IndianP):* The synthetic data was generated using the ground truth of the Indian Pines dataset[1], and the spectral signatures were extracted from the USGS digital spectral library.[2] This synthetic HSI data can be considered as clean data, which is the size of $145 \times 145 \times 224$. Some parameters for IndianP dataset are suggested as: $r^0 = (30, 30, 10)$, $r^{\max} = (130, 130, 10)$.

2) *WDC Mall:* This dataset originally comprises 1208×307 pixels and includes 191 spectral bands, which was acquired through the hyperspectral digital imagery collection experiment (HYDICE) sensor. To reduce the computational load, we only extract parts that contain the most details, the final dimension is $256 \times 256 \times 191$. Some parameters for WDC are suggested as: $r^0 = (51, 51, 5)$, $r^{\max} = (230, 230, 5)$.

3) *NC16:* This dataset[3] was collected in the Yellow River Delta National Nature Reserve in Shandong Province, China, which contains 1060×976 pixels and 270 spectral bands. To reduce the computational load, a subimage with size $300 \times 180 \times 270$ is used in our experiments. Some parameters are suggested as: $r^0 = (60, 36, 3)$, $r^{\max} = (270, 162, 3)$.

---

[1][Online]. Available: https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html

[2][Online]. Available: http://speclab.cr.usgs.gov/spectral.lib06

[3][Online]. Available: https://drive.google.com/drive/folders/1HMTIodCUcXgMIaijOco_wC8uK
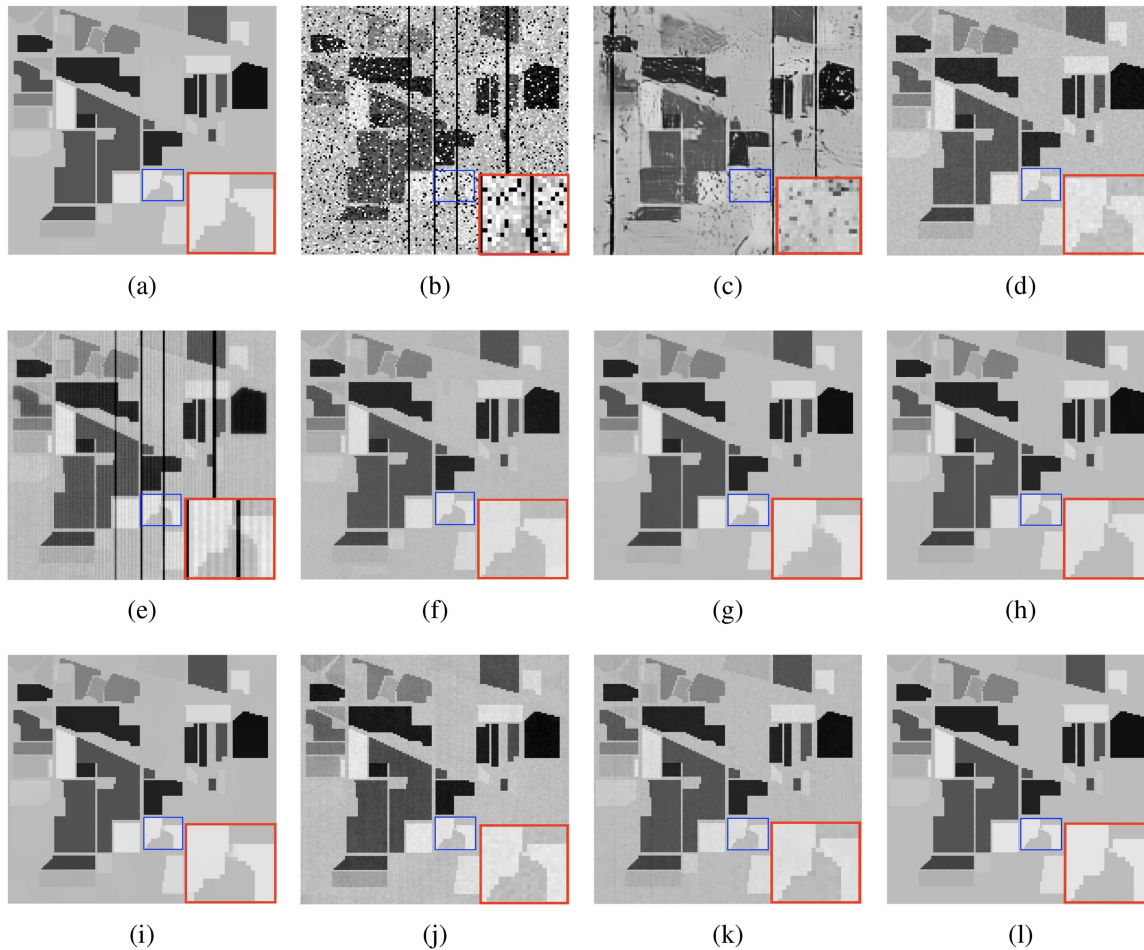
Fig. 5.   Denoised results by all the compared methods: (a) Original image, (b) the simulated noise image of Case 4, (c) FFDNet, (d) BALMF, (e) 3DlogTNN, (f) L_0-PLRTV, (g) LRTDTV, (h) LRTDGS, (i) LTHTV, (j) TLR-L_{1-2}SSTV, (k) LRTFL_0, (l) proposed.

Four types of noise are added to the synthetic datasets, specifically as follows:

Case 1: We add zero-mean Gaussian noise with the same distribution to each band, where the variance of the Gaussian noise is 0.1.

Case 2: The Gaussian noise added to each band is same as Case 1. Moreover, we introduce some stripes with the randomly selected number ranging from 3 to 10 and widths ranging from 1 to 3 in bands 91 to 130.

Case 3: We add the same zero-mean Gaussian noise distribution and percentage of impulse noise to each band. Specifically, the variance of the Gaussian noise is 0.075, and the percentage of impulse noise is 0.15. Additionally, we introduce the deadlines same as Case 2.

Case 4: Similar to Case 3, Gaussian noise, impulse noise, and deadlines are added for each band. Moreover, we introduce some stripes from band 161 to band 190, with the number of stripes being randomly selected from 20 to 40.

Next, we evaluate the HSI restoration performances of all comparison approaches from two perspectives, that is, visual and quantitative comparisons.

