**Mixed-Noise Band Selection for Hyperspectral Images**

ZHENG LI, (Student Member, IEEE), CHENWEI DENG, (Senior Member, IEEE), AND YUN HUANG, (Student Member, IEEE)

Beijing Key Laboratory of Embedded Real-Time Information Processing Technology, School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

Corresponding author: Chenwei Deng (cwdeng@bit.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 91438203 and Grant 31727901.

**ABSTRACT** Hyperspectral images (HSIs) with abundant spectral information are generally susceptible to various types of noise, such as Gaussian noise and stripe noise. Recently, a few quality-based selection algorithms have been proposed to remove noise bands from HSIs. However, these methods suffer from an inability to discriminate the mixed-noise bands of HSIs and are sensitive to image content variation and luminance changes. Here, we develop a mixed-noise band selection framework that can separate the Gaussian and stripe noise bands from HSIs effectively. We first improve tensor decomposition to reconstruct the mixed-noise components and low-rank components, which reduces the influence of image content and luminance changes. Spectral smoothness constraints and unidirectional total variation are incorporated into the decomposition model to enhance the separation performance for Gaussian and stripe noise. Then, different statistical features, including Weibull and histogram of oriented gradient (HOG) features, are applied to extract the robust parameters from mixed-noise components. More importantly, an extreme learning machine (ELM) is trained to predict the noise bands. The ELM has an extremely fast learning speed and tends to achieve better performance than other networks. Finally, by aggregating all these strategies, our methods can select the mixed-noise bands efficiently. The experimental results on both synthetic and real noise HSIs indicate that the proposed method outperforms the state-of-the-art methods.

**INDEX TERMS** Hyperspectral images, tensor decomposition, mixed-noise band selection.

I. INTRODUCTION

Hyperspectral images (HSIs) are widely applied in different fields, including hyperspectral classification [1], [2], object recognition, detection [3], [4] and superresolution [5], [6]. Some bands of HSIs are inevitably corrupted by various types of noise, degrading the image quality greatly [7], [8], which directly deteriorates the related processing of the HSIs. Therefore, it is crucial and useful to exploit band selection algorithms for HSIs. In the last few years, a number of noise band selection methods [9], [10] have been developed, which can be classified into two types: subjective and objective band selection. Since subjective judgment is time consuming and expensive, especially for multiband images, it is important to develop objective selection, which is the focus of this article.

Objective noise band selection methods employ quality metrics to remove the noise bands. According to the availability of a reference image, quality metrics can be divided into full-reference (FR) methods and no-reference (NR) methods. The former class of approaches [11]–[15], which employ reference images to assess the distorted data, have a high correlation with the human vision system. The algorithm proposed in [11] used the most popular structural similarity (SSIM) index, which combines luminance, contrast, and structural information to estimate images. In [12], Sheikh applied visual information fidelity (VIF) to quantify the loss of image information. In addition to these metrics, feature similarity (FSIM) [13] exploits a combination of phase congruency and gradients to describe the quality of an image. Wang et al. [14] advocated the use of non-negative matrix factorization (NMF) to measure image degradation. In [15], Zhu designed a simple multivariate SSIM (MvSSIM) index for HSIs that can be applied to assess spectral structural similarity via covariance among different bands. However, the reference images are unavailable in actual situations, strictly limiting the application domain of FR models.
Later algorithms [16]–[22] predict the image quality by extracting only information from the distorted images whose quality is being evaluated. These strategies also suffer from various drawbacks. For example, Saad et al. [16] computed discrete cosine transform (DCT) coefficients and built a generalized Gaussian distribution to indicate perceptual quality. In [17], Mittal proposed the blind/referenceless image spatial quality evaluation (BRISQUE) and natural image quality evaluation separately by employing the natural scene statistics (NSS) of local luminance values to quantify the distorted image quality. These models are highly sensitive to the image content and illumination variation, and approximation issues arise in the method [18]. The model in [19] utilized entropy calculated from local image blocks in the spatial and DCT domains. Xue et al. [20] subsequently developed a blind quality metric based on the bandpass image gradient magnitude (GM) and Laplacian of Gaussian (LOG) maps. Gu et al. [21] proposed an NR method using free-energy-based brain theory integrated with some classical human visual system-inspired features and NSS features by fitting the mean subtracted contrast-normalized coefficients. Liu et al. [22] collected gradient statistics features and employed an AdaBoost neural network to represent the image quality. These models are sensitive to luminance variation and cannot deal with stripe noise or be applied to HSI quality assessment directly. In [23], Yang et al. designed a spectral-spatial quality metric that extracts the quality-sensitive features from the image structure, texture, and spectra. However, this metric estimates HSIs as a whole and cannot be applied to remove mixed-noise bands.

