Deep spatio-spectral Bayesian posterior for hyperspectral image non-i.i.d. noise removal

Qiang Zhang[a], Qiangqiang Yuan[b,c,*], Jie Li[b,*], Fujun Sun[d], Liangpei Zhang[a,c]

[a] *State Key Laboratory of Information Engineering, Survey Mapping and Remote Sensing, Wuhan University, Wuhan, China*
[b] *School of Geodesy and Geomatics, Wuhan University, Wuhan, China*
[c] *Collaborative Innovation Center of Geospatial Technology, Wuhan, China*
[d] *Beijing Electro-mechanical Engineering Institute, Beijing, China*

ARTICLE INFO

ABSTRACT

The noise pollution issue seriously obstructs subsequent interpretation and application of the hyperspectral image (HSI). In this work, differing from most existing HSI denoising methods ideally assumed that noise in different bands denotes independent & identically distributed (i.i.d.), we propose a novel HSI denoising approach focusing on non-i.i.d. noise removal. The presented framework collaboratively models the non-i.i.d. noise embedding within HSI and removals them under a deep spatio-spectral Bayesian posterior (DSSBP) structure. Specifically, the non-i.i.d. noise estimation, distribution and removal procedure are both executed with the model-driven based strategy and data-driven based strategy. Through blending the Bayesian variational posterior and deep convolutional neural network, the proposed method both inherits the reliability of traditional model-driven based methods for HSI noise modeling and the high efficiency of data-driven based methods for parameters learning. Simulated and real experiments in different HSIs and non-i.i.d. noise scenarios testify that the proposed DSSBP approach outperforms other existing methods for non-i.i.d. noise removal, in terms of evaluation indexes and executive efficiency.

1. Introduction

Hyperspectral image (HSI) denotes the cube data comprised of continuous, dense, and broad spectral response for the observed objects. Due to the image-spectrum reflecting property, HSI has the unique superiority in terms of fine interpretation (Paoletti et al., 2018; Brell et al., 2019). Compared with natural image, HSI owns more spectrums from visible to infrared bands which contains more attribute reflected information. Therefore, HSI has been widely applied to surface classification (Sidike et al., 2018), anomaly detection (Wu et al., 2018), agricultural monitoring (Asaari et al., 2018) etc. Nevertheless, in the process of sensor imaging and atmospheric transmission, HSI inevitably suffers from the noise pollution issues. This phenomenon severely degrades the quality of HSI, which is adverse for subsequent processing and applications of HSI (Sun et al., 2017a; Hong et al., 2019). In consequence, how to effectively removal the disturbed noise in HSI is a vitally important and initial task before other interpretation requirements (Wang et al., 2019).

Up to now, plenty of HSI noise reduction methods have been presented (Li et al., 2015; Zhuang and Bioucas-Dias, 2018). It should be noted that differing from nature image, how to collaboratively combine with the spatial and spectral feature is a critical point for 3D-cube image denoising (Guo et al., 2013; Yue et al., 2018). According to the strategy and structure, pre-existing HSI denoising approaches can be classified into two types in this paper: model-driven and data-driven based approaches. Detailed descriptions and analysis are given below.

(1) **Model-driven based methods for HSI denoising**: This type of methods usually relies on the typical prior of HSI and then design the optimization model to generating the denoising result (Sun et al., 2018). Specifically, spatio-spectral total variation (Yuan et al., 2012; He et al., 2016), spatio-spectral non-local mean (Maggioni et al., 2012; Qian and Ye, 2013), spatio-spectral sparse representation (Zhao and Yang, 2015; Lu et al., 2016), low-rank prior (Zhang et al., 2014; Chen et al., 2017; Sun et al., 2017b; Xue et al., 2018), tensor decomposition (Karami et al., 2011; Guo et al., 2013; Chen et al., 2019a) have been developed for HSI denoising. For instance, Yuan et al. simultaneously considered the spatial and spectral dimension through total variational model (Yuan et al., 2012). Maggioni et al. stacked 4-D cubes of voxels and exploited the

---

local and non-local relevance between each or different cubes for HSI denoising (Maggioni et al., 2012). Lu et al. established spatio-spectral sparse dictionary to reduce disturbed noise in HSIs (Lu et al., 2016). Recently, considering the HSI cube as a 3D-tensor data, tensor decomposition methods have been testified that the low-rank prior is beneficial for significant improvement of HSI re-construction (Xie et al., 2018). As an example, Fan et al. developed low-rank tensor constructing method to preserve spatial semantics of HSI and concurrently reduce mixed noise (Fan et al., 2017; Fan et al., 2018).

(2) **Data-driven based methods for HSI denoising**: Different from model-driven based methods, this type of methods depends on the large numbers of the sample data (clean label data and corre-sponding noisy data), through end-to-end deep learning framework to obtain the fitting denoiser. Recently, deep learning strategy has been broadly utilized for remote sensing data processing (Zhu et al., 2017), and makes progress in several image quality improvement tasks such as missing data reconstruction (Zhang et al., 2018a), pansharpening (Xing et al., 2018), super-resolution (Lanaras et al., 2018) and despeckling (Zhang et al., 2018b) etc. Aiming at natural image denoising, several works have been developed such as DnCNN (Zhang et al., 2017), FFDNet (Zhang et al., 2018c), CBDNet (Guo et al., 2019) etc. In terms of HSI denoising task relying on deep learning, Xie and Li proposed a cascaded neural network in-tegrating mutative non-linear function for reducing noise in HSI (Xie and Li, 2017). Yuan et al. developed a spatial-spectral con-volutional neural network with multi-scale and multi-level feature for HSI denoising (Yuan et al., 2019). Chang et al. employed multi-channel convolutional filters to take in hybrid noise for model training (Chang et al., 2019). By resolving 3-D convolution to 2-D convolution and spectral vector, Dong et al. presented a modified HSI denoising 3D U-net (Dong et al., 2019). Besides, for better re-moving mixed noise in HSIs, Zhang et al. further bring spatio-spectral gradient information (Zhang et al., 2019) into the deep CNN. Besides, Liu and Lee (2019) utilized 3D dilated convolution neural network for HSI denoising, extracting deep features from spatial and spectral prospects.

Generally, data-driven based methods for HSI denoising are faster and more convenient through an end-to-end learning framework, without careful adjustment of hyper-parameters and complex optimi-zation in model-driven based methods. Conversely, model-driven based methods for HSI denoising can better conform to the dominant prior of HSI such as low rank property, without establishing large training HSI samples in data-driven based methods (Yuan et al., 2019). Besides, most existing HSI denoising methods ideally assumed that noise in HSI de-notes i.i.d. (Xiong et al., 2019) such as Gaussian noise. However, the actual noise distribution is more complicated in different bands, and usually emerges with non-i.i.d. form (Yue et al., 2019a) in most HSIs. This issue also limits the practicability and robustness of many HSI denoising methods.

Therefore, how to unite the merits of both model-driven and data-driven methods and reduce their weakness is significant for HSI de-noising especially directing at non-i.i.d. noise. The main purposes can be summarized as two points below: (1) Considering HSI spatio-spectral characteristic; (2) Aiming at mixed noise. From this prospect, we pro-pose a novel deep spatio-spectral Bayesian posterior (DSSBP) frame-work in this work for HSI non-i.i.d. noise removal. The presented DSSBP collaboratively models the HSI non-i.i.d. noise and removals them under the spatio-spectral Bayesian posterior structure. Specifically, the non-i.i.d. noise modeling and removing procedure are both executed with the deep spatio-spectral learning model. Besides, taking the spatial directionality and spectral heterogeneity of sparse noise in HSI into consideration, the anisotropic total variational term is utilized for modeling the sparse noise especially stripe noise. The highlighted in-novations of this work are summarized below:

(1) The proposed DSSBP can simultaneously estimate non-i.i.d. noise and sparse-distributed noise distribution of each band in HSI and removal noise in a Bayesian variational framework. This tactic is more practical for most realistic scenarios, contrasted with the idealized i.i.d. noise assumption in most HSI denoising methods.

(2) To simultaneously exploit spectral relevance and consider spatial difference, 3D and 2D convolution neural network are both em-ployed into the proposed HSI denoising model. Notably, non-i.i.d. noise included sparse noise estimating and removing operations are collaboratively parameterized into the end-to-end data-driven fra-mework.

(3) Through blending the Bayesian variational posterior and deep neural network, the proposed DSSBP inherits the superiority of both traditional model-driven and data-driven based methods. Experiments on different HSI non-i.i.d. noise scenarios testify that the presented approach outperforms other contrastive algorithms for noise removal.

The remaining parts are arranged below. In Sections 2 and 3, the problem formulation and the proposed method is depicted for HSI non-i.i.d. noise removal, respectively. Then the simulated and real experi-mental results are shown in Section 4. In the end, the conclusion and prospect are generalized in Section 5.

