



Article

Variational Low-Rank Matrix Factorization with Multi-Patch Collaborative Learning for Hyperspectral Imagery Mixed Denoising

Shuai Liu ^{1,*} , Jie Feng ² and Zhiqiang Tian ¹

¹ School of Software Engineering, Xi'an Jiaotong University, No.28, Xianning West Road, Xi'an 710049, China; zhiqiangtian@mail.xjtu.edu.cn

² Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, 266 Xinglong Section of Xifeng Road, Xi'an 710126, China; jiefeng@xidian.edu.cn

* Correspondence: sh_liu@mail.xjtu.edu.cn

Abstract: In this study, multi-patch collaborative learning is introduced into variational low-rank matrix factorization to suppress mixed noise in hyperspectral images (HSIs). Firstly, based on the spatial consistency and nonlocal self-similarities, the HSI is partitioned into overlapping patches with a full band. The similarity metric with fusing features is exploited to select the most similar patches and construct the corresponding collaborative patches. Secondly, considering that the latent clean HSI holds the low-rank property across the spectra, whereas the noise component does not, variational low-rank matrix factorization is proposed in the Bayesian framework for each collaborative patch. Using Gaussian distribution adaptively adjusted by a gamma distribution, the noise-free data can be learned by exploring low-rank properties of collaborative patches in the spatial/spectral domain. Additionally, the Dirichlet process Gaussian mixture model is utilized to approximate the statistical characteristics of mixed noises, which is constructed by exploiting the Gaussian distribution, the inverse Wishart distribution, and the Dirichlet process. Finally, variational inference is utilized to estimate all variables and solve the proposed model using closed-form equations. Widely used datasets with different settings are adopted to conduct experiments. The quantitative and qualitative results indicate the effectiveness and superiority of the proposed method in reducing mixed noises in HSIs.

Keywords: denoising; low-rank matrix factorization; patch learning; variational inference; hyperspectral image



Citation: Liu, S.; Feng, J.; Tian, Z. Variational Low-Rank Matrix Factorization with Multi-Patch Collaborative Learning for Hyperspectral Imagery Mixed Denoising. *Remote Sens.* **2021**, *13*, 1101. <https://doi.org/10.3390/rs13061101>

Academic Editor: Benoit Vozel

Received: 1 February 2021

Accepted: 12 March 2021

Published: 14 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hyperspectral images (HSIs) are acquired by hyperspectral sensors, represented as a 3D data-cube containing both rich spectral and spatial information. Due to the limitations of the acquisition and transmission process, HSIs unavoidably suffer from various degradations, such as noise contamination, stripe corruption, missing data relating to the voxels in the data-cube or entire spectral bands [1–5]. These degradations severely limit the quality of the images and influence the precision of the subsequent processing, including unmixing, target detection, and classification [6–9]. Therefore, image restoration is of critical importance and challenging in the preprocessing stage of HSI analysis.

Previously, the traditional 2D or 1D denoising models have been applied for reducing noises in HSIs pixel-by-pixel [10] or band-by-band [11]; however, these methods ignore the correlations between different spectral bands or adjacent pixels and often result in relatively low-quality results. To further enhance the denoising performance, more efficient methods have been proposed, of which the key point is to elaborately encode the prior knowledge on the structure underlying a natural HSI, especially the characteristic across the spatial and spectral dimensionality of the image.

Othman and Qian [12] made an initial attempt to resolve this issue by designing a hybrid spatial–spectral derivative-domain wavelet shrinkage model, which was constructed by exploring the dissimilarity of the signal regularity existing along the space and spectrum of a natural HSI. Fu et al. [13] proposed an effective restoration model by considering the underlying sparsity across the spatial–spectral domain, high correlation across spectra, and non-local self-similarity over space. Meanwhile, a series of methods expanding the wavelet-based method from 2D to 3D has been proposed, for example, the so-called “non-local means” filtering approach has become popular in image processing and extensions have been developed in order to denoise structural 3D images [14]. Letexier et al. [15] proposed a generalized multi-dimensional Wiener filter for denoising hyperspectral images. Similarly, Chen et al. [16] extended Sendur and Selesnick’s bivariate wavelet thresholding from 2D image denoising to 3D data cube denoising. To get better denoising results, as an extension of the BM3D [11] method, Maggioni et al. presented a BM4D model [17]. Utilizing highly correlated spectral information and highly similar spatial information, the spectral-spatial adaptive sparse representation model was proposed for reducing the noise in HSIs [18]. By explicitly treating HSI data as a 3D cube, denoising models based on tensor decomposition have appeared. In Reference [19], a novel coupled spectral–spatial tensor representation framework was proposed for noise reduction of hyperspectral images. Chen et al. proposed a low-rank tensor decomposition model for HSI restoration [20]; however, most of the above-mentioned approaches are limited due to their insufficient usage of the correlations in the spectral domain, which results in suboptimal performance while suppressing mixed noises.

By efficiently exploring the latent knowledge across spectral bands for HSIs, low-rank models have been proposed and widely utilized to restore the pure datasets from the degraded images, with competitive performances [21–26]. The classical low-rank matrix factorization (LRMF) model was presented by K. Mitra et al. and T. Okatani et al. [24]. Subsequently, using the low-rank matrix recovery (LRMR) framework, an HSI restoration technique was explored to simultaneously remove various noises in an HSI [25]. The global and non-local low-rank factorization (GLF) was proposed to suppress the noises in HSIs by utilizing the low dimensional sub-spaces and the self-similarity of the real HSI [26]. These approaches obtained satisfactory results by effectively exploiting the spectral information.

To sufficiently enhance the denoising performance, it is necessary to well integrate the spatial characteristics of HSIs into the low-rank-based models [27–32]. Wang et al. [29] proposed a novel low-rank constraint and spatial–spectral total variation regularization model by jointly utilizing global low-rank and local spatial–spectral smooth properties of HSIs. Wang et al. [30] developed a total variation regularized low-rank tensor decomposition (LRTDTV) method, in which HSI was regarded as a third-order tensor rather than being divided into the patches or being unfolded into the matrix. In [31], a novel robust principal component analysis approach was introduced into the spatial–spectral low-rank model for mixed noise removal by fully identifying the intrinsic structures of the mixed noise and clean HSIs. Based on the global correlation along the spectra and nonlocal self-similarity across space, a low-rank tensor dictionary learning (LTDL) approach was explored with satisfactory performance in [32]. In the spatial domain, HSI has latent consistency. Using this, patch learning has been widely applied to depict spatial information and has achieved good performance [33–37]. When the HSIs are heavily polluted by noise, the patches with little effective information are usually not directly used to recover the noise-free data. Therefore, these methods could not efficiently learn and represent the intrinsic spatial consistency and nonlocal similarities of HSIs and thus limit their denoising performance.

Deep learning has also been widely used for HSIs [38–43]. Their success suggests its effectiveness for learning and depicting latent features when denoising the HSI. Additionally, hyperspectral images are usually polluted by various noises, which often have different statistical features [44,45], such as the noises depending on the signal, the noises depending on the space domain, or spectral domain noise and mixed noise. Therefore, it is

necessary to construct a model for suppressing complex mixed noises in order to deal with real HSI scenarios.

To alleviate the above limitations, a variational low-rank matrix factorization model, combined with multi-patch collaborative learning (VLRMFmcl), is proposed in the Bayesian framework to suppress various noises in HSIs. The main contributions of this work are summarized as follows.

(1) Multi-patch collaborative learning is exploited to effectively depict and learn the spatial consistency and the nonlocal self-similarity in the HSI. The pixels within the same collaborative patches share similar spatial–spectral characteristics, which are utilized to effectively improve the performance of denoising the patches polluted by heavy noises.

(2) Variational low-rank matrix factorization is proposed to learn and characterize the collaborative patch data by exploring latent characteristics across the spatial–spectral domain, in which the latent clean image in degraded HSI has the property of low rank and the mixed noises do not. The Gaussian distribution with zero mean and the variance adjusted by gamma distribution is explored to represent the latent clean image. The Dirichlet process Gaussian mixture is exploited to depict the inherent statistical features of different noises in the HSI, which are adapted and learned by exploring the Gaussian distribution, the inverse Wishart distribution, and the Dirichlet process. Through this process, the underlying mixed noise of the HSI can be fit adaptively without needing to know specific noise types or intensity.

(3) Considering the uncertainty of information about latent variables, the posteriors of the latent clean image and the mixed noises are both explicitly parameterized and update in a closed form by utilizing the variational inference. The feasibility and validity of the VLRMFmcl method are evaluated under different experimental conditions. Compared with several popular denoising methods, VLRMFmcl can reduce the noises of the hyperspectral images while preserving the structural information.

The paper is organized as follows. Section 2 gives a detailed description of the proposed restoration model, which is performed using variational inference. In Section 3, several experimental results are presented by utilizing the real-world HSI datasets. The conclusions are given in Section 4.

2. Proposed Model

To effectively suppress the various noises in HSIs, multi-patch collaborative learning is explored to represent the intrinsic spatial consistency and non-local self-similarity of HSIs. Then, the learned patches are input into the variational low-rank matrix factorization model, which is developed to suppress the mixed noises of each patch in the Bayesian framework. Figure 1 presents the framework of the proposed VLRMFmcl.

2.1. Multi-Patch Collaborative Learning

In HSIs, the adjacent pixels have high consistency in the space domain [26–29]. Based on this fact, they are often divided into overlapping three-dimensional patches for the HSI analysis. The effective information of one patch is very scarce when most of the pixels within this patch are polluted by a large amount of noise. Therefore, it is very difficult to recover the noiseless data by directly exploiting this image patch. To solve these problems, it becomes important to effectively utilize the patches in HSIs.

Figure 2 presents some patches from the Pavia Centre data (presented in Section 3.2), in which the area marked by the red box represents a test patch, and the ones marked by three green boxes represent its neighboring patches. The similarities between the patches marked by the red box and the green boxes are very different. Inspired by this basic characteristic in the HSI, the heavily polluted patches could be restored by the patches with high similarities with them. Additionally, it has been argued that the “collaborative” nature of the approximation can improve classification accuracy [46]. Considering that the HSI denoising aims to facilitate subsequent applications (e.g., classification), a similar “collaborative” [46] nature is introduced. Based on these, a multi-patch collaborative

learning strategy is proposed by exploring the similarities between different image patches to effectively learn about the HSI.