1) Visual Quality Comparison: We select a representative band of IndianP to compare the visual quality of the results

recovered by all methods. Fig. 5 displays the visual results of band 100, which is corrupted by mixed noise in Case 4. To facilitate visual comparison, a blue box is marked for the same subregion in each subfigure, and the corresponding subregion is further magnified within a red box.

From Fig. 5, it is easy to make several conclusions. First, FFDNet and 3DlogTNN are poorly used to remove the noise with obvious structure, such as deadlines. Second, BALMF, L_0-PLRTV, TLR-L_{1-2}SSTV, and LRTFL_0 can achieve superior performance for deadlines removal, but cannot completely eliminate the other noise. Third, while LRTDTV, LRTDGS, and LTHTV remove all kinds of noise better, they also result in the loss of some details. From the enlarged box, we can observe that the proposed method is capable of better removing the Gaussian noise and sparse noise including impulse noise, stripes, deadlines, while effectively preserving the essential details of the underlying HSI. This manifests that L_0-norm is more effective in exploring the sparsity of corruption and gradient image domains.

2) Quantitative Comparison: In simulated experiments, the denoising results of HSIs can be objectively evaluated by the MPSNR, mean structural similarity index (MSSIM), and Erreur Relative Globale Adimensionnelle de synthèse (ERGAS). Their definitions can be found in [8], where MPSNR and MSSIM

TABLE II
QUANTITATIVE EVALUATION ON THE SYNTHETIC INDIAN PINES

Case	Case 1			Case 2			Case 3			Case 4		
Method	MPSNR	MSSIM	ERGAS	MPSNR	MSSIM	ERGAS	MPSNR	MSSIM	ERGAS	MPSNR	MSSIM	ERGAS
Noisy	20.00	0.3673	233.78	19.40	0.3599	254.90	12.92	0.1752	529.35	12.87	0.1736	533.05
FFDNet [1]	28.57	0.8801	87.68	26.91	0.8511	130.22	21.22	0.5604	216.46	21.01	0.5430	220.76
BALMF [42]	32.92	0.8487	55.83	32.52	0.8235	57.87	32.90	0.8485	55.91	32.73	0.8434	57.38
3DlogTNN [30]	33.86	0.9183	49.13	31.24	0.8769	104.47	31.57	0.8840	100.55	30.15	0.8392	121.58
L_0-PLRTV [22]	38.68	0.9714	28.29	38.64	0.9713	28.39	38.96	0.9829	28.37	35.11	0.8938	99.25
LRTDTV [8]	40.90	0.9898	22.55	39.86	0.9878	28.56	40.85	0.9908	22.73	40.35	0.9899	24.09
LRTDGS [15]	42.90	0.9960	18.13	40.93	0.9920	28.88	40.99	0.9929	27.25	40.73	0.9930	27.19
LTHTV [31]	42.81	0.9962	19.21	42.34	0.9959	19.87	41.01	0.9920	33.4	42.26	0.9961	**20.04**
L_{1-2}SSTV [24]	33.39	0.8920	52.17	33.00	0.8823	54.72	31.35	0.8770	66.35	31.13	0.8700	68.06
LRTFL_0 [29]	38.30	0.9587	30.80	37.23	0.9563	46.49	37.61	0.9592	44.47	37.11	0.9532	44.19
Proposed	**44.45**	**0.9973**	**15.52**	**44.34**	**0.9973**	**15.68**	**43.08**	**0.9972**	**18.61**	**42.84**	**0.9970**	20.19

TABLE III
QUANTITATIVE EVALUATION ON THE SYNTHETIC WDC

Case	Case 1			Case 2			Case 3			Case 4		
Method	MPSNR	MSSIM	ERGAS	MPSNR	MSSIM	ERGAS	MPSNR	MSSIM	ERGAS	MPSNR	MSSIM	ERGAS
Noisy	20.00	0.4182	390.90	19.88	0.4142	395.03	12.41	0.1581	966.19	12.39	0.1577	970.60
FFDNet [1]	25.17	0.5645	206.51	25.00	0.5601	211.63	18.19	0.3136	536.36	18.16	0.3140	541.40
BALMF [42]	33.97	0.9256	77.42	33.90	0.9249	77.99	31.80	0.9254	100.75	31.66	0.9218	104.24
3DlogTNN [30]	34.73	0.9418	69.84	33.64	0.9319	82.09	34.94	0.9526	72.08	33.04	0.9155	118.43
L_0-PLRTV [22]	35.70	0.9597	63.11	35.68	0.9597	63.12	31.97	0.9149	109.58	28.96	0.7885	378.77
LRTDTV [8]	34.96	0.9460	67.53	34.92	0.9460	67.86	35.14	0.9495	66.38	35.08	0.9486	66.92
LRTDGS [15]	34.47	0.9424	71.75	34.43	0.9420	72.10	34.66	0.9452	70.26	34.58	0.9442	70.88
LTHTV [31]	29.59	0.8133	125.42	29.52	0.8116	126.74	29.78	0.8224	122.79	29.72	0.8202	123.75
L_{1-2}SSTV [24]	33.04	0.9135	85.57	32.76	0.9110	88.89	30.96	0.9027	107.67	30.54	0.8868	116.29
LRTFL_0 [29]	34.43	0.9533	84.40	34.38	0.9531	84.61	34.42	0.9542	85.34	34.41	0.9540	85.56
Proposed	**36.45**	**0.9621**	**57.70**	**36.40**	**0.9620**	**58.07**	**35.83**	**0.9576**	**62.81**	**35.76**	**0.9568**	**63.46**

are the average values of PSNR and SSIM for all bands of an HSI, respectively. Better HSI denoising results are indicated by higher MPSNR, MSSIM, and lower ERGAS values.

Tables II–IV summarizes three quantitative assessments of different denoising methods under Cases 1—4 for three HSIs, respectively. The best results for each quality metric are highlighted in bold. Clearly, the proposed L_0RLRTD demonstrates an obvious advantage for the HSI denoising over other compared approaches. Additionally, combining the visual comparison, we observe that the DL-based FFDNet method could not successfully recover the HSIs contaminated by mixed noise in Case 3

and Case 4. The primary reason is that FFDnet is a single-band deep denoiser, which ignores the correlation between different bands of an HSI, consequently leading to a poor recovery effect.