2. Problem formulation

In real scenarios, HSI noises generally have more complicated sta-tistical structures (Zheng et al., 2019). The noise level and type may be different and diversity in different bands (such as mixed noise). And in the given band, the spatial distribution of noise is more complex and anisotropic (Chen et al., 2018), which not just obeys the Gaussian distribution. Therefore, in this work we considered more complex HSI noise (non-i.i.d. noise) beyond only Gaussian in HSI denoising. Let the HSI denotes 3-D cube $\mathbf{Y} \in \mathbb{R}^{w \times h \times b}$, in which w and h stand for the spatial dimension, b represents spectral dimension of the HSI. Then the de-graded model from clean HSI $\mathbf{X}$ to noisy HSI $\mathbf{Y}$ can be formulated as (Rasti et al., 2018):

$$\mathbf{Y} = \mathbf{X} + \mathbf{N} + \mathbf{S} \tag{1}$$

among this relationship $\mathbf{N}$ stands for the non-i.i.d. noise. $\mathbf{S}$ is the sparse noise especially stripe noise in observed HSI. Apparently, to acquire the noise-free result $\mathbf{X}$ from Eq. (1), the inverse problem is ill-posed and additional constraints and priors need to be imposed. Then we construct the noise model to depict the degraded procedure of the noisy HSI $\mathbf{Y}$ as following:

$$\mathbf{Y}_i \tilde{N}(\mathbf{Y}_i | Z_i, \sigma_i^2), i = 1, 2, \cdots b \tag{2}$$

where $N(\cdot | \mu, \sigma^2)$ marks as the Gaussian distribution through mean μ and variance σ^2. $\mathbf{Z}$ denotes the latent clean HSI underlying $\mathbf{Y}$. Notably, the noise distribution in different bands of HSI is independent and discrepant. Different from most HSI denoising methods assuming i.i.d. noise distribution in Eq. (2), we consider the noise in HSI as a non-i.i.d. distribution to accord with the realistic scenario. The noise variance $\sigma^2 = \{\sigma_1^2, \sigma_2^2, \cdots, \sigma_b^2\}$ is flexible embedding in spatial dimension, and a rational conjugate prior as imposed in (Yue et al., 2019b) is introduced as below:

$$\sigma_i^2 \tilde{I}G\left(\sigma_i^2 \Big| \frac{r^2}{2} - 1, \frac{r^2 \xi_i}{2}\right), i = 1, 2, \cdots b \tag{3}$$

where $IG(\cdot | \alpha, \beta)$ denotes the inverse Gamma distribution by para-meters α and β, ξ stands for the filtering result of the residual map through a Gaussian filter within $r \times r$ window. In other words, the noise variance σ^2 is pixel-wise Gaussian distribution for each band in HSI.

In addition, the original clean HSI $\mathbf{X}$ obviously exists a dominant prior with the latent clean HSI $\mathbf{Z}$. From this perspective, the conjugate Gaussian prior for $\mathbf{Z}$ is imposed as below:

$$\mathbf{Z}_i \tilde{N}(\mathbf{Z}_i | \mathbf{X}_i, \varepsilon_0^2), \; i = 1, 2, \cdots b \tag{4}$$

where ε_0^2 represents a predetermined parameter with a small value. Besides, the sparse noise $\mathbf{S}$ such as stripe noise in HSI simultaneously presents with spatial horizontal or vertical distribution. And the density of sparse noise also differs in spectral dimension. Naturally, the conjugate sparse prior K for $\mathbf{S}$ is constrained as follows:

$$\mathbf{S}_i K \tilde{} (\mathbf{S}_i | Y_i, \mathbf{Y}), \; i = 1, 2, \cdots b \tag{5}$$

Combining Eqs. (2)–(5), we can acquire a full Bayesian inference framework for the HSI non-i.i.d. noise modeling issue. Then our target focuses on estimating the posterior on the latent clean HSI $\mathbf{Z}$, noise variance σ^2 and sparse-distributed noise $\mathbf{S}$ through the noisy HSI $\mathbf{Y}$, denoted by $p(\mathbf{Z}, \sigma^2, \mathbf{S} | \mathbf{Y})$.

3. Methodology

3.1. Spatio-spectral Bayesian posterior network

To estimate the posterior $p(\mathbf{Z}, \sigma^2, \mathbf{S} | \mathbf{Y})$ through Eqs. (2)–(5), a probability distribution $q(\mathbf{Z}_i, \sigma_i^2, \mathbf{S}_i | \mathbf{Y})$ for each spectrum in HSI is established in consideration of the conditional independence between the latent clean HSI $\mathbf{Z}$, noise variance σ^2 and sparse-distributed noise $\mathbf{S}$:

$$q(\mathbf{Z}_i, \sigma_i^2, \mathbf{S}_i | \mathbf{Y}) = q(\mathbf{Z}_i | \mathbf{Y}) q(\sigma_i^2 | \mathbf{Y}) q(\mathbf{S}_i | \mathbf{Y}) \tag{6}$$

According to the conjugate priors in Eqs. (3)–(5), we can logically speculate the three estimated posteriors as below:

$$q(\mathbf{Z}_i | \mathbf{Y}) = N(\mathbf{Z}_i | \varphi_i(Y_i, \mathbf{Y}_s; W_E), m_i^2(Y_i, \mathbf{Y}_s; W_E)) \tag{7}$$

$$q(\sigma_i^2 | \mathbf{Y}) = IG(\sigma_i^2 | \alpha_i(Y_i, \mathbf{Y}_s; W_D), \beta_i(Y_i, \mathbf{Y}_s; W_D)) \tag{8}$$

$$q(\mathbf{S}_i | \mathbf{Y}) = TV(\mathbf{S}_i | K_s(\nabla \mathbf{Y}_s; W_s)) \tag{9}$$

where φ_i and m_i^2 represent the intermediate results to calculated the posterior latent clean HSI $\mathbf{Z}$ from noisy HSI $\mathbf{Y}$. α_i and β_i represent the intermediate results to calculated the posterior noise variance σ^2 from noisy HSI $\mathbf{Y}$. K_s represent the spatio-spectral gradient (Xu et al., 2019) through total variational (TV) prior to constrain the sparse-distributed noise $\mathbf{S}$. Y_i and $\mathbf{Y}_s$ stand for the i-th band and its neighboring spectral cube in $\mathbf{Y}$, respectively. W_E and W_S denote the trainable parameters within the non-i.i.d. noise estimation network and sparse noise spatio-spectral gradient network, respectively. W_D denotes the trainable parameters within the non-i.i.d. noise distribution network. The flowchart and procedure of the presented framework is depicted in Fig. 1. "$\mathbf{s}$" stands for the stride size in convolutional operation. "$\mathbf{c}$" represents

the channel number of feature maps. "$3 \times 3 \times 3 \times 32$" denotes the 3D-filters of size $3 \times 3 \times 3$ with 32 channels. "$3 \times 3 \times 64$" denotes the 2D-filters of size 3×3 with 64 channels.

Specifically, the network parameters of W_E, W_D and W_S are integrated through the aforementioned posteriors by training samples. In other words, non-i.i.d. noise and sparse noise estimating and removing operations are collaboratively parameterized into the end-to-end data-driven learning framework.

3.2. Network structure and learning

Aiming at non-i.i.d. noise estimation network and noise distribution network in Fig. 1, we separate the spatial band and its adjacent spectral cube, and utilize the multi-channel 2D convolutional neural network to generate the spatial and spectral feature maps. Then both two feature maps are concatenated with a united result for subsequent operation. In terms of non-i.i.d. noise estimation network contained parameters W_E, the layers depth is set as nine 2D convolutional and activate layer (2D Conv + ReLU). And for non-i.i.d. noise distribution network contained parameters W_D, the layers depth is set as five 2D convolutional and activate layer, as shown in Fig. 1. The middle feature maps quantity of all these layers are fixed with 64 the kernel size is all set as 3×3 with stride equal to 1.

For the sparse noise network, we employ the 3D convolutional neural network to jointly extract spatio-spectral information. Notably, the spectral gradient cube is served as the input data into the sparse noise network with parameters W_S. The layers depth of the sparse noise network is set as five 3D convolutional and activate layer (3D Conv + ReLU). The middle feature map number of all these layers is fixed with 32 and the kernel size is all set as $3 \times 3 \times 3$ with stride equal to 1.

After constructing the whole Bayesian variational framework, the lower bound of the non-i.i.d. noise (Yue et al., 2019b) in HSI can be derived:

$$L(\mathbf{Z}, \sigma^2; \mathbf{Y}) = E_{q(\mathbf{Z}, \sigma^2 | \mathbf{Y})} - D_{KL}(q(\mathbf{Z} | \mathbf{Y}) \| p(\mathbf{Z})) - D_{KL}(q(\sigma^2 | \mathbf{Y}) \| p(\sigma^2)) \tag{10}$$

where the three sub-items in Eq. (10) can be as the following form:

$$E_{q(\mathbf{Z}, \sigma^2 | \mathbf{Y})} = \int q(\mathbf{Z}, \sigma^2 | \mathbf{Y}) \log p(\mathbf{Y} | \mathbf{Z}, \sigma^2) d\mathbf{Z} d\sigma^2 \tag{11}$$

$$\log p(\mathbf{Y} | \mathbf{Z}, \sigma^2) = \sum_i^b \left\{ -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma_i^2 - \frac{(Y_i - Z_i)^2}{2\sigma_i^2} \right\} \tag{12}$$
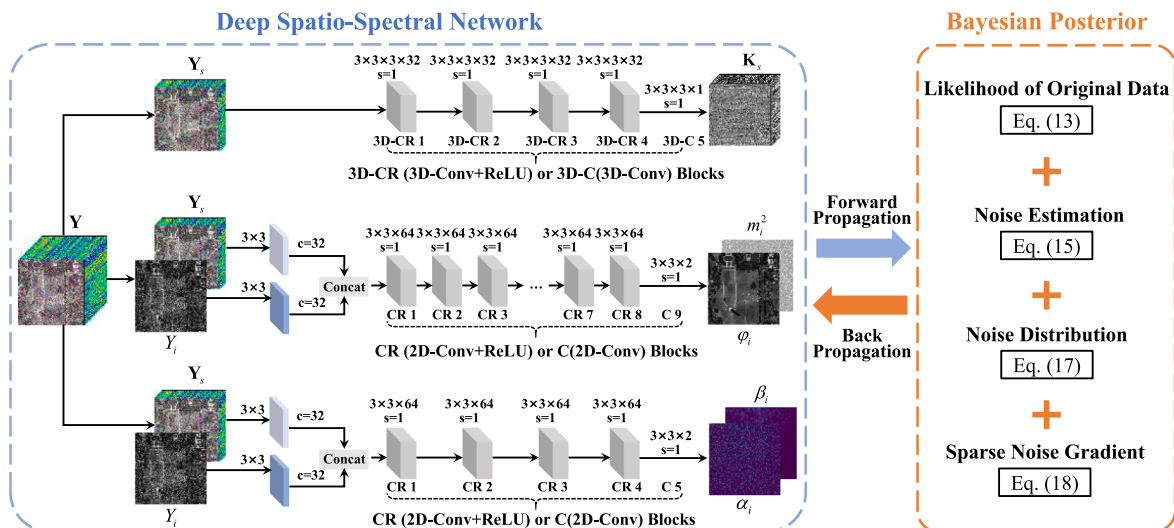