To address the abovementioned drawbacks, we propose a mixed-noise band selection algorithm for HSIs. Three major steps are involved. First, tensor factorization [24] is improved to acquire the mixed-noise components of real HSIs. Then, multiple statistical models are calculated from the noise components to extract the robust features. Finally, these features are fed to an extreme learning machine (ELM), which learns to select the noise bands contaminated by Gaussian and stripe noise. Experiments demonstrate that our model correlates highly with the noise degree and is robust to the image background and illumination.

In summary, we make the following three contributions:

1. A novel noise band selection framework is proposed, where mixed-noise bands of HSIs can be evaluated exactly and selected from real HSIs.

2. A spectral smoothness constraint is designed to enhance the reconstruction performance so that Gaussian noise can be effectively extracted from low-rank components.

3. A new unidirectional total variation method is incorporated into our model to separate the stripe noise from the noise HSIs, thereby improving the stripe noise detection rate on complicated backgrounds.

The rest of this article is organized as follows: Section II presents the related tensor decomposition model. In Section III, we give a detailed introduction and discussion of our metric. The experimental results are reported in Section IV, and related analyses are given. Finally, we conclude in Section V.
III. METHOD

To overcome the aforementioned drawbacks, we improve the tensor robust principal component analysis (TRPCA) model to enhance the separation performance for the noise components of HSIs. The workflow of our selection algorithm is shown in Fig. 2. We employ tensor decomposition to separate the low-rank components from the original images and reduce the influence of the image content variations. The spectral smoothness constraint is incorporated into the low-rank component to enhance the separation performance for luminance changes between different bands of HSIs. Additionally, we design a unidirectional total variation method to separate the stripe noise from the low-rank components. After several iterations, we can reconstruct the mixed-noise components from the HSI cube. Furthermore, we extract multiple quality-sensitive parameters from the noise components and employ an ELM to predict the noise bands. In this way, the noise bands can be detected and selected accurately.

A. RELATED NOTIONS

We briefly review some tensor-related notions and preliminaries in this subsection. Throughout this article, we employ Euler script letters, e.g., $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, to denote tensors and boldface capital letters, e.g., $A \in \mathbb{R}^{n_1 \times n_2}$, to denote matrices. The $(i, j, k)$-th entry of tensor $A$ is denoted as $a_{ijk}$, and $n_3$ is the spectral band number of a HSI tensor. Here, we use MATLAB notation to denote slices of the tensor: the $(i, j, k)$-th horizontal, lateral and frontal slices are represented as $A(\cdot, \cdot, k)$, $A(i, \cdot, \cdot)$, and $A(\cdot, \cdot, k)$, respectively. The frontal slice $A(\cdot, \cdot, :)$, is denoted as $A^{(k)}$. Similarly, the tube of tensor $A$ is denoted as $A(i, j, :)$. The inner product of tensors $A$ and $B \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as $\langle A, B \rangle = \sum_{k=1}^{n_3} \langle A^{(k)}, B^{(k)} \rangle$. The inner product of matrices $A$ and $B \in \mathbb{R}^{n_1 \times n_2}$ is defined as $\langle A, B \rangle = \text{Tr}(A^*B)$, where $\text{Tr}(\cdot)$ denotes the matrix trace and $A^*$ denotes the conjugate transpose of $A$.

The norms of a tensor are denoted as $\|A\|_1 = \sum_{ijk} |a_{ijk}|$, and the Frobenius norm is denoted as $\|A\|_F = \sqrt{\sum_{ijk} |a_{ijk}|^2}$. We apply the MATLAB command $\text{fft}$ to define $\tilde{A}$ as the result of the discrete Fourier transformation of $A$ in the 3rd dimension, i.e., $\tilde{A} = \text{fft}(A, [], 3)$. $\tilde{A}$ denotes a block diagonal matrix with each block on a diagonal as the frontal slice $A^{(k)}$ of $A$, i.e.,

$$
\tilde{A} = \text{bdiag}(A) = \begin{bmatrix}
A^{(1)} \\
A^{(2)} \\
\vdots \\
A^{(n_3)}
\end{bmatrix}.
$$

The multiplication of tensors, called the t-product [27], [28], is defined below.

**Definition 1 (t-Product):** Consider $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$; its block circulant matrix has the size $n_1 n_3 \times n_2 n_3$ and is
defined as:

\[ b \circ \text{circ}(A) = \begin{bmatrix} A^{(1)} & A^{(n_2)} & \ldots & A^{(2)} \\ A^{(2)} & A^{(1)} & \ldots & A^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(n_3)} & A^{(n_3-1)} & \ldots & A^{(1)} \end{bmatrix}. \]

The unfolding and folding operators are separately defined as:

\[ \text{unfold}(A) = \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n_3)} \end{bmatrix}, \]

\[ \text{fold (unfold}(A)) = A. \]

Given \( A \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) and \( B \in \mathbb{R}^{n_2 \times l \times n_3} \), the t-product \( A \ast B \) is defined to be a tensor of size \( n_1 \times l \times n_3 \):

\[ A \ast B = \text{fold (bcirc}(A) \cdot \text{unfold}(B)). \quad (4) \]

**Definition 2 (Conjugate Transpose):** The conjugate transpose of a tensor \( A \) of size \( n_1 \times n_2 \times n_3 \) is \( A^* \), with size \( n_2 \times n_1 \times n_3 \), which is obtained by taking the conjugate transpose of each frontal slice and then reversing the order of the transposed frontal slices through \( n_3 \).