To make a more comprehensive comparison of the performance for all restoration methods, we present the PSNR value of each band for IndianP in Fig. 6. A more visual representation of the higher PSNR values achieved by the proposed L_0RLRTD method can be observed on almost all bands. Furthermore, we present the spectral signatures before and after restoration. Fig. 7 shows the digital number (DN) values of IndianP at the (30,30) pixel in Case 4. Combining with the ERGAS values in Table II,

TABLE IV
QUANTITATIVE EVALUATION ON THE SYNTHETIC NC16

| Case	Case 1			Case 2			Case 3			Case 4		
Method	MPSNR	MSSIM	ERGAS	MPSNR	MSSIM	ERGAS	MPSNR	MSSIM	ERGAS	MPSNR	MSSIM	ERGAS
Noisy	20.00	0.0999	475.28	19.99	0.0997	475.66	11.94	0.0250	1.21e3	11.92	0.0248	1.21e3
FFDNet [1]	38.12	0.9398	60.44	37.98	0.9396	61.37	17.31	0.1293	655.20	17.25	0.1274	658.55
BALMF [42]	38.55	0.8797	58.45	38.35	0.8775	59.75	31.05	0.8584	133.07	30.99	0.8569	133.82
3DlogTNN [30]	36.69	0.8294	70.13	36.02	0.8156	77.40	38.27	0.8837	62.16	36.47	0.8346	97.71
L_0-PLRTV [22]	41.60	0.9466	43.26	41.57	0.9465	43.40	37.92	0.9436	68.36	34.18	0.8100	237.93
LRTDTV [8]	43.90	0.9602	35.91	43.40	0.9586	38.62	43.22	0.9579	42.14	42.96	0.9567	51.85
LRTDGS [15]	43.36	0.9572	37.33	43.03	0.9569	35.58	41.46	0.9524	43.93	41.40	0.9523	44.14
LTHTV [31]	43.44	0.9583	38.35	43.14	0.9578	39.35	41.92	0.9557	44.06	41.44	0.9522	45.54
L_{1-2}SSTV [24]	33.70	0.7172	98.36	33.46	0.7117	102.24	33.80	0.7477	98.22	33.44	0.7337	103.85
LRTFL_0 [29]	43.22	0.9583	39.61	43.00	0.9580	40.41	42.28	0.9557	41.87	41.65	0.9519	46.67
Proposed	**44.57**	**0.9604**	**34.67**	**44.30**	**0.9601**	**35.38**	**43.41**	**0.9590**	**38.07**	**43.14**	**0.9543**	**42.38**
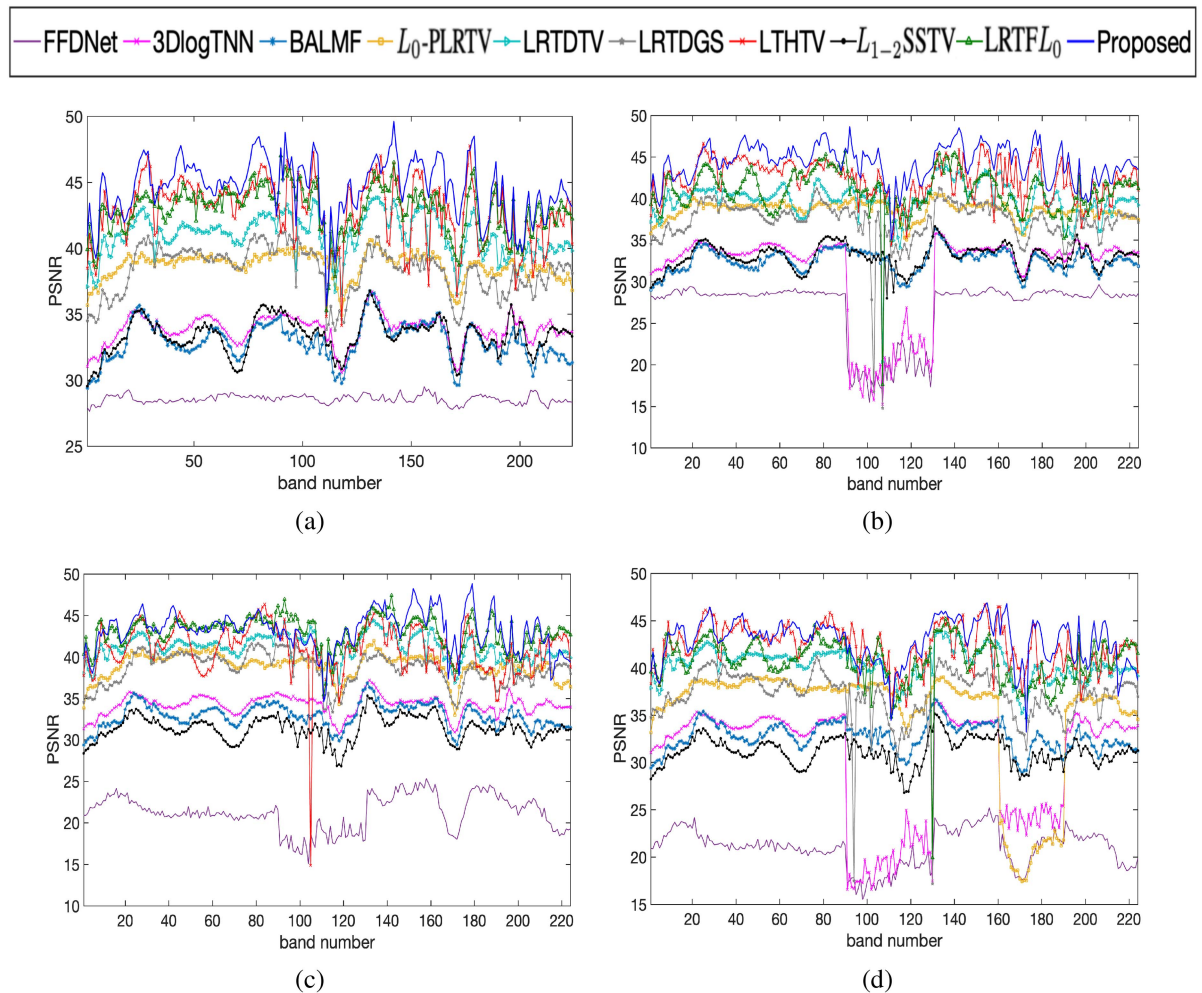
Fig. 6. PSNR values for each band: (a) Case 1, (b) case 2, (c) case 3, (d) case 4. Best viewed in ×2 sized color pdf file.
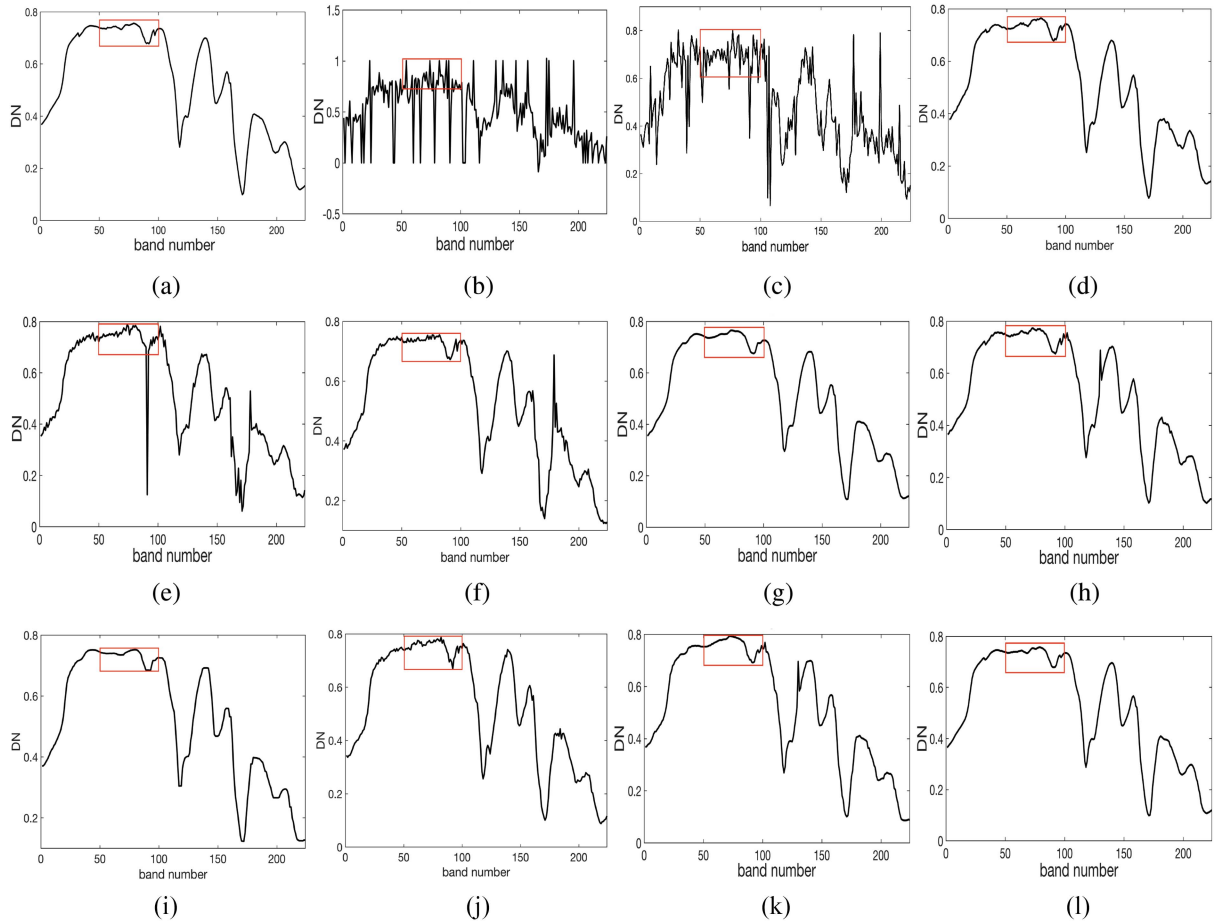
Fig. 7. Curves of spectral signatures pixel (30,30) restored by all compared methods under IndianP of Case 4. (a) Original. (b) Noise. (c) FFDNet. (d) BALMF. (e) 3DlogTNN. (f) L_0-PLRTV. (g) LRTDTV. (h) LRTDGS. (i) LTHTV. (j) TLR-L_{1-2}SSTV. (k) LRTFL_0. (l) Proposed.