Fig. 1. Flowchart and procedure of the presented approach.

simultaneously combining with Eq. (6), Eq. (11) and Eq. (12), the likelihood $E_{q(\mathbf{Z},\sigma^2|\mathbf{Y})}$ of original clean data is calculated as:

$$E_{q(\mathbf{Z},\sigma^2|\mathbf{Y})} = -\frac{1}{2}mean\left\{\log 2\pi + \log \beta_i - \psi(\alpha_i) + \frac{\alpha_i}{\beta_i}[(Y_i - \varphi_i)^2 + m_i^2]\right\}$$

(13)

In terms of the *KL* divergence $D_{KL}(q(\mathbf{Z}|\mathbf{Y})\|p(\mathbf{Z}))$ for Gaussian distribution in Eq. (10),

$$D_{KL}(q(\mathbf{Z}|\mathbf{Y})\|p(\mathbf{Z})) = \sum_i^b D_{KL}(N(\mathbf{Z}_i|\varphi_i, m_i^2)\|p(\mathbf{Z}_i|X_i, \varepsilon_0^2))$$

(14)

simultaneously combining with Eq. (2), Eq. (4), Eq. (7) and Eq. (14), the likelihood $D_{KL}(q(\mathbf{Z}|\mathbf{Y})\|p(\mathbf{Z}))$ can be calculated as:

$$D_{KL}(q(\mathbf{Z}|\mathbf{Y})\|p(\mathbf{Z})) = \frac{1}{2}mean\left\{\frac{(\varphi_i - X_i)^2}{2\varepsilon_0^2} + \left[\frac{m_i^2}{\varepsilon_0^2} - \log\frac{m_i^2}{\varepsilon_0^2} - 1\right]\right\}$$

(15)

And for the *KL* divergence $D_{KL}(q(\sigma^2|\mathbf{Y})\|p(\sigma^2))$ for inverse Gamma distribution in Eq. (10), it can be depicted as:

$$D_{KL}(q(\sigma^2|\mathbf{Y})\|p(\sigma^2)) = \sum_i^b D_{KL}(IG(\sigma_i^2|\alpha_i, \beta_i)\|IG(\sigma_i^2|\frac{r^2}{2} - 1, \frac{r^2\xi_i}{2}))$$

(16)

simultaneously uniting Eq. (3), Eq. (8), and Eq. (16), the final form of $D_{KL}(q(\sigma^2|\mathbf{Y})\|p(\sigma^2))$ can be derived as:

$$D_{KL}(q(\sigma^2|\mathbf{Y})\|p(\sigma^2)) = mean\left\{(\alpha_i - \lambda)\psi(\alpha_i) + \log\frac{\Gamma(\lambda)}{\Gamma(\alpha_i)} + \lambda\log\frac{2\beta_i}{p^2\xi_i}\right\}$$

(17)

where λ is equal to $(\frac{r^2}{2} - 1)$ in Eq. (17), $\psi(\cdot)$ represents the digamma function.

Besides, taking the spatial directionality and spectral heterogeneity of sparse noise in HSI into consideration in Eq. (9), the anisotropic TV term is utilized for modeling the sparse noise especially stripe noise:

$$TV(\mathbf{S}_i; \mathbf{Y}) = \|\nabla_v\mathbf{K}_s\|_1 + \|\nabla_h\mathbf{K}_s\|_1$$

(18)

where the vertical and horizontal gradient of spectral difference are imposed with L1 norm in the sparse noise network.

In terms of the holistic network optimization, the lower bound of the non-i.i.d. noise in Eq. (10) and sparse noise TV term in Eq. (14) are blended into the integrative loss function as below:

$$loss(W_E, W_D, W_S) = -L(\mathbf{Z}, \sigma^2; \mathbf{Y}) + \eta\cdot TV(\mathbf{S}_i; \mathbf{Y})$$

(19)

where η devotes the penalty parameter of the sparse noise TV term. Relying on Eq. (19), the proposed framework conforms to the back propagation (BP) gradient descent algorithm (LeCun et al., 1990) to optimize three network parameters W_E, W_D, and W_S.

4. Experimental results

4.1. Model training and parameters setting

For the training of the proposed HSI noise estimating and removing model, the HSI data in Xiongan (Matiwan Village), China, acquired by airborne hyperspectral imager is utilized as the training dataset. The dataset has the characteristics of high spectral-resolution, high spatial-resolution, and various ground objects after geometric and radiation correction. After eliminating the noisy bands, 242 clean bands with the size of 1580 × 3750 is set as the training samples. These training labels are subsequently cropped as the cube data with the size of 64 × 64 × 24, where the spatial stride and spectral stride are set as 32 and 3, respectively. To generate the non-i.i.d. noisy HSI data, the simulated noise and stripe (Roshan and Amr, 2011) for each band are imposed on the clean HSI cubes. The non-i.i.d. noise variance $\sigma^2 = \{\sigma_1^2, \sigma_2^2, \cdots, \sigma_b^2\}$ as the range of [0–0.5]. For each band in HSIs, the specific non-i.i.d. noise variance σ_b^2 are selected from the range of [0–0.5] through uniform random distribution. Besides, data resampling and rotation are also employed to augment the quantity of training samples and enhance model generalization ability for different testing HSIs.

200 HSI patches are separately employed from the training patches as the validation data sets. The chosen mechanism is random selection between the original noise-free HSI patches. The whole training procedure is based on the Pytorch framework with Python language. The batch size and epoch number are fixed as 64 and 300, respectively. For optimizing the model parameters in the BP procedure, Adam (Kingma and Ba, 2014) gradient descent algorithm is carried out with the given parameters 0.9 and 0.999. For per 25 epochs, learning rates are progressively descended with multiplicative factor 0.5 where the initial value is equal to 0.01 (Zhang et al., 2020). The Gaussian filter window size r and ε_0^2 are set as 5 and 1e-4, respectively. The penalty parameter η in Eq. (15) is devoted as the reciprocal of cube spectrum number. Available codes will be released at: https://github.com/WHUQZhang/DSSBP.

4.2. Testing data and performance evaluation

To validate the availability of the presented approach, both the simulative and actual scenarios are implemented for HSI denoising. For simulated experiments, three different cases on W. DC and Pavia HSIs are executed to validate the capacity on i.i.d. noise, non-i.i.d. noise, and mixed noise removal, respectively. And for real experiments, five different HSIs are employed to testify the practicability under the realistic and complicated scenarios. In addition, five state-of-the-arts HSI denoising algorithms HSSNR (Othman and Qian, 2006), BM4D (Maggioni et al., 2012), LRMR (Zhang et al., 2014), NMoG (Chen et al., 2018), SSGN (Zhang et al., 2019) are served as the contrastive methods. Considering fair comparison, SSGN is also retraining with the same training samples with the presented approach.

As to the performance evaluation, the mean PSNR for each band, mean SSIM (Chen et al., 2019b) for each band, and mean spectral angle (MSA) for each spectral vector are measured as the objective evaluation indexes for simulated experiments. MPSNR and MSSIM depict the structural recovering ability in spatial domain, and MSA measures the spectrum maintaining ability in spectral domain. In addition, the consuming times of all the HSI denoising methods are recorded under the same operating environment for the comparisons of executed efficiency.

4.3. Simulated experiments

In this section, three different cases in W. DC Mall and Pavia HSIs are executed to validate the capacity on i.i.d. noise, non-i.i.d. noise, and mixed noise removal, respectively. Detailed descriptions are listed as following.

Case 1 (i.i.d. Gaussian noise): For the original HSI data, all the bands are degraded through i.i.d. Gaussian noise with equal-distributed on spatial dimension and identical-distributed on spectral dimension. The i.i.d. Gaussian noise variance σ^2 is equal to 0.1 for all the bands.

Case 2 (non-i.i.d. Gaussian noise): For the original HSI data, all the bands are imposed with non-i.i.d. Gaussian noise, which denotes unequal-distributed on spatial dimension and unidentical-distributed on spectral dimension. We set r equal to 11, and the non-i.i.d. noise variance $\sigma^2 = \{\sigma_1^2, \sigma_2^2, \cdots, \sigma_b^2\}$ as the range of [0–0.5] in Eq. (6). The maximum level of non-i.i.d. noise is equal to 0.5. For different bands in HSIs, the specific non-i.i.d. noise variance σ_b^2 are selected from the range of [0–0.5] through uniform random distribution.