**Definition 3 (Orthogonal Tensor):** A tensor \( Q \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) is orthogonal if

\[ Q^* \ast Q = Q \ast Q^* = I, \]

where \( I \in \mathbb{R}^{n_1 \times n_1 \times n_3} \) is the identity tensor, whose first frontal slice is the identity matrix and whose other frontal slices are all zeros.

**Definition 4 (Tensor Nuclear Norm):** The tensor nuclear norm of \( A \) is defined as \( \|A\|_s \). The tensor nuclear norm is defined in the Fourier domain. It is closely related to the nuclear norm of the block circulant matrix in the original domain.

\[ \|A\|_s = \frac{1}{n_3} \sum_{i=1}^{n_3} \|A^{(i)}\|_s = \frac{1}{n_3} \|A\|_s = \frac{1}{n_3} \|b \circ \text{circ}(A)\|_s . \quad (5) \]

**Definition 5 (T-SVD):** \( A \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) can be decomposed as:

\[ A = U \ast S \ast V^*, \quad (6) \]

where \( U \in \mathbb{R}^{n_1 \times n_1 \times n_3} \) and \( V \in \mathbb{R}^{n_2 \times n_2 \times n_3} \) are orthogonal tensors. \( S \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) is a rectangular \( f \)-diagonal tensor.

To complete the decomposition, the matrix singular value decomposition (SVD) needs to be calculated in the Fourier domain [27].

Note that the tensor nuclear norm above is defined in the Fourier domain and is akin to the nuclear norm of the block circulant matrix in the original domain. For a detailed discussion, please refer to [27].

### B. Locally Smooth Spectrum Constraint

We can use Equ. (2) to recover the low-rank component and noise component of HSIs from sparse noise, but model (2) performs poorly in the separation of HSIs with mixed noise, such as Gaussian noise and stripe noise. Therefore, we modify model (2) and add different constraints to the objective function.

Since the spectra of HSIs are locally continuous [35], the two adjacent spectra are not greatly different from each other. Tikhonov regularization is employed as an extra term \( \|L_e \Gamma\|_F \) proportional to the objective function, where \( L_e \in \mathbb{R}^{(n_1 n_2) \times n_3} \) is the matricization [30] result of tensor \( L \in \mathbb{R}^{n_1 \times n_2 \times n_3} \) and \( \Gamma \) is given as follows:

\[ \Gamma = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ -1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & \ldots & 0 & -1 \end{bmatrix} \]

Essentially, this extra term “encourages” the algorithm to minimize the difference between the coefficients of adjacent spectra. As shown in Fig. 3, spectrum data learned with the smoothness constraints (Fig. 3(c)) resemble the original ones (Fig. 3(a)) much more closely than the result without smoothness constraints (Fig. 3(b)). Ultimately, this operation greatly improves the separation performance.

The optimization of our model can be implemented by a two-stage iterative algorithm: updating \( L \) with fixed \( \mathcal{E} \) and updating \( \mathcal{E} \) with fixed \( L \). When we update the low-rank component \( L \), the smoothness constraint is calculated in the spectral domain.

### C. UNIDIRECTIONAL TOTAL VARIATION

Stripe noise, which has a linear structure, is a common degradation phenomenon in space-borne and airborne remote sensing imaging systems. It is necessary to remove the stripe noise bands for HSI application. Surface texture prevents the recognition of stripe noise, as shown in Fig. 4(a), where the HOG features are disordered. The orientation of the HOG features in Fig. 4(b) is consistent with stripe noise. Therefore, it is crucial to separate the stripe noise from the low-rank component to improve the performance of stripe noise detection.

Suppose that images \( L_{a} \in \mathbb{R}^{n_1 \times n_2} \) are contaminated by vertical stripe noise \( E_{a} \in \mathbb{R}^{n_1 \times n_2} \). Unlike Gaussian noise, stripe noise is relatively sparse and has a certain direction. To better remove the stripe noise and preserve the texture information of HSIs, we incorporated a UTV constraint as follows:

\[ \min_{L_{a}} \|\nabla_{y} L_{a}\|_F + \|\nabla_{x} E_{a}\|_F. \quad (7) \]

Then, the final objective function is:

\[ \min_{L_{a}} \|L\|_s + \lambda \|\mathcal{E}\|_1 + \|L_e \Gamma\|_F + \|\nabla_{y} L_{a}\|_F \]

\[ s.t. \|D - L - \mathcal{E}\|_F \leq \sigma. \quad (8) \]
Check the convergence conditions:

$$\|\nabla \eta(\mathbf{L}_{k+1}) - \nabla \eta(\mathbf{L}_k)\|_F \leq \varepsilon,$$

where $\mathbf{L}_k \in \mathbb{R}^{m \times n_2}$ and $\varepsilon$ is the experiential threshold, which can be adjusted to obtain different separation effects. The algorithm proposed in this article is summarized as algorithm 1, which is described below.