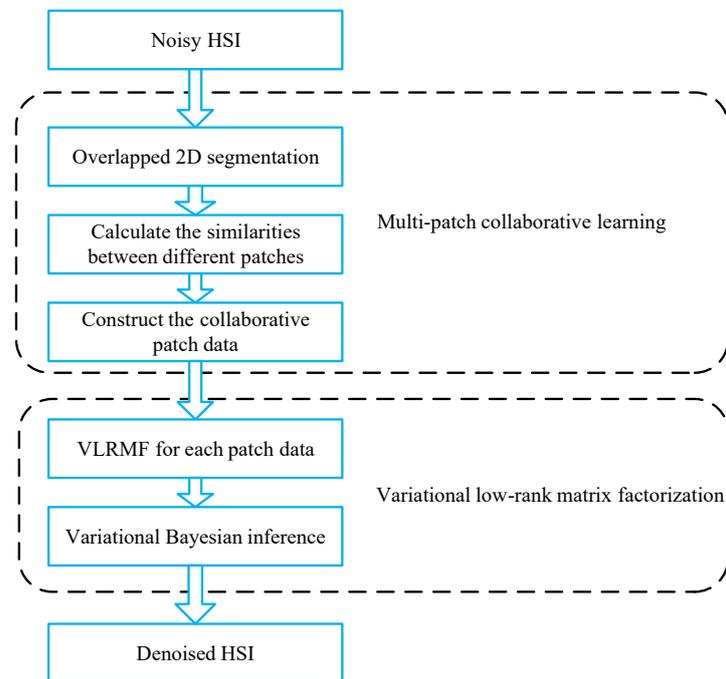


Figure 1. The framework of the proposed variational low-rank matrix factorization with multi-patch collaborative learning (VLRMFmcl) method. HSI, hyperspectral image.

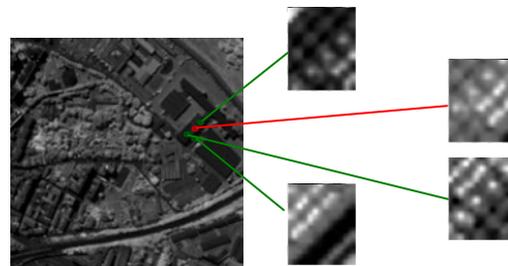


Figure 2. Some patch structures of the Pavia Centre data. Four areas are enlarged to better show the similarities of different patches. The area marked by the red box represents a test patch, and the ones marked by three green boxes are the neighboring patches of the test patch.

Supposing the hyperspectral image \mathbf{X} , it is segmented into overlapped three-dimensional patches with the size of $d \times d \times \lambda$, where d represent the spatial size of patches and λ is the total number of bands in the HSI. For a pixel x_i from \mathbf{X} , the collection $\mathbf{N}(x_i)$, $1 \leq i \leq d^2$ consists of all pixels of a patch centered at the sample x_i . The pixels in $\mathbf{N}(x_i)$ can be considered to contain similar characteristics. The \mathbf{y}_i is formed by stacking all pixels of $\mathbf{N}(x_i)$ into a vector, which can be regarded as the fusing feature of x_i . The similarities between the different patch data can be formulated as

$$\text{SimilarIndex}(\mathbf{y}_i, \mathbf{y}_j) = \exp\left(-\left(\frac{1}{d^2\lambda} \left| \mathbf{I}_{1 \times d^2\lambda} \mathbf{y}_i - \mathbf{I}_{1 \times d^2\lambda} \mathbf{y}_j \right| \right)\right), \quad (1)$$

where $\mathbf{I}_{1 \times d^2\lambda}$ is the row vector with the dimension of $d^2\lambda$, of which the elements are equal to one. Obviously, the larger the value of the $\text{SimilarIndex}(\mathbf{y}_i, \mathbf{y}_j)$ is, the higher the similarities that are observed between $\mathbf{N}(x_i)$ and $\mathbf{N}(x_j)$. According to Equation (1), we can select the most similar $(P-1)$ patches to the patch centered at x_i , and construct the non-local

patch data Y_i . When d is large enough, it can be considered as all the most similar data to be searched in the whole hyperspectral image.

2.2. Variational Low-Rank Matrix Decomposition

In the existing literature, many computer vision, machine learning, and statistical problems can be approached by solving and learning a low-dimensional linear model. In this case, the low-rank matrix decomposition has been widely concerned and applied in many fields [3–6], which can effectively explore the low-dimensional properties of the observed data. Assuming $X = [y_1, \dots, y_H] \in R^{M \times H}$ represents the observed data, M and H represent the spatial size of X , the general low-rank matrix decomposition model is formulated as

$$X = UV^T + n, \quad (2)$$

where $U = [u_1, \dots, u_L] \in R^{M \times L}$ and $V = [v_1, \dots, v_H]^T \in R^{H \times L}$ represent the decomposition matrix, $L \ll \min(M, H)$; n is the noise, which is depicted by Gaussian distribution, Laplace distribution, or polynomial distribution.

Obviously, the pixels of the same collaborative patch have similar characteristics in the spatial and spectral domains. In other words, these pixels have the low-rank property and can be effectively learned and expressed by the low-rank matrix decomposition. Additionally, the hyperspectral images are usually polluted by various noises with different statistical properties. The Gaussians mixture model can effectively learn and depict the different noises, including Gaussian noise, sparse noise, and so on. Above all, the noise model is explored to depict the complex noises in real HSIs, in which the Dirichlet process is utilized to adaptively achieve the selection of Gaussian distribution and the determination of the number of Gaussian distributions. The symbol $Y = \{y_i\}_{i=1}^P$ represents the collaborative patch data, and $M = d^2\lambda$ represents the dimensions of the sample y_i . According to (2), the proposed Bayesian low-rank matrix decomposition model for denoising the collaborative patch data can be written as follows:

$$Y = UV^T + n. \quad (3)$$

The first term is the low-rank decomposition term, in which $u_i \in R^M$ and $v_j \in R^L$ are defined as $u_i \sim N(0, \tau_{ui}^{-1}I)$ and $v_j \sim N(0, \tau_{vj}^{-1}I)$ separately. That is, u_i and v_j are drawn from the Gaussian distribution with zero mean and variances of τ_{ui} and τ_{vj} , individually. In order to improve the model robustness and reduce the sensitivity of parameters, the gamma distribution is introduced to adaptively adjust the parameter τ_{ui} and τ_{vj} . The first term can be formulated as:

$$\begin{aligned} u_i &\sim N(0, \tau_{ui}^{-1}I), & v_j &\sim N(0, \tau_{vj}^{-1}I) \\ \tau_{ui} &\sim \Gamma(a_0, b_0), & \tau_{vj} &\sim \Gamma(c_0, d_0) \end{aligned} \quad (4)$$

where I is the column vector, of which the entries are all equal to one; a_0 , b_0 , c_0 and d_0 represent the hyper-parameters of the gamma distribution.

The second term in Equation (3) represents the mixed noises in the real HSI. Considering the complex statistical properties of the noises in the HSI, the Gaussians mixture model is utilized to depict the different noises, which are displayed as

$$\begin{aligned} n_i &\sim \prod_{k=1}^{\infty} N(\mu_k, \Sigma_k)^{z_{ijk}} \\ \mu_k &\sim N(\mu_0, \Sigma_0) & \Sigma_k &\sim iWishart(e_0, f_0) \\ z_{ij} &\sim Mult(\pi) & \pi_t &\sim v_t \prod_{j=1}^{t-1} (1 - v_j) & v_t &\sim beta(1, \beta) \end{aligned} \quad (5)$$

In Equation (5), μ_k and Σ_k are the mean and variance of the k -th Gaussian distribution, which are learned and represented by the Gaussian distribution and the inverse Wishart distribution. The Gaussian distribution and the inverse Wishart distribution are conjugate.

μ_0 and Σ_0 represent the mean and variance of the parameter μ_k ; and e_0 and f_0 are the freedom degree and scale matrix of the inverse Wishart distribution. To effectively depict the various noises in a data-driven way, the indicator variable $z_{ij} \in \{0, 1\}^K$, $K \rightarrow \infty$ is introduced to determine and learn the number and mode of the Gaussian distribution, and $\sum_k z_{ijk} = 1$. z_{ij} is drawn from the polynomial distribution with the parameter π , which is learned through the Dirichlet process. Figure 3 shows the graphical representation of the Bayesian low-rank matrix decomposition model.

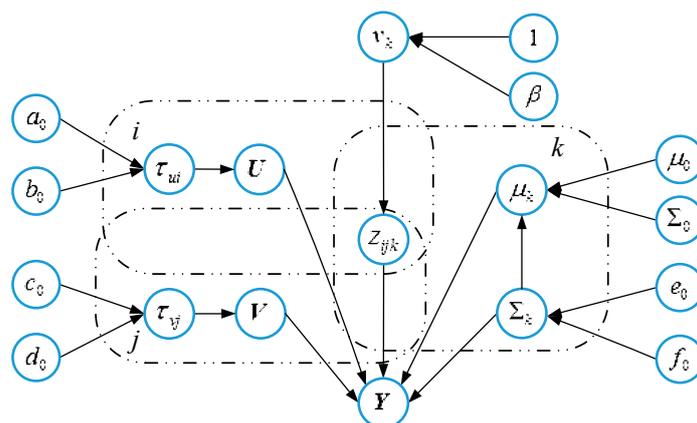


Figure 3. Graphical representation of the variational low-rank matrix decomposition model.

Additionally, $Y = \Delta \circ Y$ is introduced into Equation (3) when recovering the missing pixels. $\Delta = \{0, 1\}^{M \times P}$ is the sampling matrix whose elements are equal to 0 or 1. Therefore, $\Sigma_{f i} = 0$ ($f = 1, \dots, M$) represents the loss of the f -th element in y_i when acquiring; $\Sigma_{f i} = 1$ ($f = 1, \dots, M$) means that the f -th element in y_i is effectively collected.