Case 3 (non-i.i.d. Gaussian noise + stripe noise): Uniformly to Case 2, all the bands are contaminated with the non-i.i.d. Gaussian noise with unequal-distributed on spatial dimension and unidentical-distributed on spectral dimension. Besides, partial spectrums in HSI are imposed

Table 1
Quantitative evaluation of the simulated W.DC HSI experiments.

	Noisy HSI	HSSNR	BM4D	LRMR	NMoG	SSGN	Proposed
Case 1: i.i.d. Gaussian noise							
MPSNR	23.28	27.26	28.64	33.22	<u>34.51</u>	34.34	**34.53**
MSSIM	0.769	0.923	0.941	0.981	**0.983**	0.982	<u>0.982</u>
MSA	19.47	9.083	5.116	4.628	**4.127**	4.248	<u>4.129</u>
Time/s	–	304.4	461.8	449.6	513.8	**7.3**	<u>14.8</u>
Case 2: non-i.i.d. Gaussian noise							
MPSNR	19.78	23.51	24.24	28.73	<u>29.76</u>	25.84	**30.88**
MSSIM	0.654	0.84	0.856	0.958	<u>0.962</u>	0.893	**0.976**
MSA	23.72	11.34	10.37	6.14	<u>5.73</u>	9.28	**5.102**
Time/s	–	312.6	479.5	437.9	538.2	**7.2**	<u>15.0</u>
Case 3: non-i.i.d. Gaussian noise + stripe noise							
MPSNR	20.82	25.64	26.39	28.35	<u>29.98</u>	27.28	**30.67**
MSSIM	0.669	0.893	0.938	0.957	<u>0.967</u>	0.947	**0.974**
MSA	22.47	10.92	7.87	<u>6.142</u>	6.298	6.564	**5.436**
Time/s	–	314.8	486.3	440.7	542.5	**7.3**	<u>15.1</u>

Table 2
Quantitative evaluation of the simulated Pavia University HSI experiments.

	Noisy HSI	HSSNR	BM4D	LRMR	NMoG	SSGN	Proposed
Case 1: i.i.d. Gaussian noise							
MPSNR	24.36	29.72	31.48	34.86	36.97	**37.71**	<u>37.69</u>
MSSIM	0.783	0.951	0.963	0.971	**0.989**	0.983	<u>0.988</u>
MSA	17.34	6.942	5.624	4.583	<u>3.695</u>	3.824	**3.642**
Time/s	–	284.6	412.5	385.7	468.4	**5.5**	<u>12.3</u>
Case 2: non-i.i.d. Gaussian noise							
MPSNR	18.67	24.15	26.14	28.25	<u>29.11</u>	27.70	**30.26**
MSSIM	0.615	0.806	0.851	0.951	<u>0.962</u>	0.945	**0.971**
MSA	20.38	11.81	6.420	6.673	<u>6.074</u>	7.436	**5.589**
Time/s	–	291.3	423.8	394.2	473.5	**5.6**	<u>12.4</u>
Case 3: non-i.i.d. Gaussian noise + stripe noise							
MPSNR	17.84	22.93	25.72	27.68	<u>28.43</u>	27.14	**29.68**
MSSIM	0.593	0.824	0.839	0.887	<u>0.922</u>	0.892	**0.934**
MSA	24.68	12.73	10.46	8.634	<u>7.113</u>	8.249	**6.252**
Time/s	–	297.8	431.5	389.2	475.6	**5.5**	<u>12.6</u>

with stripe noise. 10 bands (Band 17, 27, …, 107) of the original HSIs were imposed with simulated stripe noise, where the stripe intervals, locations, and intensity are different for each selected band, through scoped random values of rate (0.3 ~ 0.6) and mean (0.05 ~ 0.25) (Zhang et al., 2019).

The quantitative evaluations and consuming-time of the five algorithms and the proposed method in the W. DC and Pavia University HSIs are listed in Table 1 and Table 2, respectively. Notably, the best indexes are highlighted with the bold type, and the second indexes are marked with underline type. Besides, the six denoising results of Case 2 (non-i.i.d. Gaussian noise) and Case 3 (non-i.i.d. Gaussian noise + stripe noise) on W. DC data, and Case 2 (non-i.i.d. Gaussian noise) on Pavia University data are given in Figs. 2–4 with pseudo-color or gray color, respectively. To better distinguish the details of the restoring results, the magnification maps of the local area are also shown in the lower of Figs. 2–4, respectively.

In Case 1 (i.i.d. Gaussian noise) experiments, several HSI denoising methods like LRMR, NMoG, SSGN and proposed method can all obtain qualified results. NMoG performs the best on MSSIM and MSA in W. DC data, and MSSIM in Pavia University data. SSGN performs the best on MPSNR in Pavia University data, due to the HSI i.i.d. noise hypothesis. Particularly, although the proposed method just employs the non-i.i.d. noise simulated training samples, it can also reach the approximate level of the i.i.d. noise-based methods like SSGN. To a certain extent, this case demonstrates that HSI non-i.i.d. noise strategy embedding the proposed method has a downward compatibility for i.i.d. noise. By means of noise estimation and distribution thoughts under the spatiospectral Bayesian posterior framework, i.i.d. noise in HSIs can also be well solved even though training with the non-i.i.d. noise HSI samples. The noise modeling operations are indirectly beneficial for noise reduction in HSI.

In Case 2 (non-i.i.d. Gaussian noise), though the HSSNR approach reveals a certain noise reduction effect under complex noise scene, as displayed in Figs. 2(b) and 4(b), it lacks the qualified capacity of disposing noisy bands with non-i.i.d. noise, and two HSI denoising outcomes still involve evident residuary noise. For BM4D method, it can overall removal the disturbed noise in HSI as responded in Figs. 2(c) and 4(c). However, the over-smoothing issue also obviously exists in the denoising results whose spatial textures are missing and blurry, on account of the non-local blocks mean thought in BM4D. LRMR and NMoG generate decent recovering outcomes in Figs. 2 and 4 through utilizing the low-rank prior of matrix or tensor. Whereas spectral distortion still exists to some degree especially in the magnified regions of Fig. 2(d)–(e) and Fig. 4(d)–(e), respectively. Besides, these model-driven based methods are time-consuming with hundreds of seconds, as recorded in Tables 1 and 2. For SSGN model, the artifacts and residual noise distribute in both global results and local amplifying regions, as
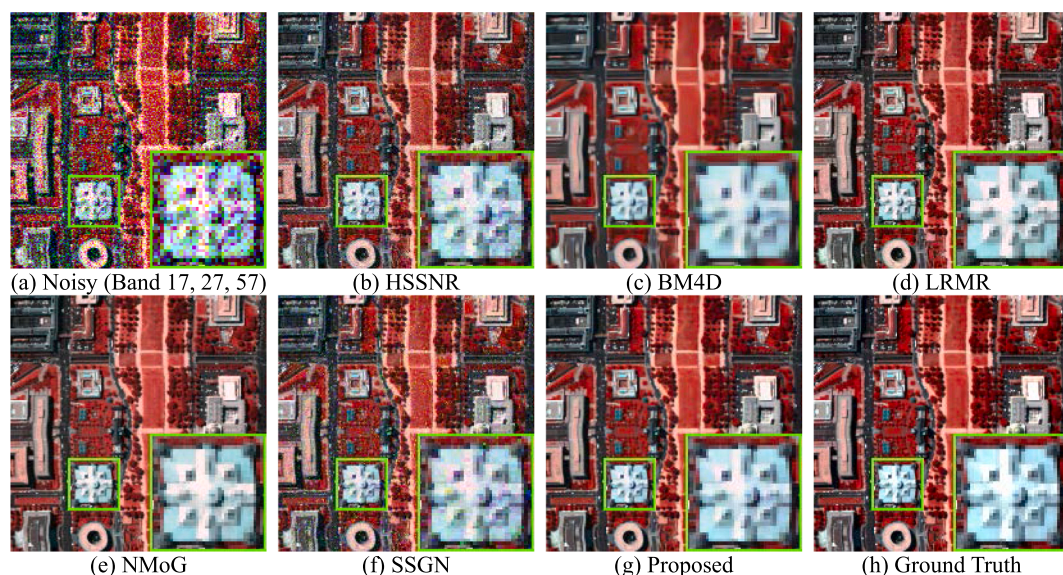

(a) Noisy (Band 17, 27, 57)   (b) HSSNR   (c) BM4D   (d) LRMR

(e) NMoG   (f) SSGN   (g) Proposed   (h) Ground Truth

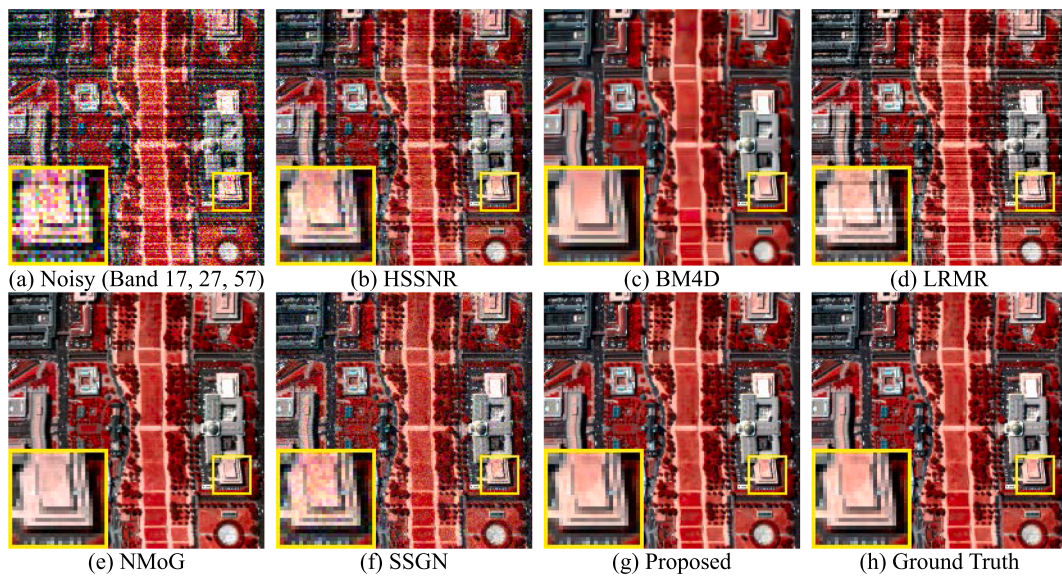**Fig. 2.** Non-i.i.d. noise removing results of Case 2 in W.DC data simulated experiments.