it is evident that the proposed L_0RLRTD method yields the best spectral signature.

C. Experiments on Real-World Data

To verify the effectiveness of the proposed method in handling actual noisy HSIs, two typical real-world datasets are used to conduct experiments including the Shanghai dataset with the size of $300 \times 300 \times 155$, which was captured by the GaoFen-5 satellite, and the Hyperspectral Digital Imagery Collection Experiment (HYDICE) Urban[4] with the size of $307 \times 307 \times 210$. Some parameters are suggested as follows: $r^0 = (60, 60, 5)$, $r^{\max} = (270, 270, 5)$ for Shanghai, and $r^0 = (61, 61, 5)$, $r^{\max} = (276, 276, 5)$ for HYDICE Urban. The remaining parameters can refer to the parameter selection used in the simulation experiment.

Visual results of band 152 in Shanghai and band 200 in Urban are shown in Figs. 8 and 9, respectively. According to the visual results, we draw the following conclusions: first, in Figs. 8(b) and 9(b), the DL-based FFDNet method caused

[4][Online]. Available: http://www.tec.army.mil/hypercube

artifacts, and some details are not well preserved, which demonstrates that the trained network is not necessarily able to deal with the complex noise in real scenes; second, the other model-based test methods can suppress amount of the noise to restore the sketch of the HSIs, however, some methods including 3DlogTNN, L_0-PLRTV, and LRTF L_0 cannot completely remove the stripes noise in the red rectangle area of the Urban dataset; third, BALMF, LRTDTV, LRTDGS, and LTHTV could obtain better visual results for two real HSIs, but they inevitably caused blurry, thus affecting the visual effect. By comparison, the superiority of the proposed L_0 RLRTD method is evident in its ability to remove noise and preserve details.