**Algorithm 1** Separation Algorithm With ADMM

**Input:** tensor data $\mathcal{D}$, parameter $\lambda$.

**Initialization:** Parameters $\mathcal{L}_0 = \mathcal{E}_0 = \mathcal{Y}_0$, $\rho = 1.1$, $\mu_0 = 1e^{-3}$, $\varepsilon = 1e - 6$.

**While** not converged **do**

1. Update $\mathcal{L}_{k+1}$ by
   
   $$\mathcal{L}_{k+1} = \arg \min_{\mathcal{L}} \|\mathcal{L}\|_w + \frac{\mu}{2} \|\mathcal{L} + \mathcal{E}_k - \mathcal{D} + \frac{Y_k + \nabla \eta(\mathbf{L}_k)}{\mu}\|_F^2;$$

2. Update $\mathcal{E}_{k+1}$ by
   
   $$\mathcal{E}_{k+1} = \arg \min_{\mathcal{E}} \lambda \|\mathcal{E}\|_1 + \frac{\mu}{2} \|\mathcal{L}_{k+1} + \mathcal{E} - \mathcal{D} + \frac{Y_k}{\mu}\|_F^2;$$

3. $\mathcal{Y}_{k+1} = Y_k + \mu_k(\mathcal{L}_{k+1} + \mathcal{E}_{k+1} - \mathcal{D});$

4. $\Delta \mathcal{L} = \mathcal{L} - \mathcal{L}_k - \mathcal{E}_k - \nabla \mathcal{Y} \mathcal{L};$

5. Update $\mu_{k+1} = \min(\rho \mu_k, 1e10);$ 

6. Check the convergence conditions
   
   $\|\mathcal{L}_{k+1} - \mathcal{L}_k\| \leq \varepsilon$, $\|\mathcal{E}_{k+1} - \mathcal{E}_k\| \leq \varepsilon$, $\|\nabla \mathcal{Y}(\mathcal{L}_{k+1}) - \nabla \mathcal{Y}(\mathcal{L}_k)\|_F \leq \varepsilon;$

**end while**

During processing, the low-rank components $\mathcal{L}$ in the $y$ direction are reduced with a certain ratio to $\mathcal{L}_{k+1} = \text{prox}_{\tau \mathcal{G}}(\mathcal{L}_k - \tau \nabla \mathcal{Y} \mathcal{L}_k)$, where $\text{prox}_{\tau \mathcal{G}}$ is the proximity operator described in [36], [37], to keep the main structure of the HSI. The termination condition is:

$$\|\nabla \eta(\mathcal{L}_{k+1}) - \nabla \eta(\mathcal{L}_k)\|_F \leq \varepsilon,$$

where $\mathcal{L}_k \in \mathbb{R}^{m \times n_2}$ and $\varepsilon$ is the experiential threshold, which can be adjusted to obtain different separation effects. The algorithm proposed in this article is summarized as algorithm 1, which is described below.

D. **MULTIPLE STATISTICAL FEATURES OF THE NOISE COMPONENT**

Distortion changes the regular statistical features of the image, and measuring the deviations of the statistical features makes it possible to predict the image quality without a reference [23]. To describe the image quality, several models are built, and multiple statistical parameters are extracted from mixed-noise components.

1) **WEIBULL STATISTICS**

Fig. 5(a) shows an example noise map histogram that obviously follows a Weibull distribution, and the curves are distinct for images with various levels of Gaussian noise. Steven et al. [38] have also suggested that human vision systems seem to be tuned to the shape and scale parameters of images with a Weibull distribution. Therefore, we built a Weibull model to fit noise maps:

$$f(x; k, \lambda_1) = \begin{cases} \frac{k}{\lambda_1} \left(\frac{x}{\lambda_1}\right)^{k-1} \exp\left(\frac{x}{\lambda_1}\right) & x \geq 0 \\ 0 & x \leq 0 \end{cases}$$

where $x$ is an element of the noise components and $k$ and $\lambda_1$ are the shape and scale parameters, respectively. Fig. 5(b) shows the shape parameter $k$ under Gaussian noise $\sigma = 0.001, 0.005, 0.010, 0.050$. Clearly, the acquired parameter $k$ increases with increasing Gaussian noise. More importantly, each curve is relatively stable for different channel images, proving that our feature is illumination and content insensitive.
FIGURE 5. Illustrations of the noise component distributions and the corresponding Weibull fitting parameters. (a) The distribution of noise components with various Gaussian noise levels, and (b) the corresponding Weibull shape parameters for (a).