2.3. Variational Bayesian Inference

According to Equations (3) and (4) and Figure 3, it can be observed that all the variances in the proposed Bayesian low-rank matrix decomposition model satisfy the conjugation. Therefore, variational Bayesian inference can be used to solve the model. Assuming the symbol $\Psi = \{u_i, v_i, \mu_k, \Sigma_k, z_i, \tau_{ui}, \tau_{vj}, v_t\}$ represents the variables of the proposed model; the symbol $\Theta = \{a_0, b_0, c_0, d_0, \mu_0, \Sigma_0, e_0, f_0, \beta\}$ is the corresponding hyperparameter in the model. The variational Bayesian inference can be achieved by estimating the posterior distribution of the latent variable Ψ with the observation data Y and the hyperparameter Θ given. To solve the proposed model, the real posterior distribution $p(\Psi|Y, \Theta)$ of the latent variables Ψ is approximated based on the distribution $q(\Psi)$. Then, we can obtain

$$\begin{aligned}
 \ln p(Y|\Theta) &= \ln \int p(Y, \Psi|\Theta) d\Psi \\
 &= \ln \int \frac{p(Y, \Psi|\Theta)q(\Psi)}{q(\Psi)} d\Psi \\
 &\geq \int \ln \frac{p(Y, \Psi|\Theta)q(\Psi)}{q(\Psi)} d\Psi \\
 &= \ln p(Y|\Theta) - \text{KL}(q(\Psi)|p(\Psi|Y, \Theta))
 \end{aligned}
 \tag{6}$$

where $\text{KL}(q(\Psi)|p(\Psi|Y, \Theta))$ is utilized to represent the KL divergence distance between the variational approximation $q(\Psi)$ and the true joint probability distribution $p(\Psi|Y, \Theta)$. It can be easily seen that the expression $\ln p(Y|\Theta)$ has a strict lower bound because $\text{KL}(q(\Psi)|p(\Psi|Y, \Theta)) \geq 0$. Therefore, the optimal solution of the proposed model can be calculated by minimizing $\text{KL}(q(\Psi)|p(\Psi|Y, \Theta))$. Algorithm 1 presents the pseudocode of the VLRMFmcl method.

Algorithm 1. The VLRMFmcl Method

Input: the noisy HSI image X ; the spatial size d of patches; the total number λ of bands; the hyperparameter $\Theta = \{a_0, b_0, c_0, d_0, \mu_0, \Sigma_0, e_0, f_0, \beta\}$;

Output: the denoised image Y ;

Process Source-Code of multi-patch collaborative learning:

Divide X into the overlapping patches with the size of $d \times d \times \lambda$;

for each pixel x_i in X **do**

Obtain the collection $N(x_i)$ and y_i , where $1 \leq i \leq d^2\lambda$

Calculate the similarities *Similar Index*(y_i, y_j) between $N(x_i)$ and $N(x_j)$;

Select the most similar ($P-1$) patches to the patch centered at x_i , and construct the collaborative patch data Y_i ;

Process Source-Code of variational low-rank matrix factorization:

Obtain the variables $\Psi = \{\mathbf{u}_i, \mathbf{v}_i, \mu_k, \Sigma_k, z_i, \tau_{ui}, \tau_{vj}, \mathbf{v}_t\}$;

Calculate $p(\Psi | Y, \Theta)$ by $q(\Psi)$; update Ψ by minimizing $\text{KL}(q(\Psi) | p(\Psi | H, \Theta))$;

return denoised image Y ;

The updating equations of the model variances are listed as follows.

(1) updating z_{ij} and \mathbf{v}_t :

The posterior probability of \mathbf{v}_t is still drawn from the beta process. Supposing $\mathbf{v}_t \sim \text{beta}(g_t, h_t)$, g_t and h_t can be calculated by

$$g_{tk} = 1 + \sum_i \sum_j q_{ijk}(t), h_{tk} = \beta + \sum_{j=t+1}^T \sum_i q_{ijk}(t). \quad (7)$$

For the variance z_{ij} ,

$$q(z_{ij}) = \prod_k q_{ijk}^{z_{ijk}}, \quad (8)$$

where

$$q_{ijk}(t) = \frac{\rho_{ijk}(t)}{\sum_{l=1}^T \rho_{ijk}(l)} \quad (9)$$

$$\rho_{ijk}(l) = \exp(\gamma_{l,1}^{ik} + \gamma_{l,2}^{ik})$$

Supposing Φ is the digamma function, $\gamma_{l,1}^{ik}$ and $\gamma_{l,2}^{ik}$ are expressed as:

$$\gamma_{l,1}^{ik} = \Phi(g_{lk}) + \Phi(h_{lk})$$

$$\gamma_{l,2}^{ik} = -0.5 \text{tr} \left\{ \left\langle \Sigma_l^{-1} \right\rangle \left[\langle \mu_l \mu_l^T \rangle + y_{ij} y_{ij}^T + \langle \mathbf{v}_i \mathbf{v}_i^T \rangle \langle \mathbf{U}^T \mathbf{U} \rangle \right. \right. \\ \left. \left. - 2 \langle \mathbf{U} \rangle \langle \mathbf{v}_i \rangle (y_{ij} - \mu_l)^T - 2 y_{ij} \langle \mu_l^T \rangle \right] + \ln |\Sigma_l| \right\} \quad (10)$$

(2) updating μ_k :

The posterior probability of μ_k is still drawn from the Gaussian distribution, which satisfies the following conditions:

$$\langle \mu_i \rangle = [\Sigma_0^{-1} + \sum_{i,k} q_{ijk}(t) \langle \Sigma_i^{-1} \rangle]^{-1} [\Sigma_0^{-1} \mu_0 \\ + \sum_{i,k} q_{ijk}(t) \langle \Sigma_i^{-1} \rangle (y_{ij} - \langle \mathbf{U} \rangle \langle \mathbf{v}_i \rangle)] \quad (11)$$

$$\langle \mu_i \mu_i^T \rangle = [\Sigma_0^{-1} + \sum_{i,k} q_{ijk}(t) \langle \Sigma_i^{-1} \rangle]^{-1} + \langle \mu_i \rangle \langle \mu_i \rangle^T$$

(3) updating Σ_k :

The posterior probability of Σ_k is still drawn from the inverse-Wishart distribution, which satisfies the following conditions:

$$e_t = e_0 + \zeta_t$$

$$f_t = f_0 + 0.5 \sum_i \sum_j q_{ijk}(t) (\langle \mu_i \mu_i^T \rangle + y_{ij} y_{ij}^T + \langle \mathbf{u}_i \mathbf{u}_i^T \rangle \langle \mathbf{v}_j^T \mathbf{v}_j \rangle \\ - 2(y_{ij} - \mu_t) \langle \mathbf{U} \rangle \langle \mathbf{v}_j \rangle - 2 \langle \mu_i^T \rangle y_{ij}) \quad (12)$$

From Equation (12), we can obtain the following expression:

$$\begin{aligned} \langle \Sigma_t^{-1} \rangle &= e_t(f_t)^{-1} \\ \langle \ln|\Sigma_t| \rangle &= \frac{1}{\Phi(0.5e_t) + d^2\lambda \ln 2 + \ln|f_t^{-1}|} \end{aligned} \quad (13)$$

(4) updating u_i :

The posterior probability of u_i is still drawn from the Gaussian distribution, which is formulated as

$$\mathbf{u}_i \sim N(\boldsymbol{\mu}_{u_i}, \boldsymbol{\Omega}_{u_i}^{-1}), \quad (14)$$

where the mean $\boldsymbol{\mu}_{u_i}$ and the variance $\boldsymbol{\Omega}_{u_i}$ are shown as

$$\begin{aligned} \boldsymbol{\Omega}_{u_i} &= \langle \tau_{ui} \rangle I + \sum_k \boldsymbol{\Sigma}_k^{-1} \sum_j \langle z_{ijk} \rangle \langle \mathbf{v}_j^T \mathbf{v}_j \rangle \\ \boldsymbol{\mu}_{u_i} &= \boldsymbol{\Omega}_{u_i}^{-1} \sum_k \boldsymbol{\Sigma}_k^{-1} \sum_j \langle z_{ijk} \rangle (\mathbf{y}_{ij} - \langle \boldsymbol{\mu}_k^{-1} \rangle) \langle \mathbf{v}_j \rangle \end{aligned} \quad (15)$$

(5) updating τ_{ui} :

The posterior probability of τ_{ui} is still drawn from the gamma distribution, which satisfies the following conditions:

$$\tau_{ui} \sim \Gamma(a, b_i), \quad (16)$$

where the parameters a and b_i are shown as

$$a = a_0 + 0.5LM, \quad b_i = b_0 + 0.5 \langle \mathbf{u}_i \mathbf{u}_i^T \rangle. \quad (17)$$

(6) updating v_j :

The posterior probability of v_j is still drawn from the Gaussian distribution, which can be written as follows:

$$\mathbf{v}_j \sim N(\boldsymbol{\mu}_{v_j}, \boldsymbol{\Omega}_{v_j}^{-1}), \quad (18)$$

where the mean $\boldsymbol{\mu}_{v_j}$ and the variance $\boldsymbol{\Omega}_{v_j}$ are:

$$\begin{aligned} \boldsymbol{\Omega}_{v_j} &= \langle \tau_{vj} \rangle I + \sum_k \boldsymbol{\Sigma}_k^{-1} \sum_i \langle z_{ijk} \rangle \langle \mathbf{u}_i \mathbf{u}_i^T \rangle \\ \boldsymbol{\mu}_{v_j} &= \boldsymbol{\Omega}_{v_j}^{-1} \sum_k \boldsymbol{\Sigma}_k^{-1} \sum_i \langle z_{ijk} \rangle (\mathbf{y}_{ij} - \langle \boldsymbol{\mu}_k^{-1} \rangle) \langle \mathbf{u}_i \rangle \end{aligned} \quad (19)$$

(7) updating τ_{vj} :

The posterior probability of τ_{vj} is still drawn from the gamma distribution, which satisfies the following conditions:

$$\tau_{vj} \sim \Gamma(c, d_j), \quad (20)$$

where the parameters c and d_j are shown as

$$c = c_0 + 0.5LP, \quad d_j = d_0 + 0.5 \langle \mathbf{v}_j^T \mathbf{v}_j \rangle. \quad (21)$$

3. Experiments

To validate the effectiveness of the proposed VLRMFmcl model, three popular hyperspectral images were chosen as the experimental datasets, the Beads, Pavia Centre, and Urban datasets. In addition, BM3D [11], ANLM3D [14], BM4D [17], LRMR [25], GLF [26], LRTDTV [30], LTDL [32], DnCNN [42], and HSID-CNN [43] were chosen as the compared methods. The necessary parameters in the BM3D, ANLM3D, and LRTDTV methods, were automatically/manually adjusted to generate the optimal denoising results, as the references suggested. In BM4D, the noise variance was selected from the set {0.01, 0.03, 0.04, 0.05, 0.07, 0.09, 1.1}. In LRMR, the rank of the noiseless matrix was chosen from {3, 5, 6, 7, 9,

11}, and the cardinality of the sparse term was chosen from the set {0, 500, 1000, 1500, 2000, 3000, 4000, 5000}. In GLF, the number of subspaces was chosen from the set {5, 8, 9, 11, 13}. In LTDL, the noise variance was selected from the set {0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4}. In the DnCNN and HSID-CNN methods, the pre-trained weights and the related settings were utilized to conduct the experiments as the references suggested.