Moreover, we employ the Q-metric [43], a blind image content measurement index, to assess the effectiveness of all compared methods on real-world datasets. The Q-metric value is calculated for each band, and the average is taken as the final evaluation result. Table V presents the evaluation results, we can observe that the proposed L_0RLRTD yields the best evaluation performance. Based on both the visual results and Q-metric values, we can conclude that the proposed L_0RLRTD method could successfully maintain the structural information and suppress
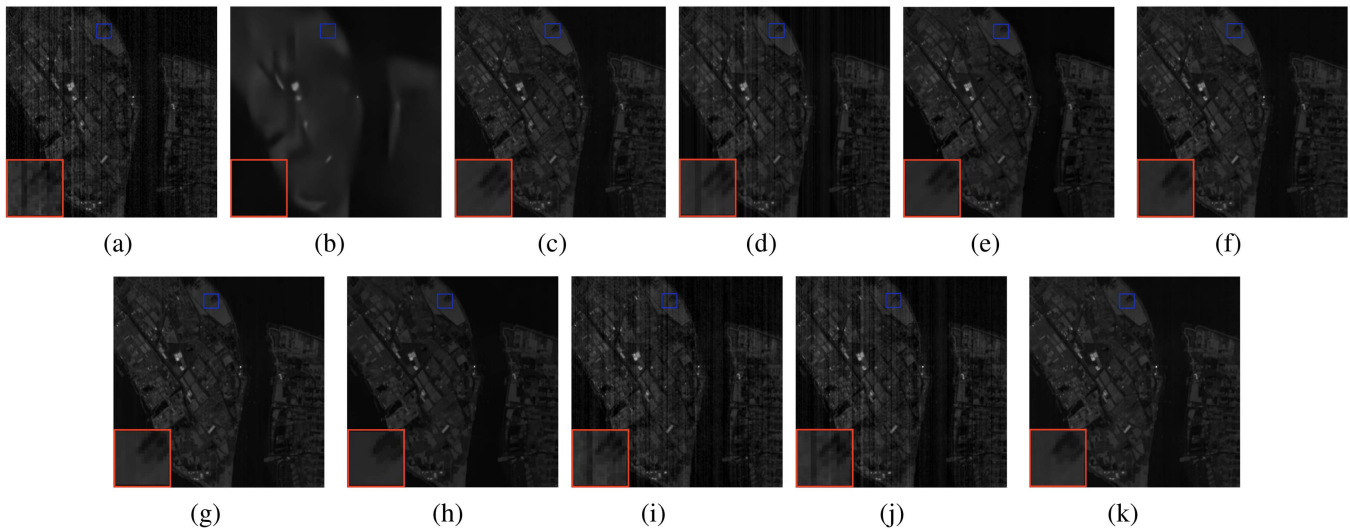
Fig. 8. Denoised visual results of band 152 in real HSI Shanghai. (a) Noise. (b) FFDNet. (c) BALMF. (d) 3DlogTNN. (e) L_0-PLRTV. (f) LRTDTV. (g) LRTDGS. (h) LTHTV. (i) L_{1-2}SSTV. (j) LRTFL_0. (k) Proposed.
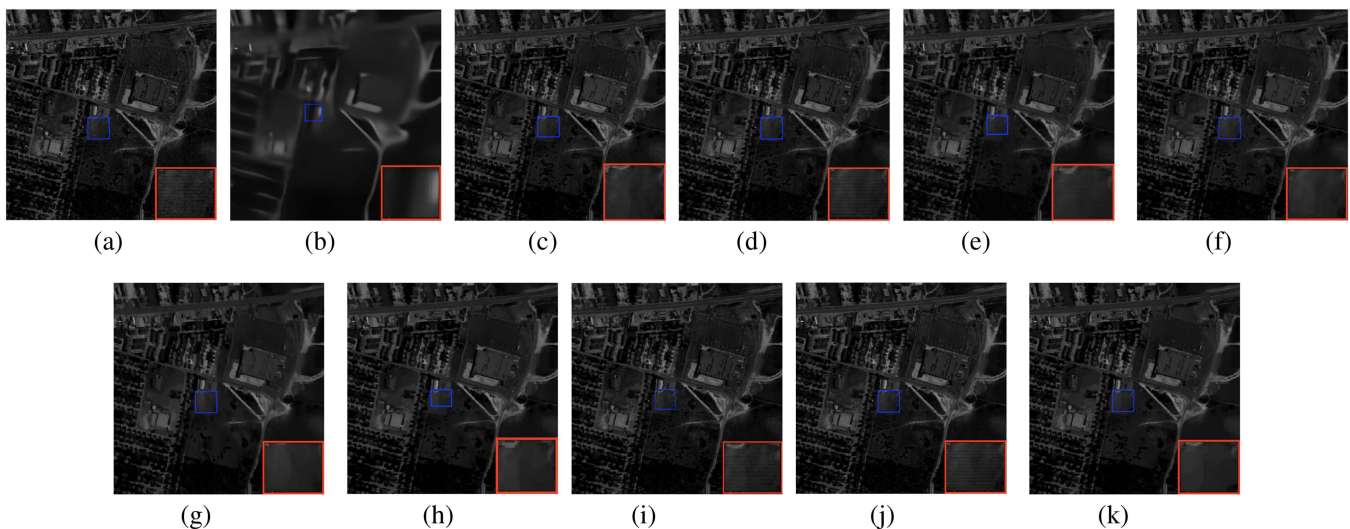

Fig. 9. Denoised visual results of band 200 in real HSI HYDICE Urban. (a) Noise. (b) FFDNet. (c) BALMF. (d) 3DlogTNN. (e) L_0-PLRTV. (f) LRTDTV. (g) LRTDGS. (h) LTHTV. (i) L_{1-2}SSTV. (j) LRTFL_0. (k) Proposed.

noise for HSIs, even when ground truth HSI is not available as a reference.

D. Technique Validation of the L_0RLRTD

1) Ablation Experiments: To validate the effectiveness of two L_0-based regularizers, we compare the proposed L_0RLRTD model (6) with two models on the synthetic IndianP dataset, including 1) Model 1: L_0RLRTD model (6) without the TV regularizer term $\| \nabla \mathcal{X} \|_0$, that is $\tau = 0$; 2) Model 2: L_0RLRTD model (6) without the sparse regularizer term $\|\mathcal{S}\|_0$, that is $\lambda = 0$. The comparison results of three evaluation indexes are presented via a histogram in Fig. 10(a)–(c). The superiority of the proposed L_0RLRTD is evident, which validates that

the L_0-based regularizer constraints for sparse noise and gradient image domains could promote the recovery results of HSIs.

2) Effectiveness of the Rank-Increasing Scheme: Compared to the commonly used rank-fixed strategy, the rank-increasing strategy has two following advantages:

1) Using the rank-increasing strategy not only alleviates the burden of selecting a proper rank, but also guarantees the recovery accuracy.
2) The denoising algorithm with the rank-increasing strategy iterates from a small rank, which reduces the time cost of the corresponding tensor decomposition. Therefore, the running time under the rank-increasing scheme is faster than the rank-fixed strategy when executing the ADMM-HT algorithm once.