FIGURE 6. Structure of the HOG descriptor [39].

2) HOG FEATURES
Since stripe noise has a linear structure, we adopt the HOG features [39] to detect it in the noise components. The structure of the HOG descriptor applied in our algorithm is shown in Fig. 6. In this article, the images have dimensions of $64 \times 64$, a cell is $8 \times 8$ pixels in size, and each cell consists of 9 orientation bins. Each block, which contains $2 \times 2$ cells, yields a 36-dimensional feature vector. Then, the whole pass, which includes $7 \times 7$ blocks, has a 1764-dimensional descriptor vector. These high-dimensional descriptor vectors are used to train an ELM model to determine whether the images are affected by stripe noise. The model can be further optimized in practical applications.

E. NOISE BAND SELECTION BASED ON AN EXTREME LEARNING MACHINE
In this subsection, we introduce the band selection process of this work. Different approaches for feature mapping are compared, such as support vector regression (SVR) [16], [17] and neural networks (NNs) [22]. Usually, the extracted features are input into these models to fit a complex relationship. However, learning-based tools are often time consuming and suffer from problems such as overfitting and local optimization. Therefore, we employ an emergent machine technique, i.e., an extreme learning machine, to predict the noise bands.

In [40]–[42], Huang et al. originally proposed an ELM for generalized single-hidden-layer feed-forward neural networks (SLFNs), which has been used in various applications [43]–[45]. The ELM aims to learn an approximation function based on the training data. Suppose that SLFNs with $K$ hidden nodes can be represented by the following equation:

$$f L(\mathbf{x}_i) = \sum_{j=1}^{K} g(\mathbf{x}_i, a_{ij}, b_{ij}) \beta_j,$$

(11)

where $a_{ij}$ is the input weight connecting the input $\mathbf{x}_i$ to the $j$-th hidden node, $b_{ij}$ is the bias connecting the input $\mathbf{x}_i$ with the $j$-th hidden node, $g(\cdot)$ is the activation function, and $\beta_j$ is the output weight of the $j$-th hidden node. The activation function $g(\cdot)$ can be any nonlinear piecewise continuous function, such as the sigmoid function (equation 12) or radial basis function (RBF, equation 13)

$$g(\mathbf{x}; \theta) = \frac{1}{1 + \exp(-a^T X + b)}$$

(12)

$$g(\mathbf{x}; \theta) = \exp(-b \|X - a\|_2^2),$$

(13)

where $\theta = (a, b)$ are the parameters of the mapping function and $\| \cdot \|_2$ denotes the Euclidean norm.

In [41], Huang et al. proved that SLFNs can approximate any continuous target function over any compact subset $X$ with the above sigmoid and RBF functions. Training ELMs is equivalent to solving a regularized least-squares problem, which is considerably more efficient than training an SVM or learning with back-propagation. Therefore, in our model, an ELM is employed for mapping the HSI features $[k \lambda]$ into different noise degrees and classifying the HOG features of stripe noise bands in various scenes.

IV. EXPERIMENT
We first introduce the test data and evaluation criteria. Then, we compare the performance of our proposed method with other denoising methods, including TV [36], TGV [36], BM4D [46], NMoG [47], and TRPCA [28] for the low-rank component. We report the mean peak signal-to-noise ratio (MPSNR) and mean structural similarity index (MSSIM) values of the reference images and reconstructed images. We also conduct an experiment on the data collected from the real sensor, and the corresponding results are presented. Furthermore, we compare the performance of
our proposed method with other no-reference indices, including BLINDS2 [16], BRISQUE [17], NIQE [18], SSEQ [19], GM-LOG [20] and OG [22]. We report the obtained Spearman rank-order correlation coefficient (SRCC) and Pearson linear correlation coefficient (PLCC) values between the image prediction result and the corresponding noise degree. The scatter plots of the predictions versus the image noise degree are provided. The detection results for the stripe noise are also presented.

A. EXPERIMENTAL IMAGE AND EVALUATION CRITERIA

Two datasets are used in the experiment. The first dataset is from Pavia University. The size of the images we tested is 300 × 300 × 103. Gaussian noise and mixed noise (Gaussian and stripe noise) are added to the test images. The second dataset was acquired by an airborne visible infrared imaging spectrometer (AVIRIS) sensor, and it contains 224 spectral bands in the range of 400 ~ 2500nm. The size of each hyperspectral image is 200 × 200 × 224. We randomly selected 120 HSIs with diverse image contents, including rural, urban, port, mountain, agricultural, and water areas, as shown in Fig. 7. The pixel values of the HSIs are normalized to [0, 1].

The MPSNR [48] measures the mean of the peak signal-to-noise ratio over all bands between the original HSI and reconstructed HSI.