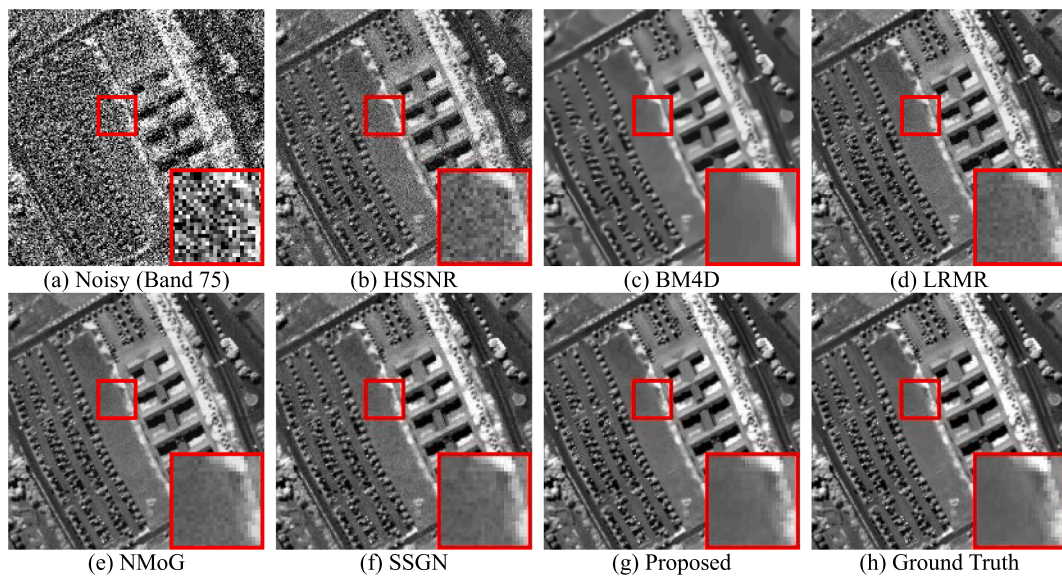**Fig. 3.** Non-i.i.d. + stripe noise removing results of Case 3 in the W.DC data simulated experiments.

Fig. 4. Non-i.i.d. noise removing results of Case 2 in the Pavia data simulated experiments.

(1) Simulated noise level for each band (2) SSIM denoising result for each band
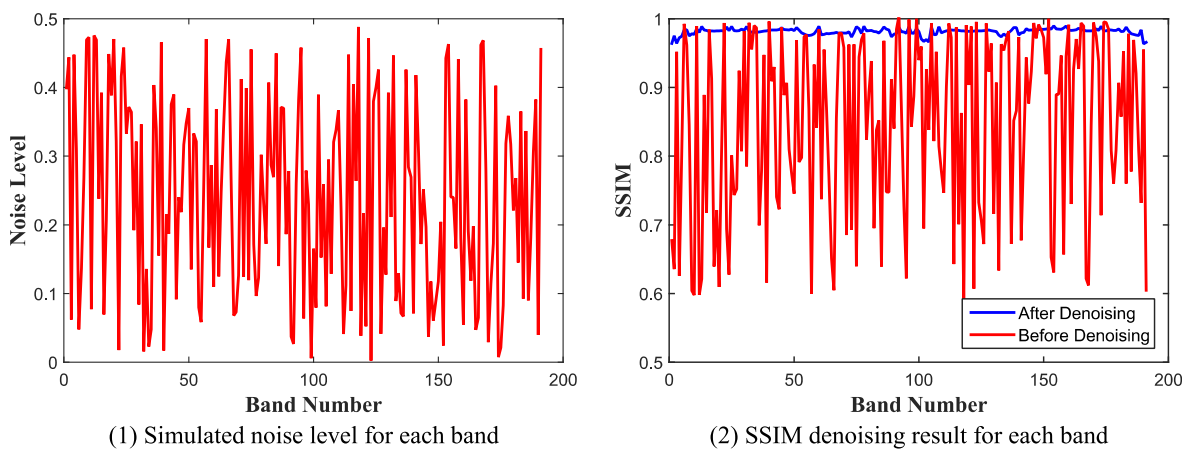
Fig. 5. The range of noise level and SSIM denoising result in WDC simulated HSI datasets.

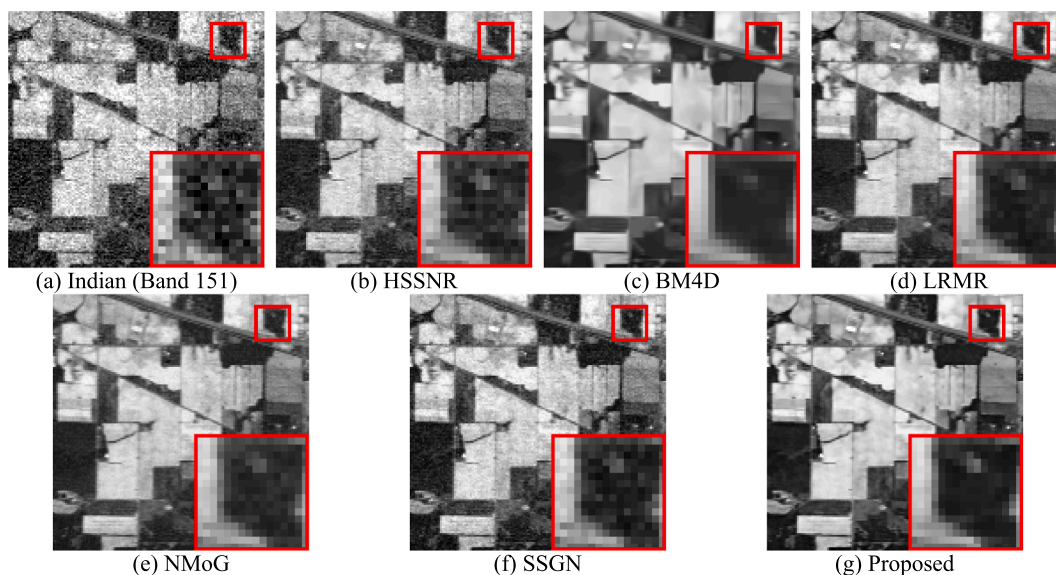Fig. 6. Noise removing pseudo-color outcomes of the Indian data in the real experiments.

Fig. 7. Noise removing gray-color outcomes of the Indian data in the real experiments.

shown in Figs. 2(f) and 4(f). It seems to apparently overfit in training samples by the hypothesis of i.i.d. noise in HSIs. In terms of the proposed method for non-i.i.d. noise removal, it performs best on MPSNR, MSSIM and MSA in both W.DC and Pavia University data, as listed in Tables 1 and 2. In addition, the proposed method also behaves stronger on spectral preservation and noise removal compared with other five HSI denoising algorithms, as depicted in Figs. 2(g) and 4(g).

In Case 3 (non-i.i.d. Gaussian noise + stripe noise) experiments, HSSNR and LRMR cannot eliminate the stripe noise, as depicted in Fig. 3(b) and (d). For BM4D approach, it can overall removal the mixed noise in HSI as displayed in Fig. 3(c). However, the over-smoothing issue still obviously exists in the denoising results whose spatial textures are missing and blurry. NMoG can well deal with the mixed noise in Fig. 3(e). Nevertheless, the spectral distortion obviously emerges to some degree especially within magnified areas of Fig. 3(e). In terms of the proposed method for non-i.i.d. noise and stripe noise removal, it performs best on MPSNR, MSSIM and MSA in both W.DC and Pavia University data, as listed in Tables 1 and 2. In addition, the proposed method behaves stronger on spectral preservation and noise removal

compared with other five HSI denoising algorithms, as depicted in Fig. 3(g). This also verifies the availability of sparse noise TV term to constraint stripe noise, and the effectiveness of non-i.i.d. noise estimation and distribution thoughts by the spatio-spectral Bayesian posterior framework.