In addition, five metrics were chosen to numerically evaluate the denoising performance of the different algorithms—the peak signal-to-noise ratio (PSNR), feature similarity (FSIM) [47], the mean spectral angle (MSA), noise reduction (NR) [48,49], and the mean relative deviation (MRD) [48,49]. At the same time, the visual effect was utilized as an intuitive way to determine the denoising performance. Suppose I^{den} and I^{ref} represent the denoised and reference images, respectively; I_1 and I_2 represent the spatial size of the image.

(a) The greater the value of PSNR is, the better the denoising image quality is. PSNR (in dB) is formulated as:

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2 * I_1 I_2}{\|I^{ref} - I^{den}\|^2} \right) \quad (22)$$

(b) The greater the value of FSIM is, the better the denoising image quality is. FSIM is formulated as:

$$\text{FSIM} = \frac{\sum_{x=\Omega} S_l(x) * PC_m(x)}{\sum_{x=\Omega} PC_m(x)}, \quad (23)$$

where $S_l(x)$ is derived from the phase congruency and the image gradient magnitude of I^{den} and I^{ref} ; $PC_m(x)$ is the maximum phase congruency of PC_{den} (for I^{den}) and PC_{ref} (for I^{ref}); and Ω represents the entire airspace of the image.

(c) MSA was used to estimate spectral fidelity between the denoising images and reference images in the spectral domain. The smaller the value of MSA is, the better the spectral fidelity of the restored algorithms. The MSA is calculated by:

$$\text{MSA} = \frac{1}{I_1 I_2} \sum_{i=1}^{I_1} \sum_{j=1}^{I_2} \cos^{-1} \left(\frac{I_{ij}^{den} \cdot I_{ij}^{ref}}{\|I_{ij}^{den}\| \cdot \|I_{ij}^{ref}\|} \right) \quad (24)$$

(d) NR was used to evaluate the noise reduction of different restored methods in the frequency domain. The greater the value of NR is, the better the performance of the denoising algorithms. NR is formulated as

$$\text{NR} = N_0 / N_1, \quad (25)$$

where N_1 is the power of the frequency components generated by stripes in the restored image and N_0 is for the reference image. N_1 and N_0 can be obtained by

$$N_c = \sum_{\varphi} P_c(D), \quad (26)$$

where $P_c(D)$ is the averaged power spectrum down the columns of an image with D being the distance from its reference image in Fourier space, and φ is the stripe noise region of the spectrum.

(e) MRD was utilized to compare the degree of distortion between the selected noiseless region of the restored images and reference images. The smaller the value of MRD is, the smaller the image distortion. In the experiment, a 10×10 window was selected to calculate the MRD value. MRD is formulated as:

$$\text{MRD} = \frac{1}{I_1 I_2} \sum_{i=1}^{I_1} \sum_{j=1}^{I_2} \frac{|I_{ij}^{den} - I_{ij}^{ref}|}{I_{ij}^{ref}} \times 100\%. \quad (27)$$

3.1. Experiment on the Beads Data Set

The Beads data set, acquired from Columbia University, has a spectral resolution of 10 nm and a spectral range from 400 nm to 700 nm. The data set has a total number of 31 consecutive spectral bands. The spatial resolution in each band is 512×512 .

Three kinds of noises were considered in the simulation experiment. The detailed descriptions are listed as follows.

(1) Gaussian white noise with the mean 0 and fixed variance.

(2) Poisson noise is added by adjusting the ratio between the maximum brightness and the initial image brightness, which can be written as $X_{poission} = X \cdot peak$. $X_{poission}$ represents the image polluted by Poisson noise; X is the initial image data; $peak$ refers to the intensity of the Poisson noise. To reduce the passion noise, we utilized settings similar to those in [50]. The variance-stabilizing transformation (VST) was utilized to convert Poisson noise into Gaussian noise before implementing a denoising approach. The final denoising images were obtained by the inverse variance stability transformation.

(3) Sparse noise is added to the randomly selected pixels by utilizing uniform distribution with the interval $[-10, 10]$.

Mixed noise, consisting of zero-mean Gaussian noise with variance $\sigma = 0.1$ and Poisson noise with $peak = \{5, 10, 20, 30, 50, 70, 100, 130, 160\}$, was added to the Beads data. Then the nine compared methods and the proposed one were utilized to restore the noisy Beads data. The performance curves of the simulation experiments are shown in Figure 4, where the vertical coordinates represent the values of PSNR, FSIM, and MSA, respectively. The horizontal coordinates represent the value of parameter $peak$. Comparing the curves of PSNR, FSIM, and MSA, it can be clearly observed that both the PSNR and FSIM values of the VLRMFmcl method were higher than those of BM3D, ANLM3D, BM4D, LRMR, GLF, LRTDTV, DnCNN, and HSID-CNN methods. At the same time, the MSA of VLRMFmcl was lower than these eight compared algorithms. Compared with LTDL, the proposed model is superior in PSNR. For FSIM and MSA, it showed better values than LTDL in most cases. These facts indicate that VLRMFmcl can effectively improve the quality of the noisy HSI by better maintaining the image feature information and restoring the spectral information in the HSI. In addition, the performance curve of the VLRMFmcl method is smoother than the nine comparison algorithms, which means VLRMFmcl is more stable when denoising the HSI.

Figure 5 shows the restored images of band 27 obtained by different models, which are polluted by Gaussian noise, sparse noise, and missing pixels. Compared with the noisy image in Figure 5b, the quality of Figure 5c–l is significantly improved. According to Figure 5, it can be seen that VLRMFmcl can effectively reduce the various forms of noise in the HSI with a large difference in brightness. The denoising results can preserve the structural information and the edges of the homogeneity region. Using patch-matching three-dimensional filtering, BM3D smoothed out some feature structures and blurred the visual effect while suppressing the different noises and restoring the missing pixels. ANLM3D showed better visual performance than BM3D and DnCNN; however, ANLM3D had a weaker ability to recover the detailed information in the HSI. The restored images obtained by BM4D were too smooth and lost some information. As shown in Figure 5f, the results of the LRMR method were significantly better than BM3D, ANLM3D, and DnCNN, but the results of LRMR still showed obvious sparse noise. Utilizing low-rank factorization of tensors constructed by nonlocal similar 3D patches, GLF was able to recover the basic shapes of the Beads dataset, but its result lacked sharpness. As shown in Figure 5, the results of LRTDTV, LTDL, and VLRMFmcl were much better than those of BM3D, ANLM3D, BM4D, LRMR, GLF, DnCNN, and HSID-CNN. In general, the VLRMFmcl method can remove the mixed noises and restore the missing pixels of the Beads data.

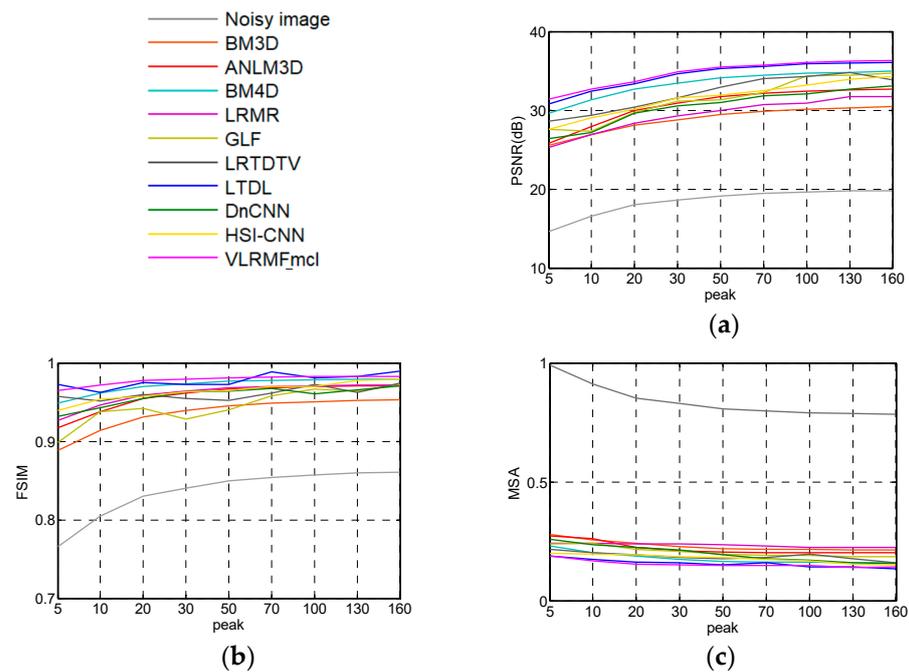


Figure 4. Quantitative evaluation results for Beads data versus different noises: (a) peak signal-to-noise ratio (PSNR); (b) feature similarity (FSIM); (c) mean spectral angle (MSA).