The MSSIM [11] measures the mean of the structural similarity (SSIM) between a reference HSI and the recovered HSIs over all channels.

The PLCC [49] measures the prediction accuracy of different quality metrics.

The SRCC [49] measures the prediction monotonicity of different quality metrics. A better quality metric has both a higher PLCC and higher SRCC [50].

B. SEPARATION PERFORMANCE FOR GAUSSIAN NOISE DATA

To compare our model with other separation algorithms [28], [36], [46], [47], we first add Gaussian noise with a standard deviation of σ = 0.1 to the Pavia University dataset. Fig. 8 shows the separation performance of different algorithms, and the MPSNR and MSSIM values of the compared methods are presented in Table 1. Fig. 8(a) is the Gaussian-noise image, and Fig. 8(b)-(e) are the corresponding separation results of TGV, BM4D, NMoG and TRPCA. They show that BM4D can filter Gaussian noise effectively but has defects related to the computational complexity of block-based processing. Furthermore, BM4D performs poorly on stripe noise. NMoG can preserve structure and texture perfectly, while the MPSNR is relatively low. Our method performs well on both the MPSNR and MSSIM. The denoised results prove that our method can separate the low-rank and noise components effectively.

C. SEPARATION RESULTS FOR MIXED-NOISE DATA

We also add mixed noise, including Gaussian noise and stripe noise, to the Pavia University dataset. The standard deviation of Gaussian noise is 0.1. The stripe noise is added as follows: Suppose the input image is \( I \in \mathbb{R}^{n_1 \times n_2} \) and \( m_r \in \mathbb{R}^{1 \times n_2} \) is a matrix of \( \mathcal{N}(0, 1) \). The stripe noise image \( I_{str} \) is given as follows:

\[
I_{str}(i, :) = (I(i, :) + \mu_r(i) \ast \varphi \ast m_r),
\]

where \( \mu_r(i) \) is the mean value of image \( I(i, :) \), i.e., \( \mu_r(i) = \sum_{j=1}^{n_2} I(i, j)/n_2 \), and \( \varphi = 0.3 \) is the percentage of stripe noise that we add. The mixed-noise image is shown in Fig. 9, and Fig. 9(b)-(e) are the corresponding separation results of TGV, BM4D, NMoG and TRPCA. The MPSNR and MSSIM of the compared methods are presented in Table 2. BM4D can remove Gaussian noise but performs poorly on stripe noise. NMoG achieves a similar MSSIM as our model, while the MPSNR remains relatively low. It is obvious that our approach is superior to the others, which demonstrates the efficiency of the multiple constraints adopted in our method.

D. SEPARATION RESULTS FOR ACTUAL NOISY DATA

We compare our model with BM4D, NMoD, and TRPCA on real HSI datasets acquired from the AVIRIS sensor. Some bands of AVIRIS hyperspectral images are seriously polluted by the atmosphere and water absorption. We use all of them without removing any band data to demonstrate the robustness of the proposed method for denoising. Fig. 10 shows bands 2, 107 and 108 of AVIRIS for all competing methods. The original images are polluted by complex noise, such as Gaussian noise, stripe noise, impulse noise and blur. Our model still obtains a better separation visualization and preserves the structure and texture well from the original
images. The reconstructed results substantiate the robustness of the proposed method in noise component separation.

E. GAUSSIAN NOISE QUALITY PERFORMANCE COMPARISON

To compare the proposed algorithm with other quality metrics [16]–[20], [22] on high-frequency noise assessment, we add white Gaussian noise to the second database, which contains 120 HSIs, where each HSI consists of 189 spectral bands (without the water absorption bands and noisy bands). The variance $\sigma^2$ of Gaussian noise increases from 0 to 0.004 by 0.0002 in each band. We randomly select 5%, 10% and 15% of the images from the noisy data to train the ELM model, and the rest are used for testing. The PLCCs and SRCCs of
the objective quality scores and the corresponding Gaussian noise degrees are presented in Table 3. Clearly, our proposed model delivers a better PLCC and SRCC than the compared methods. Moreover, the PLCCs and SRCCs of the models grow with increasing training data, which accords with the actual conditions.

Fig. 11 shows the scatter plots of the Gaussian noise degree versus the compared model in the second dataset. Note that the scatter plots of the proposed model exhibit good linearity, tight clustering, and a relatively uniform density. They are consistent with the high PLCC and SRCC shown in the bottom row of Table 3.