TABLE V
BLIND QUALITY ASSESSMENT USING THE Q-METRIC IN REAL DATA
EXPERIMENT

	Urban	Shanghai
FFDNet [1]	0.0423	0.0226
BALMF [42]	0.0519	0.0333
3DlogTNN [30]	0.0616	0.0347
L_0-PLRTV [22]	0.0513	0.0319
LRTDTV [8]	0.0545	0.0340
LRTDGS [15]	0.0493	0.0341
LTHTV [31]	0.0577	0.0347
L_{1-2}SSTV [24]	0.0588	0.0334
LRTFL_0 [29]	0.0606	0.0340
Proposed	**0.0749**	**0.0354**

TABLE VI
TIME COMPARISON OF NINE METHODS ON THE SYNTHETIC WDC MALL

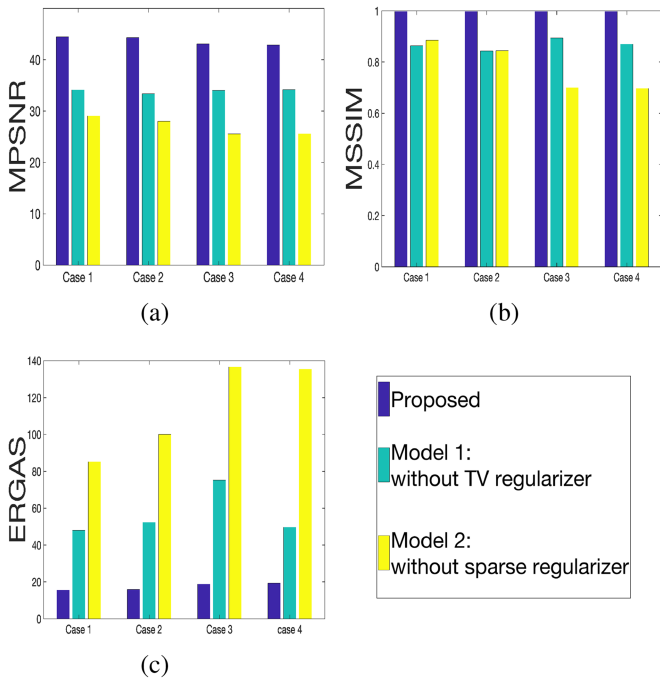| Time(s)	Case 1	Case 2	Case 3	Case 4
BALMF [42]	184	182	183	183
3DlogTNN [30]	582	583	575	575
L_0-PLRTV [22]	8050	7964	8419	8339
LRTDTV [8]	706	736	763	758
LRTDGS [15]	367	373	388	383
LTHTV [31]	533	533	562	559
L_{1-2}SSTV [24]	530	521	402	394
LRTFL_0 [29]	2293	2226	2136	2116
Proposed	523	532	538	545

Fig. 10. Performance comparison of Model 1, Model 2 and the proposed L_0RLRTD: (a) MPSNR, (b) MSSIM, (c) ERGAS.

To validate these advantages, we select the synthetic IndianP dataset as the objective and then show the relative square error (RSE [44]) and the time cost using Algorithm 1 with two rank selection schemes in Fig. 11. As suggested in [8], we set the rank as [120,120,10] for the rank-fixed scheme. The lower RSE values in Fig. 11(a) indicate that the ADMM-HT algorithm using the rank-increasing scheme achieves higher accuracy than using the rank-fixed scheme. From Fig. 11(b), we can see that the ADMM-HT algorithm with the rank-increasing scheme takes less time than the rank-fixed scheme. These results conclusively demonstrate the superiority of the rank-increasing strategy.

3) Time Comparison: In Table VI, we record the running time of the experiments on the synthetic WDC dataset in seconds. For fairness, we do not compare the FFDnet method, because it needs to run in a server with GPU, while other methods

are performed in MATLAB (2017a) on a desktop computer with 3.3 GHz CPU and 16 GB RAM.

We can clearly see that the low-rank matrix factorization-based BALMF takes the shortest time, but it is hard to handle the mixed noise removal. The time comparison among LRTDTV, LTHTV, and our L_0RLRTD confirms that using the rank-increasing scheme partly reduces the computational complexity. It also can be seen that the proposed L_0RLRTD has an obvious advantage compared to the 3DlogTNN, and L_0-PLRTV. Although our L_0RLRTD is slightly slower than LRTDGS and TLR-L_{1-2}SSTV, it achieves the best performances in the HSI restoration experiments.

4) Robustness of the Rank-Increasing Scheme: In our rank-decreasing scheme, the initial rank r^0 and the maximum rank r^{max} should be given in advance, we shall demonstrate that our algorithm is robust to the selection of r^0 and r^{max}.

We use the synthetic IndianP dataset as our objective. For the spectral mode, we directly adopt the HySime [41] to estimate the rank. For two spatial modes, we take 5%, 10%, 15%, 20%, 25%, 30%, 35%, and 40% of dimension size I_i as $r_i^0 (i = 1, 2)$, respectively. We compare the results when selecting different r^0 via the histogram in Fig. 12(a). It is not difficult to observe that the MPSNR value changes small in each case, which means that the selection of r^0 is robust to the performance of the algorithm. Similarly, we set the r_i^{max} as 75%, 80%, 85%, 90%, and 95% of dimension sizes I_i, respectively. The results of four cases are shown in Fig. 12(b). The little difference between the maximum and minimum MPSNR value indicates the robustness of the r^{max} to our algorithm.