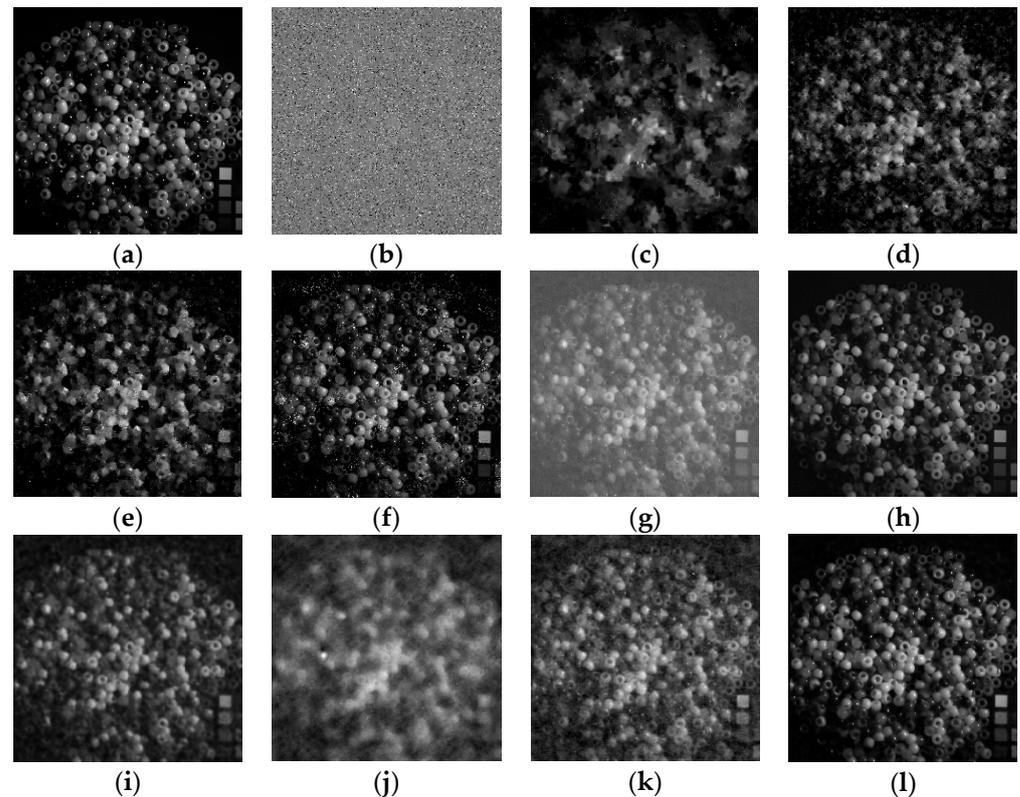


Figure 5. Restored images of band 27 corrupted with Gaussian noise, sparse noise and missing pixels: (a) clean HSI; (b) noisy HSI; (c) BM3D; (d) ANLM3D; (e) BM4D; (f) LRM3; (g) GLF; (h) LRTDTV; (i) LTDL; (j) DnCNN; (k) HSI-CNN; (l) VLRMFmcl.

To make a more intuitive comparison of different algorithms, Figure 6 shows the pseudo-color images of the restored images (R: 3, G: 12, B: 25). As can be seen in Figure 6, the denoising results of VLRMFmcl were better than those of BM3D, ANL3D, BM4D,

LRMR, GLF, LRTDTV, LTDL, DnCNN, and HSID-CNN. Additionally, the restored results of VLRMFmcl were very similar to those of the reference images, which can be easily observed by comparing Figure 5a,l.

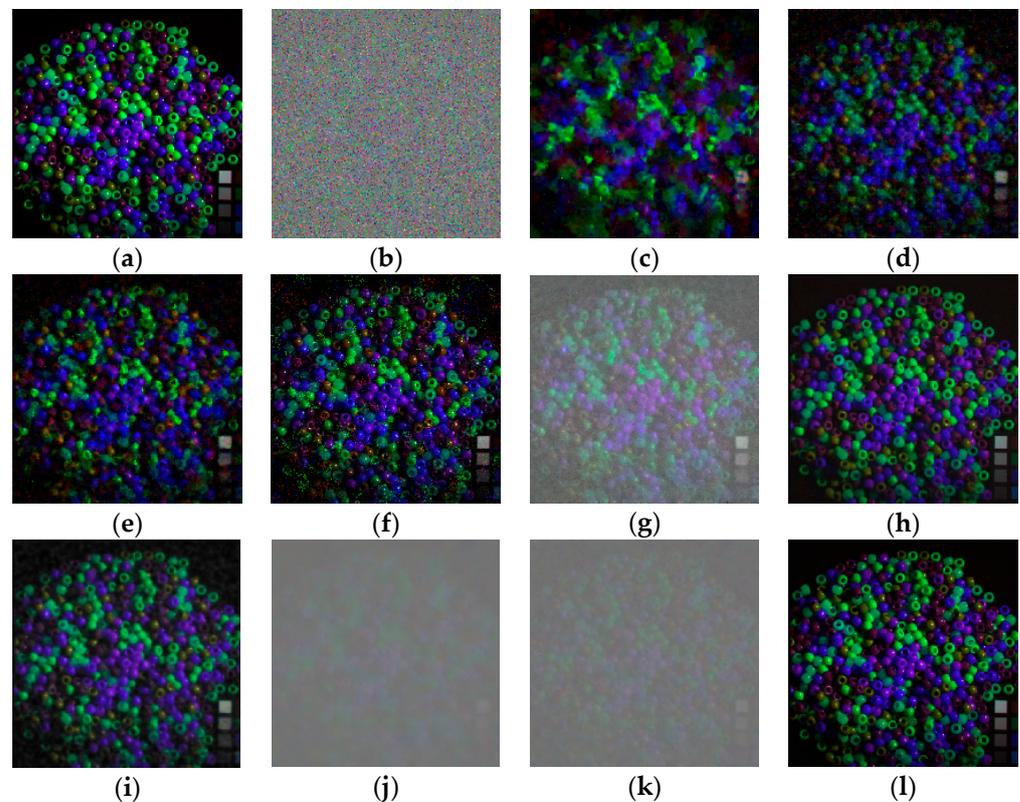


Figure 6. Restored images of Beads corrupted with Gaussian noise, sparse noise and missing pixels: (a) clean HSI; (b) noisy HSI; (c) BM3D; (d) ANLM3D; (e) BM4D; (f) LRMR; (g) GLF; (h) LRTDTV; (i) LTDL; (j) DnCNN; (k) HSID-CNN; (l) VLRMFmcl.

3.2. Experiment on the Pavia Centre Dataset

The Pavia Centre dataset was acquired from the Reflective Optics System Imaging Spectrometer. It contains 115 bands, and each band consists of 1096×715 pixels. The spectral range of the Pavia Centre dataset is from 0.43 micrometers to 0.86 micrometers. After removing 13 noisy bands, the remaining 102 bands were used in the following analysis. In the experiment, a subset of the Pavia Centre with the size of $400 \times 400 \times 102$ was used.

Three kinds of noises were considered for the Pavia Centre dataset: (1) Gaussian white noise with the mean 0 and the noise variance $\sigma = 0.1$. (2) Sparse noise was added to the randomly selected pixels by utilizing the uniform distribution with the interval $[-5, 5]$. (3) Deadlines were added to the same position as the selected bands in the HSI. Their width varied from one line to three lines.

Table 1 shows the evaluation results for PSNR, FSIM, and MSA calculated by different approaches. All the bold numbers in Table 1 indicate the best results. By utilizing the nonlocal self-similarity and adaptively learning the noises in the HSI, the VLRMFmcl method shown the best PSNR/MSA values and the suboptimal FSIM value than the compared methods. Compared with the noisy image, the PSNR and FSIM values obtained by VLRMFmcl were increased by 21.66 and 0.1443, respectively, and MSA was reduced by 0.264. HSID-CNN simultaneously assigned the spatial information and adjacent correlated bands to the network, where multiscale feature extraction was employed to capture both the multiscale spatial feature and spectral feature. Its FSIM value was optimal. Instead of learning the noise variance, BM3D and DnCNN denoised the HSI with the predefined fixed noise variance band by band, which could not efficiently utilize the spectral correlations of

the HSI. As shown in Table 1, the PSNR, FSIM, and MSA values of BM3D and DnCNN were significantly lower than those of the other methods while reducing the mixed noises. Noticing that, ANLM3D denoised the HSI by using the high nonlocal self-similarity and making a balance between the smoothing and details preservation; BM4D adopted the 3D nonlocal self-similarity data cube to exploit the local correlations between some neighboring bands; and GLF reduced the mixed noise by utilizing low-rank factorization of tensors constructed by nonlocal similar 3D patches. In Table 1, the ANLM3D, BM4D, and GLF methods presented relatively good results by exploring the spatial–spectral information of the HSI. LRM, LRTDTV, and LTDL took advantage of the low-rank property in HSI, and their PSNR and FSIM values were better than those of BM3D, ANLM3D, GLF, and DnCNN.

Table 1. Quantitative evaluation results for the Pavia Centre data set.

	Band 109	BM3D	ANLM3D	BM4D	LRMR	GLF	LRTDTV	LTDL	DnCNN	HSID-CNN	VLRMFmcl
PSNR(dB)	13.97	17.92	30.25	33.65	34.57	34.12	34.91	35.06	22.79	35.57	35.63
FSIM	0.8458	0.7903	0.9117	0.9739	0.9835	0.9702	0.9832	0.9551	0.8125	0.9907	0.9901
MSA	0.3616	0.3019	0.1101	0.1014	0.1181	0.1067	0.1025	0.0991	0.2216	0.0983	0.0976

Figure 7 shows the results of band 90 obtained by different denoising approaches. To make a better visual evaluation, Figure 8 shows the comparison of the pseudo-color images (R: 60, G: 30, B 2). It can be seen that the image qualities of the BM3D, ANLM3D, BM4D, LRM, GLF, LRTDTV, TDTL, DnCNN, HSID-CNN, and VLRMFmcl methods were significantly improved compared to the noise images as shown in Figures 7b and 8b. As can be seen in Figures 7c and 8c, the denoising results obtained by the BM3D method were relatively fuzzy, and the method could not effectively inhibit the strip noise. ANLM3D, BM4D, LRM, GLF, LRTDTV, LTDL, DnCNN, and HSID-CNN could only suppress some of the noise. It can be easily seen in Figures 7l and 8l that the proposed VLRMFmcl model was able to effectively suppress Gaussian noise, sparse noise, and deadlines, and its results were better than those of the compared methods.