**F. STRIPE NOISE BAND SELECTION USING HOG FEATURES**

We apply our model to the HSIs and separate the noise components of the HSIs. We use 15% of the data to train the ELM model, and the confusion matrixes of our models are shown in Fig. 12, indicating that our algorithm effectively distinguishes stripe noise. The recall for stripe noise detection on the HSIs and for the corresponding reconstructed noise components are 0.924 and 0.994, respectively, satisfying the demands of practical application. More importantly, the recall of the experiment clearly increased by using a noise component, demonstrating that tensor factorization plays a significant role in stripe noise detection.
In this work, we study mixed-noise band selection for HSIs. First, the tensor decomposition model is improved by adding different constraints for separating the mixed-noise components from the low-rank tensor. Then, multiple statistical features are calculated to extract the robust parameters. Finally, an ELM is trained to select the noise bands from real HSIs. The experimental results show that our model detects noise images more accurately and robustly than the compared methods.

**REFERENCES**

[1] C.-I. Chang, “Statistical detection theory approach to hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 4, pp. 2057–2074, Apr. 2019.

[2] H. Xu, H. Zhang, W. He, and L. Zhang. “Superpixel-based spatial-spectral dimension reduction for hyperspectral imagery classification,” Neurocomputing, vol. 360, pp. 138–150, 2019.

[3] P. Sidike, V. K. Asari, and M. S. Alam. “Multiclass object detection with single query in hyperspectral imagery using class-associative spectral fringe-adjusted joint transform correlation,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 2, pp. 1196–1208, Feb. 2016.

[4] L. Zhang, Y. Zhang, H. Yan, Y. Gao, and W. Wei. “Salient object detection in hyperspectral imagery using multi-scale spectral-spatial gradient,” Neurocomputing, vol. 291, pp. 215–225, May 2018.

[5] Y. Zhao, J. Yang, and J. C.-W. Chan. “Hyperspectral imagery super-resolution by spatial–spectral joint nonlocal similarity,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 6, pp. 2671–2679, Jun. 2014.

[6] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li. “Hyperspectral image super-resolution using deep convolutional neural network,” Neurocomputing, vol. 266, pp. 29–42, Nov. 2017.

[7] M. Wang, J. Yu, J.-H. Xue, and W. Sun. “Denoising of hyperspectral images using group low-rank representation,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 9, no. 9, pp. 4420–4427, Sep. 2016.

[8] G. Ma, T.-Z. Huang, J. Huang, and C.-C. Zheng. “Local low-rank and sparse representation for hyperspectral image denoising,” IEEE Access, vol. 7, pp. 79850–79865, 2019.

[9] K. Sun, X. Geng, L. Ji, and Y. Lu. “A new band selection method for hyperspectral image based on data quality,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 6, pp. 2697–2703, Jun. 2014.

[10] X. Cao, X. Li, Z. Li, and L. Jiao. “Hyperspectral band selection with objective image quality assessment,” Int. J. Remote Sens., vol. 38, no. 12, pp. 3656–3668, Jun. 2017.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

[12] H. R. Sheikh and A. C. Bovik. “Image information and visual quality,” IEEE Trans. Image Process., vol. 15, no. 2, pp. 430–444, Feb. 2006.

[13] L. Zhang, L. Zhang, X. Mou, and D. Zhang. “FSIM: A feature similarity index for image quality assessment,” IEEE Trans. Image Process., vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[14] S. Wang, C. Deng, W. Lin, G.-B. Huang, and B. Zhao. “NMF-based image quality assessment using extreme learning machine,” IEEE Trans. Cybern., vol. 47, no. 1, pp. 232–243, Jan. 2017.

[15] R. Zhu, F. Zhou, and J.-H. Xue. “MvSSIM: A quality assessment index for hyperspectral images,” Neurocomputing, vol. 272, pp. 250–257, Oct. 2018.

[16] M. A. Saad, A. C. Bovik, and C. Charrier. “Blind image quality assessment: A natural scene statistics approach in the DCT domain,” IEEE Trans. Image Process., vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[17] A. Mittal, A. K. Moorthy, and A. C. Bovik. “No-reference image quality assessment in the spatial domain,” IEEE Trans. Image Process., vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[18] A. Mittal, R. Soundararajan, and A. C. Bovik. “Making a ‘completely blind’ image quality analyzer,” IEEE Signal Process. Lett., vol. 20, no. 3, pp. 209–212, Mar. 2013.

[19] L. Liu, B. Liu, H. Huang, and A. C. Bovik. “No-reference image quality assessment based on spatial and spectral entropies,” Signal Process., Image Commun., vol. 29, no. 8, pp. 856–863, Sep. 2014.

[20] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng. “Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features,” IEEE Trans. Image Process., vol. 23, no. 11, pp. 4850–4862, Nov. 2014.

[21] K. Gu, G. Zhai, X. Yang, and W. Zhang. “Using free energy principle for blind image quality assessment,” IEEE Trans. Multimedia, vol. 17, no. 1, pp. 50–63, Jan. 2015.
Liu, Hua, Q. Zhao, H. Huang, and A. C. Bovik, “Blind image quality assessment by relative gradient statistics and AdaBoosting neural network,” Signal Process., Image Commun., vol. 40, pp. 1–15, Jan. 2016.