5) Convergence Discussion: It is worth emphasizing that the L_0RLRTD model proposed in this study is nonconvex, the convergence to the theoretical guarantee for the global minimum cannot be provided. Next, we demonstrate the convergence of the proposed Algorithm 1 by numerical experiments. Taking three simulated HSI datasets for example, the changes of the MPSNR and RSE values for the L_0RLRTD method under Case 1 are shown in Fig. 13 (the experiments under other noise cases have similar results). With an increasing number of iterations, the MPSNR and RSE curves of IndianP, WDC, and NC16 reach stable values, indicating the convergence of the proposed Algorithm 1. At the same time, we can see that the curves do not necessarily increase or decrease monotonically. This behavior
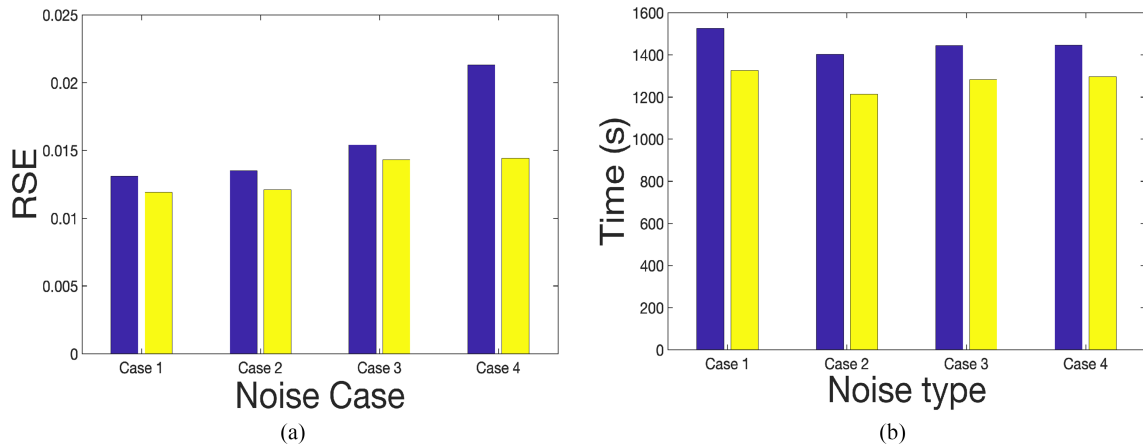
Fig. 11. Performance comparison between two rank selection schemes on the synthetic INDIAN PINES: (a) RSE, (b) time.
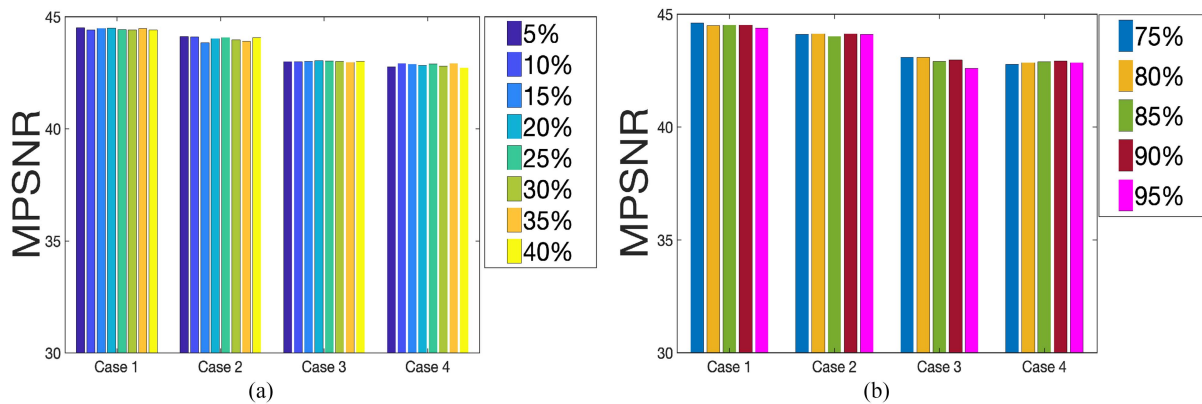

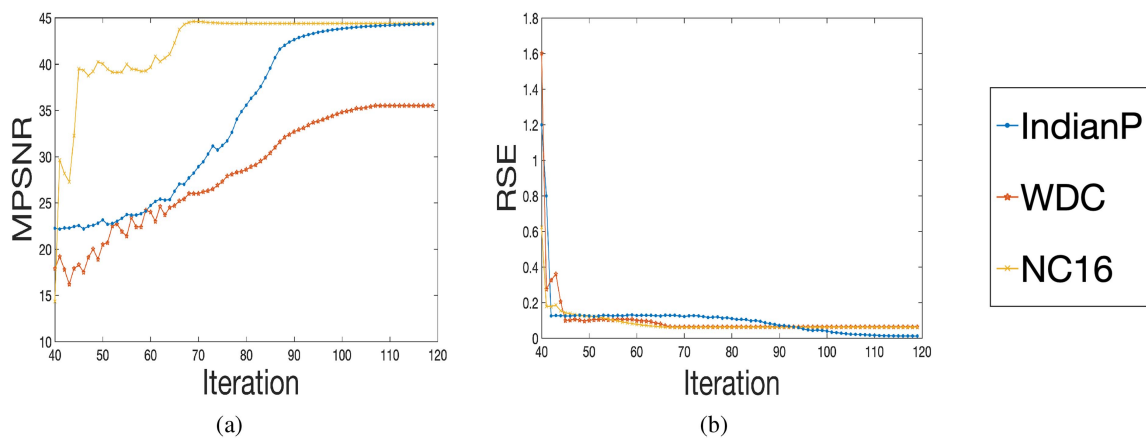Fig. 12. Robust of the rank selection: (a) r^0, (b) $r^{\max}$.

Fig. 13. Changes in MPSNR and RSE values with an increasing iteration number: (a) MPSNR, (b) RSE.

can be attributed to the nonconvexity of the proposed model and the dynamic updates of the penalty parameter β in Algorithm 1.

V. CONCLUSION

In this study, we introduce a novel method for HSIs denoising via L_0 regularized low-rank tensor decomposition. Specifically,

our approach utilizes the low-rank TKD with the Stiefel manifold to capture the spatial-spectral correlation of HSIs. For the noise corruption domain, we use the Frobenius norm to remove the Gaussian noise. Instead of the L_1-norm, we employ the L_0-norm for the sparse noise removal, which can better utilize the essence of sparse information. For the gradient image domain, we introduce a weighted L_0-norm on the gradient of

each pixel to explore the local piecewise smoothness structure of HSIs. Then, we design an algorithm named ADMM-HT to deal with the L_0-norm-based problem, meanwhile, we incorporate the rank-increasing scheme into ADMM-HT, so as to avoid the rank selection burden and improve the computational efficiency. Finally, through substantial experiments conducted on both simulated and real datasets, we demonstrate the superior performance of the proposed L_0RLRTD over other methods, as evidenced by both visual comparison and quantitative evaluation.

Furthermore, we plan to integrate DL techniques [2], [3], [4], [5], [6] with our model, leveraging the powerful learning and representation abilities of neural network models to improve recovery accuracy.

REFERENCES

[1] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[2] H. Pan, F. Gao, J. Dong, and Q. Du, "Multiscale adaptive fusion network for hyperspectral image denoising," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3045–3059, Mar. 2023.