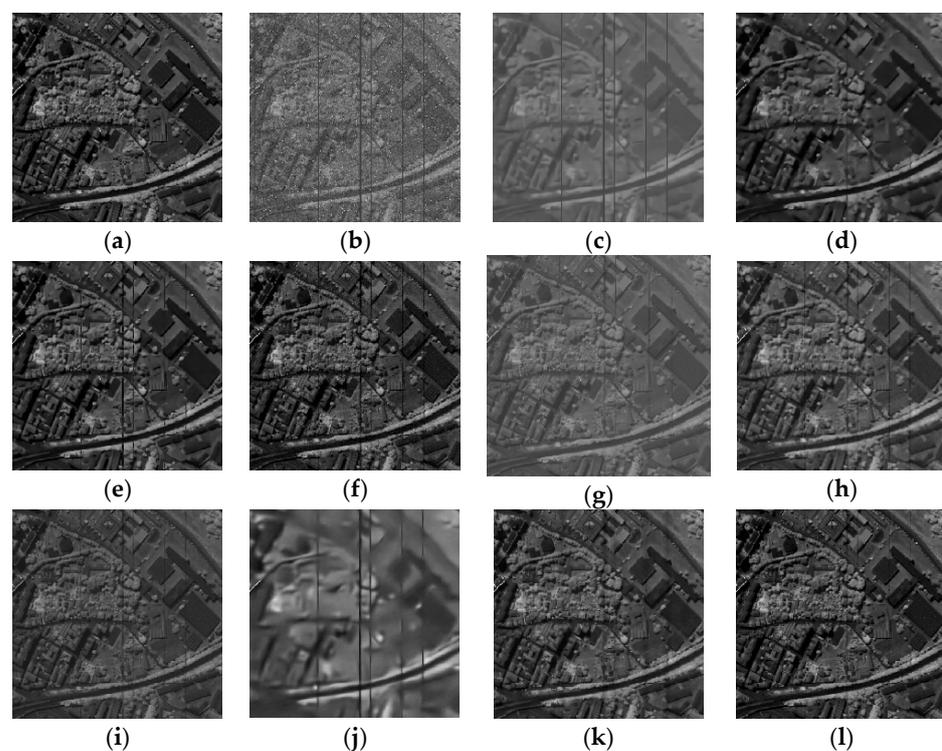


Figure 7. Restored results of band 90 corrupted with mixed noises: (a) clean HSI; (b) noisy HSI; (c) BM3D; (d) ANLM3D; (e) BM4D; (f) LRM; (g) GLF; (h) LRTDTV; (i) LTDL; (j) DnCNN; (k) HSID-CNN; (l) VLRMFmcl.

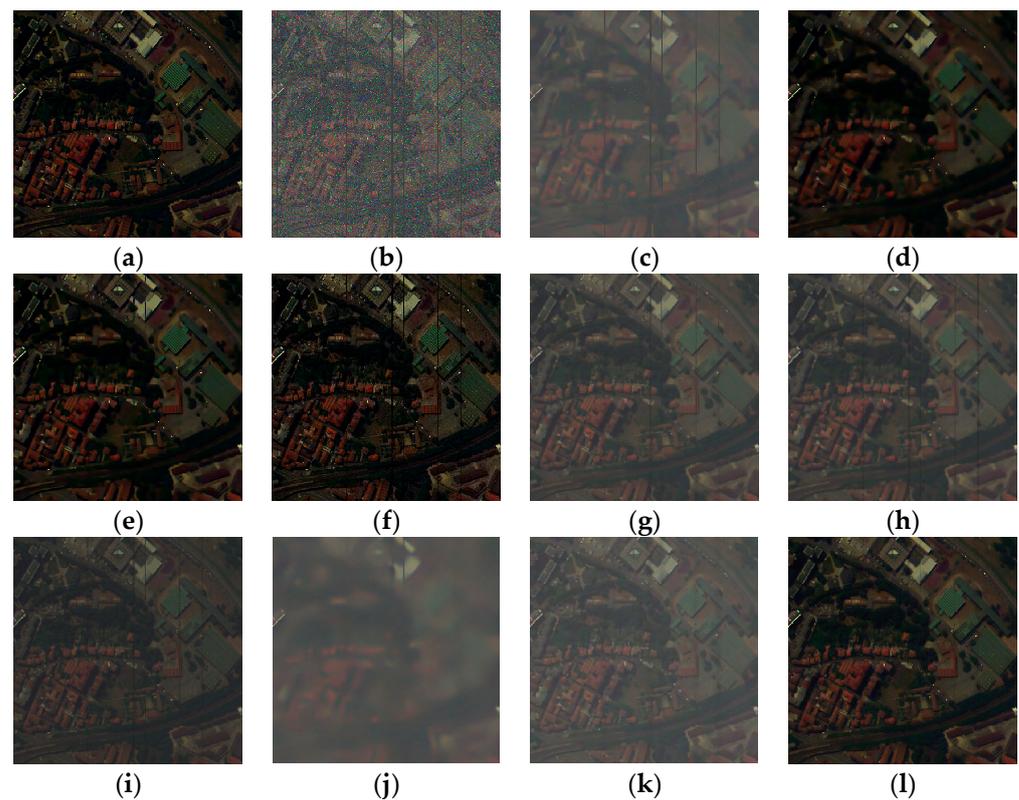


Figure 8. Restored results of the Pavia Centre data corrupted with mixed noises: (a) clean HSI; (b) noisy HSI; (c) BM3D; (d) ANLM3D; (e) BM4D; (f) LRM; (g) GLF; (h) LRTDTV; (i) LTDL; (j) DnCNN; (k) HSID-CNN; (l) VLRFmcl.

3.3. Experiment on the Urban Dataset

Urban data, with a size of $307 \times 307 \times 210$, was acquired from the HYDICE sensor. Due to the detector-to-detector difference, it has different strips and mixed noises versus bands. The image contains 210 bands, and each band consists of 307×307 pixels. Its spectral range is from 0.4 to 2.5 micrometers. Table 2 gives the NR and MRD values of band 109 for the Urban dataset, in which all the bold numbers indicate the best results. In Table 2, it can be seen that the proposed approach was able to effectively reduce the noise of the Urban data, and could retain the detailed information well, which means that VLRFmcl can effectively reduce noises with low resolution rates.

Table 2. Quantitative evaluation of noise reduction (NR) and mean relative deviation (MRD) for band 109.

	Band 109	BM3D	ANLM3D	BM4D	LRM	GLF	LRTDTV	LTDL	DnCNN	HSID-CNN	VLRFmcl
NR	1	1.8614	2.1051	2.4648	2.5173	2.7051	2.5936	2.6759	2.4973	2.6993	2.8386
MRD	0	3.2201	3.9113	3.5576	4.2165	3.3927	3.5261	3.6017	3.5162	3.3965	3.1976

Figures 9 and 10 show the denoising results of band 109 and band 151 for the Urban dataset, respectively. As shown in Figures 9a and 10a, these two images were heavily polluted with stripes and mixed noise. As shown in the blue rectangles in Figure 9, obvious stripes can be observed in the results obtained by BM3D, ANLM3D, BM4D, LRM, LRTDTV, and LTDL. Their structure and edge information were also fuzzy. This fact indicates that BM3D, ANLM3D, BM4D, LRM, LRTDTV, and LTDL showed weaker performance in denoising the severely polluted bands for the Urban dataset. The LRM method performed better in the target and detail recovery, but its denoising results still showed obvious stripes and mixed noises. As shown in Figure 9i, the result of DnCNN smoothed out some structures and blurred the visual effect. As shown in Figures 9 and 10,

GLF, HSID-CNN, and VLRMFmcl could effectively restore the edges and textures of the image, while suppressing the mixed noise.

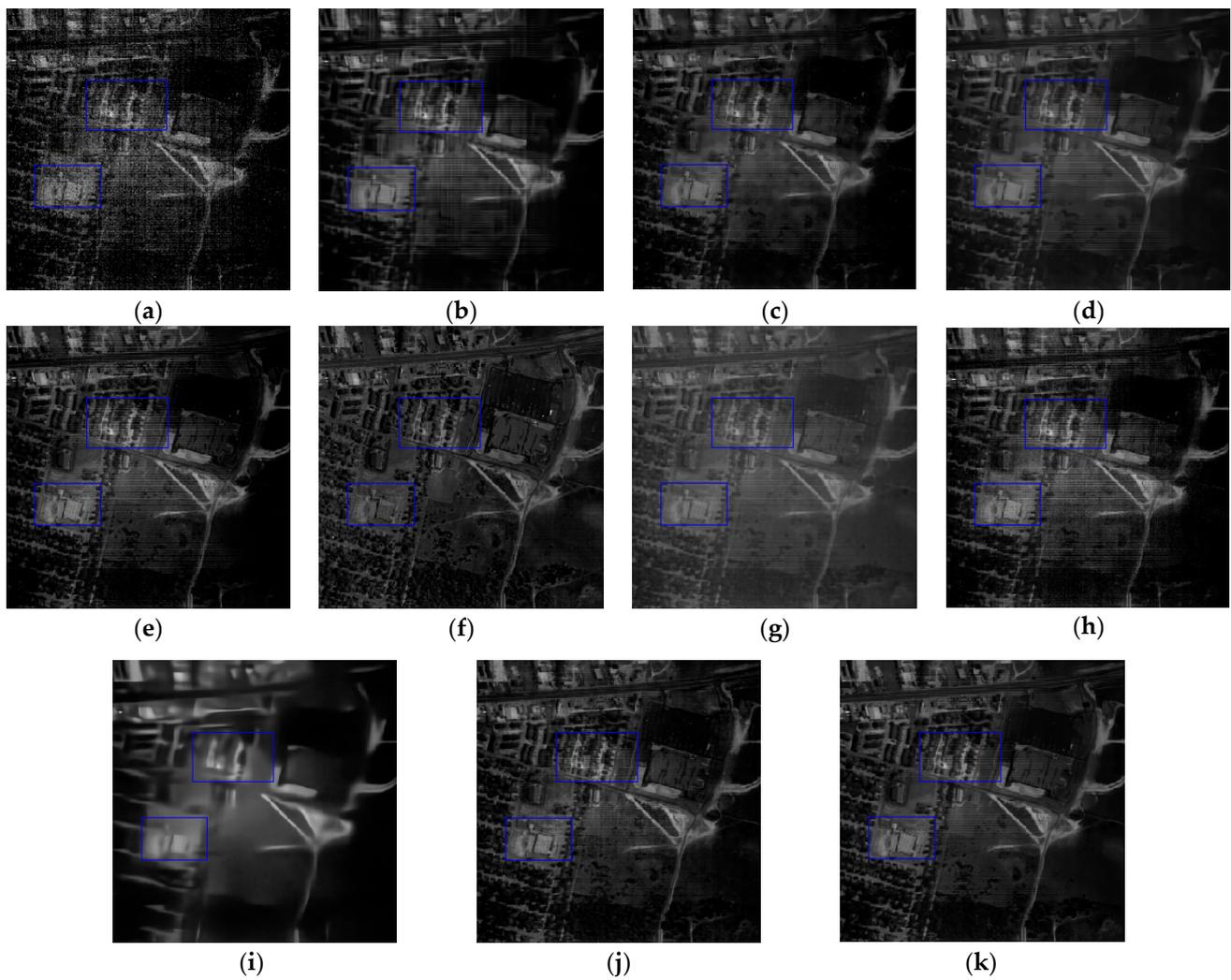


Figure 9. Restored results of the Urban image: (a) original band 109; (b) BM3D; (c) ANLM3D; (d) BM4D; (e) LRM; (f) GLF; (g) LRTDTV; (h) LTDL; (i) DnCNN; (j) HSID-CNN; (k) VLRMFmcl. The blue boxes are used to facilitate the comparison by marking the differences from the results obtained by different algorithms.