Besides, consuming-time of the five algorithms and the proposed method in the W. DC and Pavia University HSI are both recorded in Tables 1-2, respectively. On the one hand, contrasted with the model-driven based approaches like LRMR or NMoG, the presented framework manifests great superiority in terms of processing efficiency for HSI denoising. On the other hand, compared with the data-driven based approaches like SSGN method, the presented framework outperforms in terms of the HSI non-i.i.d. noise removal under the same environment. For each band in WDC simulated HSI, the specific non-i.i.d. noise variance σ_b^2 are selected from the range of [0–0.5] through uniform random distribution, as shown in Fig. 5 (1). The SSIM results of before and after denoising for each band are displayed in Fig. 5 (2). The proposed method shows the stabilization for both high- and low-level noise removal in HSI denoising procedure.
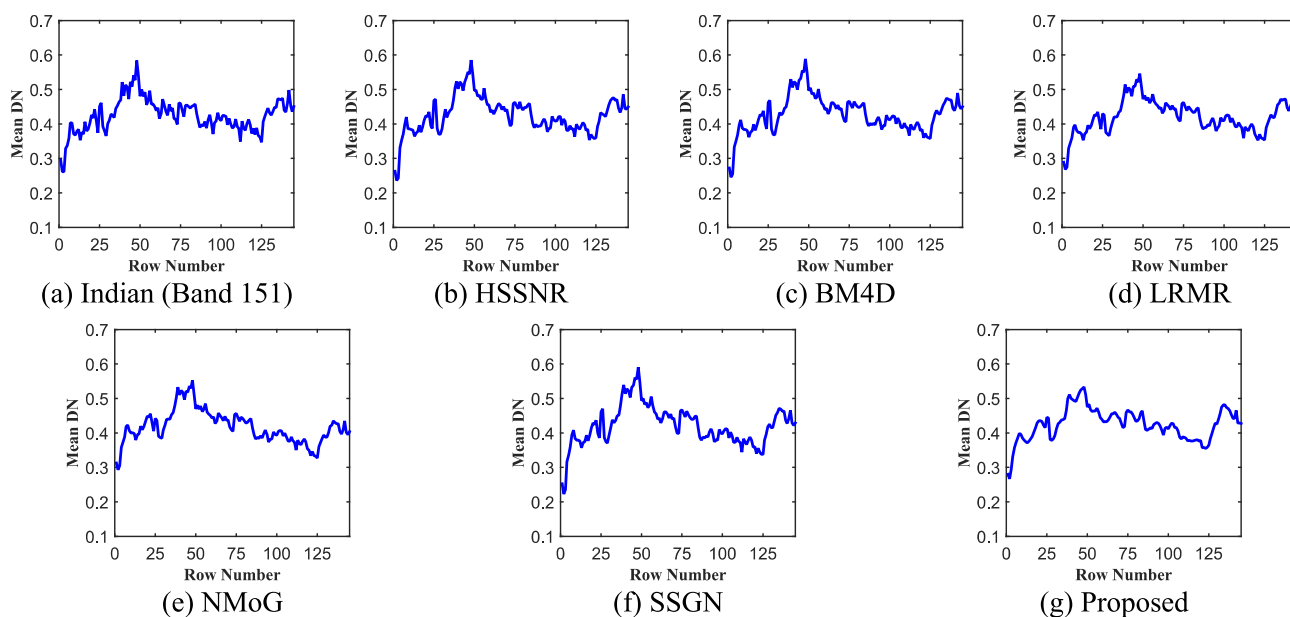
(a) Indian (Band 151)   (b) HSSNR   (c) BM4D   (d) LRMR

(e) NMoG   (f) SSGN   (g) Proposed

Fig. 8. Horizonal mean digital number curves within band 151 of Indian.

(a) Urban (Band 24, 104,   (b) HSSNR   (c) BM4D   (d) LRMR
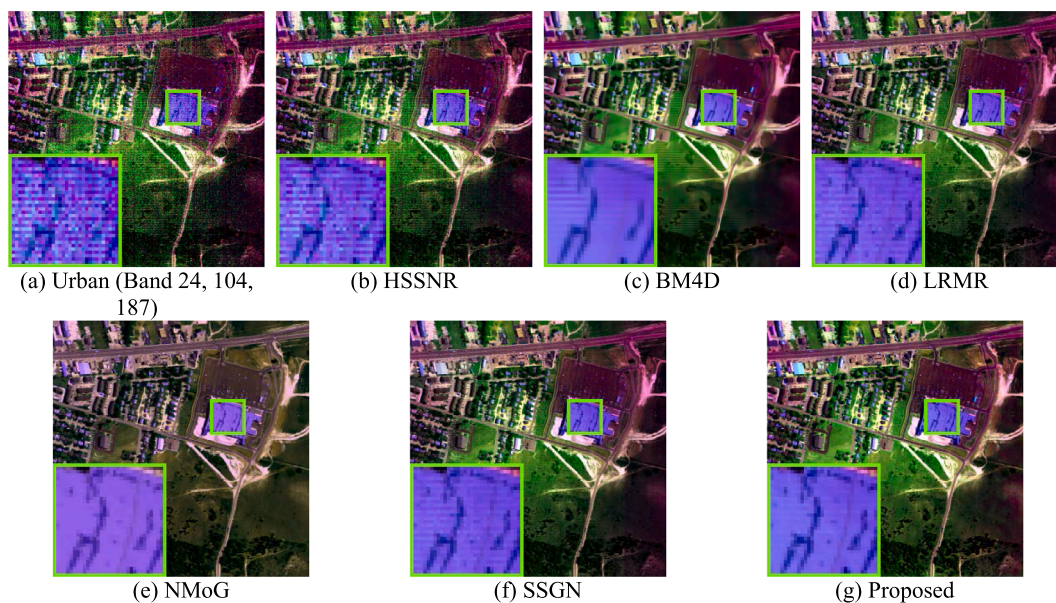187)

(e) NMoG   (f) SSGN   (g) Proposed

Fig. 9. Noise removing pseudo-color results of the Urban data real experiments.

4.4. Real experiments

In the real experiments, five actual HSIs (Indian Pines, Urban, Mars, EO-1 and Zhuhai-1 data set) (Huang et al., 2019) polluted by noise are employed as the testing data to inspect the practicability of the presented approach. It must be noted that noise types and distribution are all complicated and non-i.i.d. in these three data sets. The Indian HSI is principally contaminated with Gaussian and Poisson mixed noise. The Urban HSI is mainly degraded by stripe and Gaussian mixed noise. While in the Mars data, due to its complicated observation conditions, it includes more complex stripe noise and Gaussian noise. Through testifying these data sets embedding with different type noise, we can better distinguish the ability of the proposed method for non-i.i.d. noise removal in different HSIs.

In addition, five HSI denoising methods HSSNR, BM4D, LRMR, NMoG and SSGN are served as the contrast algorithms for better comparison. The false-color and gray-color results are both shown to reveal

the ability of spatial structure recovery and spectral information preservation. To further investigate the reconstruction quality of the HSI denoising results, the horizonal normalized mean digital number (DN) profiles of special band in the two data sets are given through different denoising methods. And finally, the consuming times of each algorithm are checked to demonstrate the efficiency for HSI denoising under the same operating environment.

(1) Indian Pines data set: The Indian HSI is obtained through the Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) and mainly contaminated with Gaussian and Poisson mixed noise. The initial Indian HSI composes of 220 spectrums, whose spatial size denotes 145×145. While some bands are severely corrupted by atmosphere scattering and water absorbing phenomenon. After eliminating these useless spectrums, 206 bands of the original Indian Pines data are testified through six algorithms HSSNR, BM4D, LRMR, NMoG, SSGN and the proposed method. The noise removing pseudo-color results of bands (2, 24, 145) and gray-color results of band 151 in this data set are given in
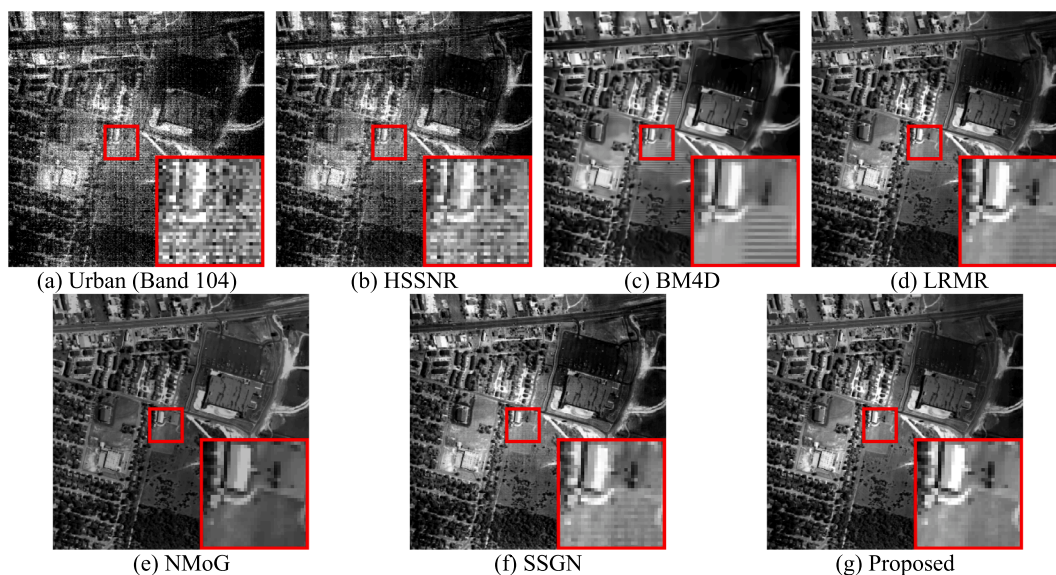
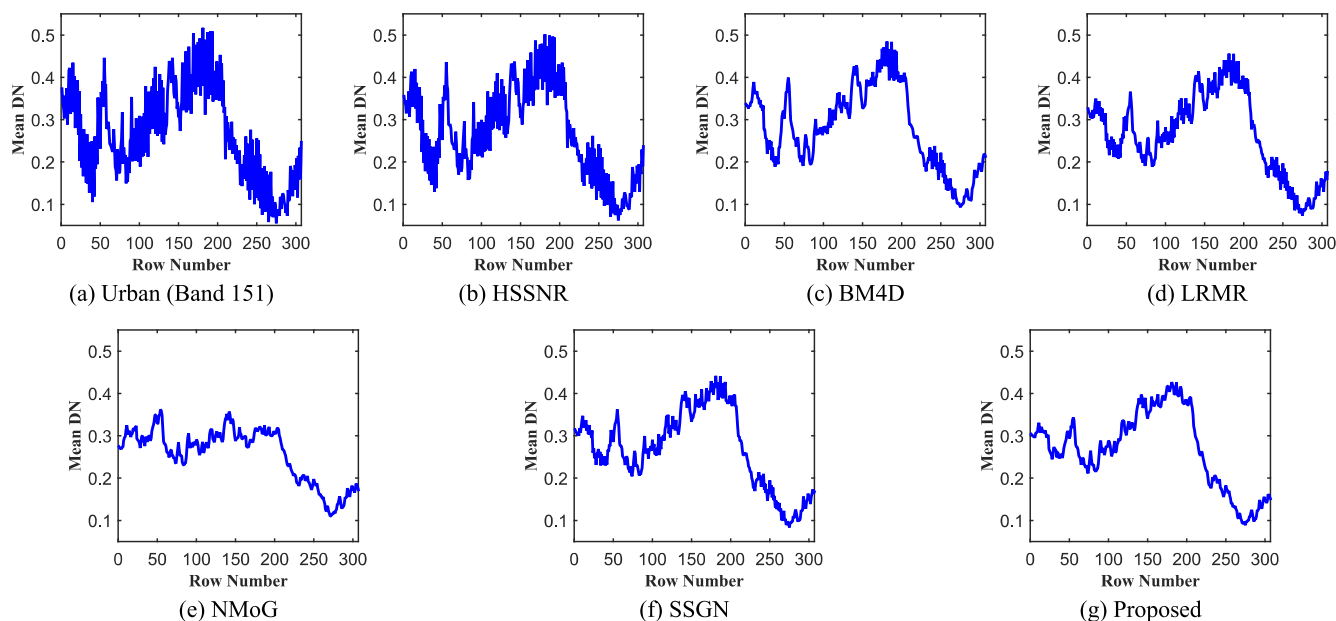**Fig. 10.** Noise removing gray-color results of the Urban data real experiments.