[3] H. Sun, M. Liu, K. Zheng, D. Yang, J. Li, and L. Gao, "Hyperspectral image denoising via low-rank representation and CNN denoiser," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 716–728, Dec. 2022.

[4] S. Li, H. Xu, S. Gao, W. Liu, C. Li, and A. Liu, "An interferometric phase noise reduction method based on modified denoising convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4947–4959, Aug. 2020.

[5] Y. Luo, X. Zhao, T. Jiang, Y. Zheng, and Y. Chang, "Hyperspectral mixed noise removal via spatial-spectral constrained unsupervised deep image prior," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9435–9449, Sep. 2021.

[6] W. Ji, Y. Chen, K. Li, and X. Dai, "Multicascaded feature fusion-based deep learning network for local climate zone classification based on the So2Sat LCZ42 benchmark dataset," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 449–467, Nov. 2023.

[7] 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.

[8] 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.

[9] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[10] A. Cichocki et al., "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[11] I. V. Oseledets, "Tensor-train decomposition," *SIAM J. Sci. Comput.*, vol. 33, no. 5, pp. 2295–2317, 2011.

[12] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, "Tensor ring decomposition," 2016, *arXiv:1606.05535*.

[13] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1888–1902, Aug. 2018.

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

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

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

[17] M. Nikolova, "Model distortions in bayesian MAP reconstruction," *Inverse Problems Imag.*, vol. 1, pp. 399–422, 2007.

[18] C. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, 2010.

[19] X. Tian, K. Xie, and H. Zhang, "A low-rank tensor decomposition model with factors prior and total variation for impulsive noise removal," *IEEE Trans. Image Process.*, vol. 31, pp. 4776–4789, Apr. 2022.

[20] B. Zhang, G. Zhu, and Z. Zhu, "A TV-log nonconvex approach for image deblurring with impulsive noise," *Signal Process.*, vol. 174, 2020, Art. no. 107631.

[21] T. Xie, S. Li, and B. Sun, "Hyperspectral images denoising via nonconvex regularized low-rank and sparse matrix decomposition," *IEEE Trans. Image Process.*, vol. 29, pp. 44–56, Jun. 2020.

[22] C. Zhong, J. Zhang, Q. Guo, and Y. Zhang, "Improving sparse noise removal via L_0-Norm optimization for hyperspectral image restoration," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Feb. 2022.

[23] M. Wang, Q. Wang, J. Chanussot, and D. Hong, "l_0-l_1 hybrid total variation regularization and its applications on hyperspectral image mixed noise removal and compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7695–7710, Sep. 2021.

[24] 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 Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3309–3325, Apr. 2021.

[25] Y. Chen, T. Xu, X. Zhao, H. Zeng, Y. Xu, and J. Chen, "Asymmetry total variation and framelet regularized nonconvex low-rank tensor completion," *Signal Process.*, vol. 206, 2023, Art. no. 108901.

[26] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5249–5257.

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

[28] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, Apr. 2020.

[29] F. Xiong, J. Zhou, and Y. Qian, "Hyperspectral restoration via L_0 gradient regularized low-rank tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10410–10425, Dec. 2019.

[30] Y. Zheng, T. Huang, X. Zhao, T. Jiang, T. Ma, and T. Ji, "Mixed noise removal in hyperspectral image via low-fibered-rank regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 734–749, Jan. 2020.

[31] S. Xu, J. Zhang, and C. Zhang, "Hyperspectral image denoising by low-rank models with hyper-Laplacian total variation prior," *Signal Process.*, vol. 201, 2022, Art. no. 108733.

[32] Q. Shao, M. Bai, and X. Zhang, "Adaptive correction procedure for TVL1 image deblurring under impulse noise," *Inverse Problems*, vol. 32, no. 8, 2016, Art. no. 085004.

[33] J. Pan, M. K. Ng, Y. Liu, X. Zhang, and H. Yan, "Orthogonal nonnegative tucker decomposition," *SIAM J. Sci. Comput.*, vol. 43, no. 1, pp. B55–B81, 2021.

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

[35] G. Yuan and B. Ghanem, "l0TV: A new method for image restoration in the presence of impulse noise," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5369–5377.

[36] L. Zhang, Y. Qian, J. Han, P. Duan, and P. Ghamisi, "Mixed noise removal for hyperspectral image with $l_0 - l_{1-2}$ SSTV regularization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5371–5387, Jun. 2022.

[37] S. P. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[38] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L_0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, 2011, Art. no. 174.

[39] X. Zhang and M. K. Ng, "Sparse nonnegativetensor factorization and completion with noisy observations," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2551–2572, Apr. 2022.

[40] M. Ishteva, L. D. Lathauwer, P. Absil, and S. V. Huffel, "Differential-geometric newton method for the best rank-(R_1, R_2, R_3) approximation of tensors," *Numer. Algorithms*, vol. 51, no. 2, pp. 179–194, 2009.

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

[42] S. Xu, X. Cao, J. Peng, Q. Ke, C. Ma, and D. Meng, "Hyperspectral image denoising by asymmetric noise modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Dec. 2022.

[43] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3116–3132, Dec. 2010.

[44] X. Li, Y. Ye, and X. Xu, "Low-rank tensor completion with total variation for visual data inpainting," in *Proc. Thirty-First AAAI Conf. Artif. Intell.*, 2017, pp. 2210–2216.

Kun Xie (Member, IEEE) received the Ph.D. degree in computer application from Hunan University, Changsha, China, in 2007.

She is currently a Professor with Hunan University. She has authored or coauthored more than 60 articles in major journals and conference proceedings, including the IEEE/ACM TON, the IEEE TMC, the IEEE TC, the IEEE TWC, the IEEE TSC, the SIGMOD, INFOCOM, ICDCS, SECON, DSN, and IWQoS. Her research interests include network measurement, network security, Big Data, and AI.

Xin Tian is currently working toward the Ph.D. degree in computer science and technology with Hunan University, Changsha, China.

Her research interests include image processing and tensor optimization.

Hanling Zhang received the M.S. degree in applied mathematics from University of Central South, Changsha, China, in 2000, and the Ph.D. degree in signal and information process from University of Northwestern Polytechnic, Xi'an, China, in 2003.

He is currently a Professor with Hunan University. He has authored or coauthored more than 60 journal articles and conference papers. His research interests include image and video signal processing, Big Data analysis and visualization, Human—Computer Interaction.