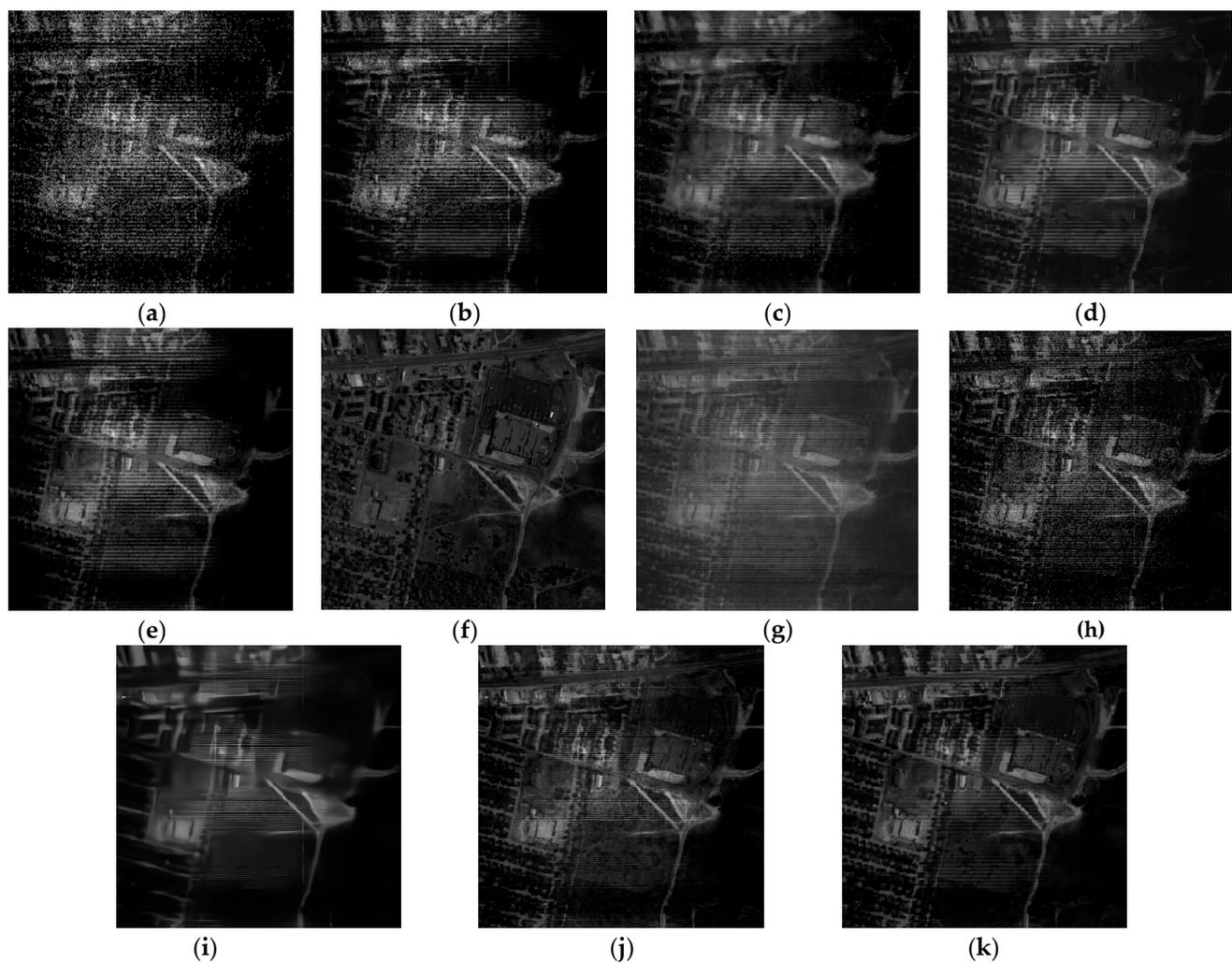


Figure 10. Restored results of the Urban image: (a) original band 151; (b) BM3D; (c) ANLM3D; (d) BM4D; (e) LRMIR; (f) GLF; (g) LRTDTV; (h) LTDL; (i) DnCNN; (j) HSID-CNN; (k) VLRMFmcl.

To facilitate the visual comparison, Figure 11 presents the pseudo-color images of the restored results calculated by different approaches (R: 55, G: 103, B: 207). Comparing the white oval regions, it can be easily seen that the proposed VLRMFmcl method was able to effectively suppress the noises in the smooth area. Meanwhile, it could effectively restore the edge and structure information. Therefore, VLRMFmcl was superior to BM3D, ANLM3D, BM4D, LRMIR, GLF, LRTDTV, LTDL, DnCNN, and HSID-CNN when denoising the Urban data.

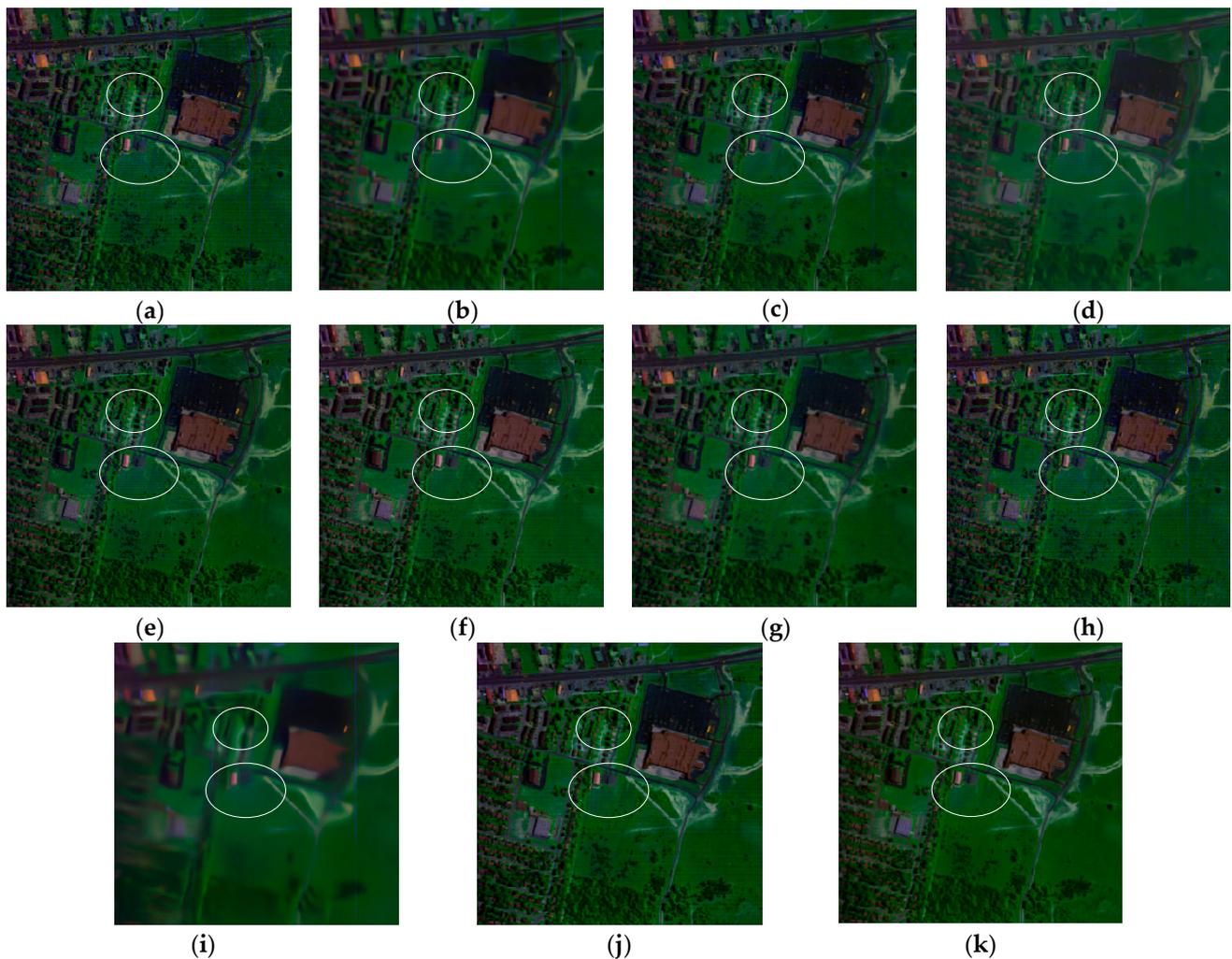


Figure 11. Restored results of the Urban image: (a) original pseudo-color image; (b) BM3D; (c) ANLM3D; (d) BM4D; (e) LRM; (f) GLF; (g) LRTDTV; (h) LTDL; (i) DnCNN; (j) HSID-CNN; (k) VLRMFmcl. The white ovals are used to facilitate the comparison by marking the differences from the results obtained by different algorithms.

4. Conclusions

Introducing multi-patch collaborative learning into low-rank matrix decomposition, a variational model was proposed under the Bayesian framework to achieve the reduction of three kinds of noise in HSIs. The non-local self-similarity of HSIs was explored by developing multi-patch collaborative learning. Through this process, the pixels from edges and heterogeneous regions could be effectively depicted. Then, the variational low-rank matrix decomposition model was constructed to separate the latent noise-free data and mixed noises for collaborative patches. Gaussian distribution with the zero mean and variance adaptively regulated by a gamma distribution was exploited to learn and represent the low-rank property of collaborative patches in the spatial-spectral domain and obtain the related clean data. To sufficiently suppress the mixed noise, their statistical characteristics were effectively depicted by the Dirichlet process Gaussian mixture model, which was constructed using the Gaussian distribution, the inverse Wishart distribution, and the Dirichlet process. Variational Bayesian inference was used to solve the model, having the advantages of simple calculation and high stability. Simulation experiments with different combinations of Gaussian noise, Poisson noise, deadlines, and stripe noise demonstrated the effectiveness of the proposed method. Compared with the BM3D, ANLM3D, BM4D, LRM, GLF, LRTDTV, LTDL, DnCNN, and HSID-CNN methods, the proposed VLRMFmcl method showed superior performance in both the quantitative and qualitative evaluations.