Fig. 11. Horizonal mean digital number curves within band 104 of Urban.

Fig. 6(b)–(g) and Fig. 7(b)–(g), respectively. To better distinguish the details of the restoring results, the magnification maps of the selected local area are also shown in the lower of Figs. 6–7, respectively.

HSSNR method reveals a certain noise reduction effect under complex noise scene, as displayed in Figs. 6(b) and 7(b). It lacks qualified capacity of disposing noisy bands with complex noise, as well as the two HSI denoising outcomes still involve evident residuary noise. For BM4D, it can overall removal the disturbed noise in HSI as responded in Figs. 6–7(c). However, the over-smoothing issue also obviously exists in the denoising results whose spatial textures are missing and blurry, on account of the non-local blocks mean thought in BM4D. LRMR and NMoG generate decent recovering outcomes in Figs. 6–7 through utilizing the low-rank prior of matrix or tensor, whereas obvious spectral distortion still exists to some degree especially in the magnified regions of Fig. 6(d)–(e), respectively. Besides, For SSGN method, the artifacts and residual noise distribute in the global results and local amplifying regions, as shown in Figs. 6(f) and 7(f). It seems to apparently overfit in training samples by the hypothesis of i.i.d. noise in HSIs. In terms of the

proposed method aiming at realistic noise removal, it performs the best on spatial structure recovery and spectral information preservation in Indian Pines data, as displayed in Figs. 6(g) and 7(g).

In addition, the horizonal mean digital number curves of special spectrum within the Indian Pines HSI are given through different denoising methods in Fig. 8. The horizontal axis in Fig. 8(a)–(g) stands for the row number of Indian Pines data set. And the ordinate coordinate denotes as the average normalized DN in per row. As depicted in Fig. 8(a), due to the severe noise interference, violent undulations can be noticed within the original profile. Experiencing the recovery procedure, the reconstruction profile of the presented approach behaves smoother on horizonal mean DN profiles in band 151 compared with other five HSI denoising algorithms, as depicted in Fig. 8(g). This also affirms the effectivity of the presented approach for non-i.i.d. noise reduction in Indian Pines data set.

(2) Urban data set: The Urban HSI is captured through the Hyperspectral Digital Imagery Collection Experiment (HYDICE) and mainly degraded by stripe and Gaussian mixed noise. The noise
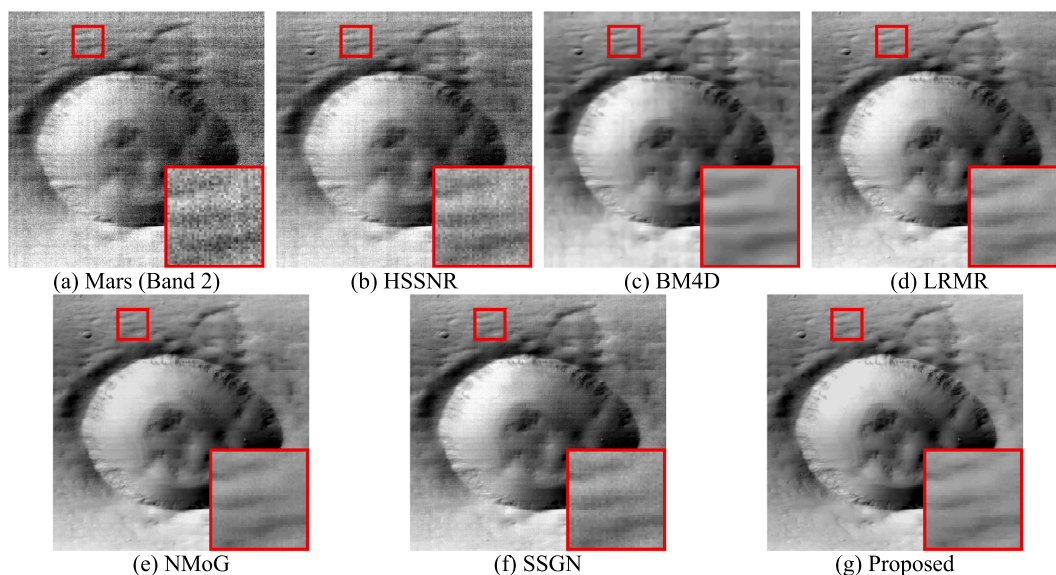
Fig. 12. Noise removing results in Band 2 of the Mars data real experiments.
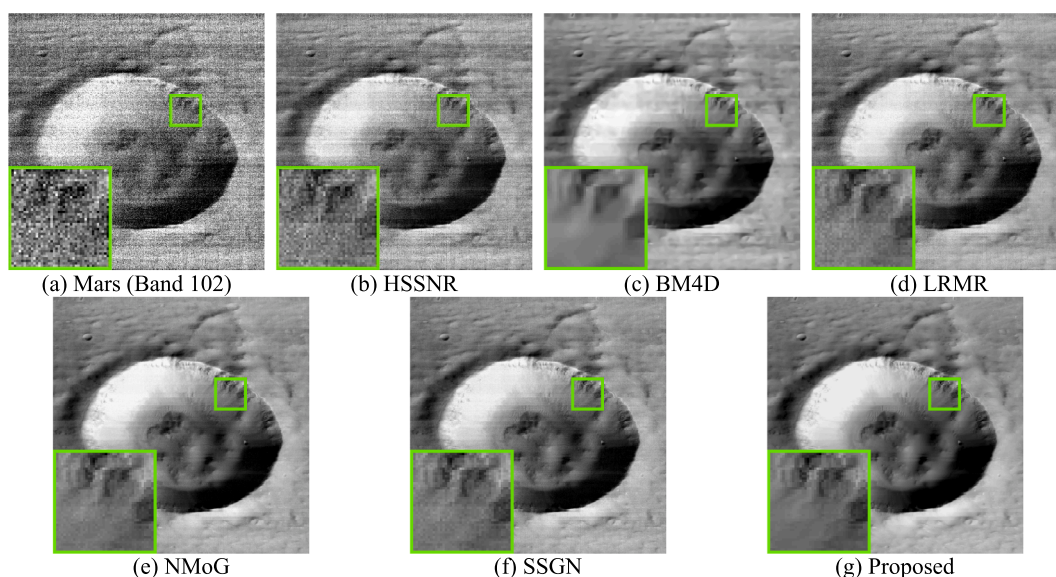

Fig. 13. Noise removing results in Band 102 of the Mars data real experiments.

removing pseudo-color results of bands (24, 104, 187) and gray-color results of band 104 in this data set are given in Figs. 9 and 10, respectively.

HSSNR still contains large area noise as shown in Figs. 9(b) and 10(b). For BM4D approach, it can partly removal the mixed noise in HSI as displayed while residual stripe noise still exists. Overall, LRMR and SSGN perform well on spectral preservation and noise suppression, while they are not able to well eliminate stripe noise, as shown in Fig. 9(d) and (f). NMoG can well deal with the mixed noise in Figs. 9(e) and 10(e). Nevertheless, the spectral distortion obviously emerges to some degree, especially highlighted within the magnified areas of Fig. 9(e). And in terms of the proposed approach, it performs the best on spatial structure recovery and spectral information preservation, as displayed in Figs. 9(g) and 10(g).