J. Yang, Y. Zhao, Y. Chen, and C. Chan, “No-reference hyperspectral image quality assessment via quality-sensitive features learning,” Remote Sens., vol. 9, no. 4, pp. 305–329, Mar. 2017.

J. Chen, W. Zhang, Y. Qian, and M. Ye, “Deep tensor factorization for hyperspectral image classification,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., Jul. 2018, pp. 4788–4791.

A. Karami, M. Yazdi, and A. Z. Asli, “Noise reduction of hyperspectral images using kernel non-negative tucker decomposition,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 3, pp. 487–493, Jun. 2011.

X. Liu, S. Bourennane, and C. Fossati, “Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 10, pp. 3717–3724, Oct. 2012.

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. (CVPR), Jun. 2016, pp. 5249–5257.

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.

E. J. Candés, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” J. ACM, vol. 58, no. 3, pp. 1–37, 2011.

T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” SIAM Rev., vol. 51, no. 3, pp. 455–500, Aug. 2009.

C. Mu, B. Huang, J. Wright, and D. Goldfarb, “Square deal: Lower bounds and improved relaxations for tensor recovery,” in Proc. 31st Int. Conf. Mach. Learn., vol. 2, Jul. 2013, pp. 73–81.

B. Huang, C. Mu, D. Goldfarb, and J. Wright. (2014). Provable Low-Rank Tensor Recovery. [Online]. Available: http://www.optimization-online.org/DB_HTML/2014/02/4252.html

F. Fan, Y. Ma, C. Li, X. Mei, J. Huang, and J. Ma, “Hyperspectral image denoising with superpixel segmentation and low-rank representation,” Inf. Sci., vol. 397–398, pp. 48–68, Aug. 2017.

S. Boyd, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1–122, 2010.

S. Jia, J. Hu, J. Zhu, X. Jia, and Q. Li, “Three-dimensional local binary patterns for hyperspectral imagery classification,” IEEE Trans. Geosci. Remote Sens., vol. 55, no. 4, pp. 2399–2413, Apr. 2017.

L. Condat, “A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,” J. Optim. Theory Appl., vol. 158, no. 2, pp. 460–478, 2013.

L. Condat, “A generic proximal algorithm for convex optimization—Application to total variation minimization,” IEEE Signal Process. Lett., vol. 21, no. 8, pp. 985–989, Aug. 2014.

H. S. Scholte, S. Ghebreab, L. Waldoor, A. W. M. Smeulders, and V. A. F. Lamme, “Brain responses strongly correlate with Weibull image statistics when processing natural images,” J. Vis., vol. 9, no. 4, pp. 1–15, Apr. 2009.

N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, Jun. 2005, pp. 886–893.

G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: Theory and applications,” Neurocomputing, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

G.-B. Huang, L. Chen, and C.-K. Siew, “Universal approximation using incremental constructive feedforward networks with random hidden nodes,” IEEE Trans. Neural Netw., vol. 17, no. 4, pp. 879–892, Jul. 2006.

G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme learning machine for regression and multiclass classification,” IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 42, no. 2, pp. 513–529, Apr. 2012.

Y. Xu, Z. Y. Dong, K. Meng, R. Zhang, and K. P. Wong, “Real-time transient stability assessment model using extreme learning machine,” IET Gener. Transmiss. Distrib., vol. 5, no. 3, pp. 314–322, Mar. 2011.

S. Wang, C. Deng, B. Zhao, G.-B. Huang, and B. Wang, “Gradient-based no-reference image blur assessment using extreme learning machine,” Neurocomputing, vol. 174, pp. 310–321, Jan. 2016.

S. Zhou, C. Deng, W. Wang, G.-B. Huang, and B. Zhao, “GenELM: Generative extreme learning machine feature representation,” Neurocomputing, vol. 362, pp. 41–50, Sep. 2019.

M. Maggioni, V. Katiyarov, K. Egiezarian, and A. Foi, “Nonlocal transform-domain filter for volumetric data denoising and reconstruction,” IEEE Trans. Image Process., vol. 22, no. 1, pp. 119–133, Jan. 2013.

Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, “Denoising hyperspectral image with non-iid noise structure,” IEEE Trans. Cybern., vol. 48, no. 3, pp. 1054–1066, Mar. 2018.

Q. Huyhn-Thu and M. Ghanbari, “Scope of validity of PSNR in image/video quality assessment,” Electron. Lett., vol. 44, no. 13, pp. 800–801, Jun. 2008.

A. K. Moorthy and A. C. Bovik, “A two-step framework for constructing blind image quality indices,” IEEE Signal Process. Lett., vol. 17, no. 5, pp. 513–516, May 2010.

C. Deng, S. Wang, A. C. Bovik, G.-B. Huang, and B. Zhao, “Blind noisy image quality assessment using sub-band kurtosis,” IEEE Trans. Cybern., vol. 50, no. 3, pp. 1146–1156, Mar. 2020.