Author Contributions: Conceptualization, S.L.; methodology, S.L.; software, S.L.; validation, S.L. and Z.T.; formal analysis, Z.T. and J.F.; writing—original draft preparation, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 61703328, 61876148), the China Postdoctoral Science Foundation funded project (No. 2018M631165), Shaanxi Province Postdoctoral Science Foundation (No. 2018BSHYDZZ23) and the Fundamental Research Funds for the Central Universities (No. XJJ2018253).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this study are available through: <https://rslab.ut.ac.ir/data>, <https://www.cs.columbia.edu/CAVE/databases/multispectral/>.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rasti, B.; Scheunders, P.; Ghamisi, P. Noise reduction in hyperspectral imagery: Overview and application. *Remote Sens.* **2018**, *10*, 482. [[CrossRef](#)]
2. Xie, T.; Li, S.; Sun, B. Hyperspectral Images Denoising via Nonconvex Regularized Low-Rank and Sparse Matrix Decomposition. *IEEE Trans. Image Process.* **2019**, *29*, 44–56. [[CrossRef](#)] [[PubMed](#)]
3. Kong, X.; Zhao, Y.; Xue, J.; Chan, J.C.W.; Ren, Z.; Huang, H.; Zang, J. Hyperspectral image denoising based on nonlocal low-rank and TV regularization. *Remote Sens.* **2020**, *12*, 1956. [[CrossRef](#)]
4. Zeng, H.; Xie, X.; Ning, J. Hyperspectral image denoising via global spatial-spectral total variation regularized nonconvex local low-rank tensor approximation. *Signal Process.* **2021**, *178*, 107805. [[CrossRef](#)]
5. Lin, B.; Tao, X.; Lu, J. Hyperspectral Image Denoising via Matrix Factorization and Deep Prior Regularization. *IEEE Trans. Image Process.* **2019**, *29*, 565–578. [[CrossRef](#)]
6. Sun, L.; Wu, F.; Zhan, T.; Liu, W.; Wang, J.; Jeon, B. Weighted nonlocal low-rank tensor decomposition method for sparse unmixing of hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1174–1188. [[CrossRef](#)]
7. Zhang, S.; Agathos, A.; Li, J. Robust minimum volume simplex analysis for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6431–6439. [[CrossRef](#)]
8. Sun, X.; Qu, Y.; Gao, L.; Sun, X.; Qi, H.; Zhang, B.; Shen, T. Target Detection Through Tree-Structured Encoding for Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2020**. [[CrossRef](#)]
9. Song, X.; Jiang, X.; Gao, J.; Cai, Z. Gaussian Process Graph-Based Discriminant Analysis for Hyperspectral Images Classification. *Remote Sens.* **2019**, *11*, 2288. [[CrossRef](#)]
10. Elad, M.; Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **2006**, *15*, 3736–3745. [[CrossRef](#)] [[PubMed](#)]
11. Dabov, K.; Foi, A.; Katkovnik, V. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [[CrossRef](#)]
12. Othman, H.; Qian, S.-E. Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 397–408. [[CrossRef](#)]
13. Fu, Y.; Lam, A.; Sato, I. Adaptive spatial-spectral dictionary learning for hyperspectral image restoration. *Int. J. Comput. Vis.* **2017**, *122*, 228–245. [[CrossRef](#)]
14. Manjn, J.V.; Coup, P.; Mart-Bonmat, L. Adaptive non-local means denoising of MR images with spatially varying noise levels. *J. Magn. Reson. Imaging* **2010**, *31*, 192–203. [[CrossRef](#)] [[PubMed](#)]
15. Letexier, D.; Bourennane, S. Noise removal from hyperspectral images by multidimensional filtering. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2061–2069. [[CrossRef](#)]
16. Chen, G.; Qian, S.-E. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 973–980. [[CrossRef](#)]
17. Maggioni, K.E.M.; Katkovnik, V.; Foi, A. Nonlocal transform domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Process.* **2013**, *22*, 119–133. [[CrossRef](#)]
18. Lu, T.; Li, S.; Fang, L.; Ma, Y.; Benediktsson, J.A. Spectral–spatial adaptive sparse representation for hyperspectral image denoising. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 373–385. [[CrossRef](#)]
19. Zhao, L.; Xu, Y.; Wei, Z. Hyperspectral Image Denoising via Coupled Spectral-Spatial Tensor Representation. *IEEE Int. Geosci. Remote Sens. Symp.* **2018**, 4784–4787.
20. Chen, Y.; He, W.; Yokoya, N.; Huang, T.Z. Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition. *IEEE Trans. Cybern.* **2019**, *50*, 3556–3570. [[CrossRef](#)] [[PubMed](#)]
21. Sun, L.; Jeon, B.; Zheng, Y.; Wu, Z. Hyperspectral image restoration using low-rank representation on spectral difference image. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1151–1155. [[CrossRef](#)]

22. Fan, H.; Chen, Y.; Guo, Y.; Zhang, H.; Kuang, G. Hyperspectral image restoration using low-rank tensor recovery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4589–4604. [[CrossRef](#)]
23. Chen, Y.; Guo, Y.; Wang, Y.; Wang, D.; Peng, C.; He, G. Denoising of hyperspectral images using nonconvex low rank matrix approximation. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5366–5380. [[CrossRef](#)]
24. Okatani, T.; Yoshida, T.; Deguchi, K. Efficient algorithm for low-rank matrix factorization with missing components and performance comparison of latest algorithms. *IEEE ICCV* **2011**, 842–849.
25. Zhang, H.; He, W.; Zhang, L. Hyperspectral image restoration using low-rank matrix recovery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4729–4743. [[CrossRef](#)]
26. Zhuang, L.; Bioucas-Dias, J.M. Hyperspectral image denoising based on global and non-local low-rank factorizations. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 1900–1904.
27. Xue, J.; Zhao, Y.; Liao, W.; Kong, S.G. Joint spatial and spectral low-rank regularization for hyperspectral image denoising. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 1940–1958. [[CrossRef](#)]
28. Fan, H.; Li, C.; Guo, Y.; Kuang, G.; Ma, J. Spatial–spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6196–6213. [[CrossRef](#)]
29. Wang, Q.; Wu, Z.; Jin, J. Low rank constraint and spatial spectral total variation for hyperspectral image mixed denoising. *Signal Process.* **2018**, *142*, 11–26. [[CrossRef](#)]
30. Wang, Y.; Peng, J.; Zhao, Q.; Leung, Y.; Zhao, X.-L.; Meng, D. Hyperspectral image restoration via total variation regularized low-rank tensor decomposition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1227–1243. [[CrossRef](#)]
31. Cao, W.; Wang, K.; Han, G. A robust PCA approach with noise structure learning and spatial–spectral low-rank modeling for hyperspectral image restoration. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3863–3879. [[CrossRef](#)]
32. Gong, X.; Chen, W.; Chen, J. A low-rank tensor dictionary learning method for hyperspectral image denoising. *IEEE Trans. Signal Process.* **2020**, *68*, 1168–1180. [[CrossRef](#)]
33. Pappayan, V.; Elad, M. Multi-scale patch-based image restoration. *IEEE Trans. Image Process.* **2016**, *25*, 249–261. [[CrossRef](#)]
34. Zhang, Q.; Zhang, L.; Yang, Y. Local patch discriminative metric learning for hyperspectral image feature extraction. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 612–616. [[CrossRef](#)]
35. Haut, J.M.; Paoletti, M.E.; Plaza, J. Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6440–6461. [[CrossRef](#)]
36. Zhang, M.; Li, W.; Du, Q. Feature Extraction for Classification of Hyperspectral and LiDAR Data Using Patch-to-Patch CNN. *IEEE Trans. Cybern.* **2018**, *50*, 2168–2267. [[CrossRef](#)]
37. Xu, Y.; Wu, Z.; Chanussot, J. Nonlocal patch tensor sparse representation for hyperspectral image super-resolution. *IEEE Trans. Image Process.* **2019**, *28*, 3034–3047. [[CrossRef](#)]
38. Zhang, Q.; Yuan, Q.; Li, J.; Sun, F.; Zhang, L. Deep spatio-spectral Bayesian posterior for hyperspectral image non-iid noise removal. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 125–137. [[CrossRef](#)]
39. Wei, K.; Fu, Y.; Huang, H. 3-D Quasi-Recurrent Neural Network for Hyperspectral Image Denoising. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 363–375. [[CrossRef](#)] [[PubMed](#)]
40. Zhang, Q.; Yuan, Q.; Li, J. Hybrid Noise Removal in Hyperspectral Imagery with a Spatial-Spectral Gradient Network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *59*, 7317–7329. [[CrossRef](#)]
41. Ma, H.; Liu, G.; Yuan, Y. Enhanced Non-Local Cascading Network with Attention Mechanism for Hyperspectral Image Denoising. In Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Virtual Conference, Barcelona, Spain, 4–9 May 2020; pp. 2448–2452.
42. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [[CrossRef](#)] [[PubMed](#)]
43. Yuan, Q.; Zhang, Q.; Li, J.; Shen, H.; Zhang, L. Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1205–1218. [[CrossRef](#)]
44. Zhao, Q.; Meng, D.; Xu, Z. L1-Norm Low-Rank Matrix Factorization by Variational Bayesian Method. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 825–839. [[CrossRef](#)] [[PubMed](#)]
45. Wei, K.; Fu, Y. Low-rank Bayesian tensor factorization for hyperspectral image denoising. *Neurocomputing* **2019**, *331*, 412–423. [[CrossRef](#)]
46. Li, W.; Du, Q.; Zhang, F.; Hu, W. Hyperspectral image classification by fusing collaborative and sparse representations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4178–4187. [[CrossRef](#)]
47. Zhang, L.; Mou, X.; Zhang, D. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [[CrossRef](#)] [[PubMed](#)]
48. Yuan, Q.; Zhang, L.; Shen, H. Hyperspectral image denoising with a spatial–spectral view fusion strategy. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2314–2325. [[CrossRef](#)]
49. Shen, H.; Zhang, L. A MAP-based algorithm for destriping and inpainting of remotely sensed images. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1492–1502. [[CrossRef](#)]
50. Qian, Y.; Ye, M. Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 499–515. [[CrossRef](#)]