In addition, the horizontal mean digital number curves of special spectrum within Urban HSI are given through different denoising methods. The abscissa axis in Fig. 11(a)–(g) stands for the row number of Urban data set. And the vertical axis denotes as the average normalized DN in per row. As depicted in Fig. 11(a), because of hybrid

noise pollution, intense undulations can be observed within the original profile. Undergoing the noise removal procedure, the reconstruction profile of the presented approach behaves smoother in band 104, compared with other five HSI denoising algorithms as depicted in Fig. 11(g). This also manifests the availability of the presented framework for non-i.i.d. noise reduction in Urban HSI.

(3) Mars data set: The Mars HSI is observed through the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM), whose cube size of denotes $420 \times 420 \times 102$. Due to its complicated observation conditions, it includes more complex stripe noise and Gaussian noise. Then the noisy data is processed by six algorithms HSSNR, BM4D, LRMR, NMoG, SSGN and the proposed method. The noise removing gray-color results of band 2 and 102 embedding with mixed noise in this data set are given in Figs. 12(b)–(g) and 13(b)–(g), respectively. To better distinguish the details of the restoring results, the magnification maps of the selected local area are also shown in the lower of Figs. 12–13, respectively.

Overall, HSSNR, BM4D, LRMR, NMoG and SSGN perform well on noise reduction, but they cannot integrally eliminate the mixed noise
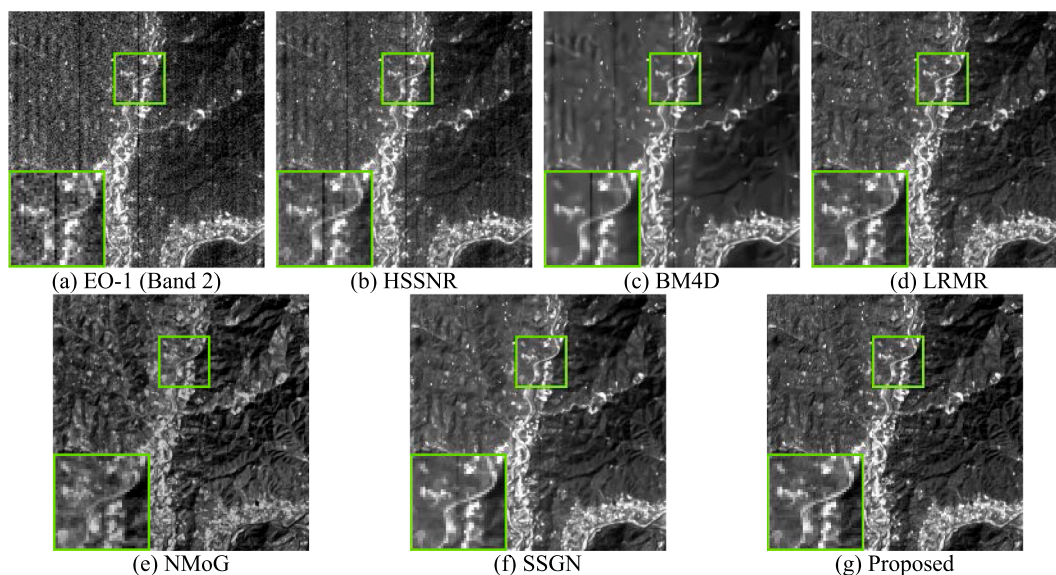
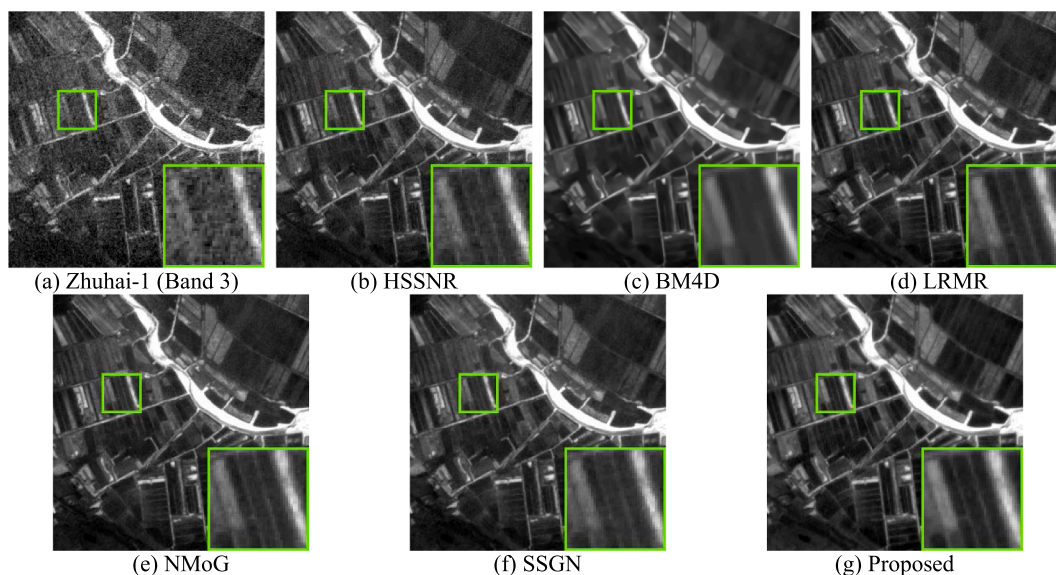Fig. 14. Noise removing results in Band 2 of the EO-1 data real experiments.

Fig. 15. Noise removing results in Band 3 of the Zhuhai-1 data real experiments.

especially for HSSNR method, as displayed in Figs. 12 and 13(b)–(f). It indicates that these model through i.i.d. noise assumption or training samples intend to overfit the non-i.i.d. noise in Mars data set, to different degrees. Compared with all the contrast methods, the proposed method steadily removals the blended noise, while concurrently ensures the spatial texture and spectral consistency, without drawing into evident artifacts or residual noise, as depicted in Figs. 12(g) and 13(g).

(4) EO-1 dataset: The EO-1 HSI is captured through the Hyperion sensor and mainly degraded by stripe and Gaussian mixed noise. The noise removing gray-color results of band 2 in this data set are given in Fig. 14, with the size of $200 \times 200 \times 166$.

HSSNR still contains large area noise as shown in Fig. 14(b). For BM4D, it can partly removal the mixed noise in HSI as displayed while residual stripe noise still exists. LRMR and SSGN still have some stripe noise, as shown in Fig. 14(d) and (f). NMoG can well deal with the mixed noise in Fig. 14(e). Nevertheless, some spatial artifacts obviously emerge, especially in the magnified areas of Fig. 14(e). And for the proposed approach, it outperforms on spatial structure recovery and mixed noise removal, as displayed in Fig. 14(g).

(5) Zhuhai-1 dataset: The Zhuhai-1 HSI is obtained with 32 bands and mainly contaminated with random noise. The noise removing results of band 3 in this data set are given in Fig. 15(b)–(g), respectively. To better distinguish the details, the magnification maps of the selected area are also shown in the lower of Fig. 15.

HSSNR lacks the qualified capacity of disposing noisy bands with complex noise, whose outcome still involves evident residuary noise in Fig. 15(b). For BM4D, it can overall removal the disturbed noise in HSI as responded in Fig. 15(c). However, the over-smoothing issue also obviously exists in the denoising result. For LRMR, NMoG and SSGN methods, the residual noise still distributes in the global results and local amplifying regions, as shown in Fig. 15(d)–(f). In terms of the proposed method aiming at realistic noise, it performs the best both on spatial details preservation and non-i.i.d. noise reduction in Zhuhai-1 data, as displayed in Fig. 15(g).

(6) Blind quality contrast: The blind performance evaluation comparison index (mean Q-metric) of five real experiments has been provided in this section, as listed in Table 3. Q-metric is used to evaluate the image quality of spatial information preservation. The higher

Table 3
Blind quality assessment using the mean Q-metric in the real experiments

Data set	HSSNR	BM4D	LRMR	NMoG	SSGN	Proposed
Urban	0.043	0.064	0.061	0.072	0.068	**0.075**
Indian Pines	0.056	0.071	0.075	0.079	0.082	**0.084**
Mars	0.048	0.062	0.066	0.074	0.069	**0.076**
EO-1	0.046	0.043	0.065	0.059	0.068	**0.071**
Zhuhai-1	0.059	0.066	0.074	0.078	0.075	**0.086**

Table 4
Time-consuming contrast in the real experiments (seconds)

Data set	HSSNR	BM4D	LRMR	NMoG	SSGN	Proposed
Urban	597.8	936.2	1155.6	1325.4	**15.6**	34.2
Indian Pines	162.7	236.3	268.7	348.2	**3.9**	10.7
Mars	2279.4	4185.7	3723.5	5297.6	**21.4**	48.6
EO-1	183.6	305.4	291.2	484.3	**4.2**	13.5
Zhuhai-1	224.7	156.2	184.3	261.9	**3.1**	8.7

Q-metric is, the better denoising result is. Compared with model-driven and data-driven based approaches, the proposed method manifests the HSI non-i.i.d. noise removing ability, by means of the combination of both model-driven and data-driven strategy.

(7) Time-consuming: To measure the execution efficiency by above-mentioned six HSI denoising methods, the consuming times of the five data sets are collected through the uniform executing conditions (Operating system: Windows 10, MATLAB R2018a, Python 3.7.0, CPU: Intel E5-2609, GPU: NVIDA TITAN X), as depicted in Table 4. Compared with model-driven based approaches like LRMR or NMoG, the presented approach manifests great superiority in terms of processing efficiency for HSI denoising. Comprehensively, the proposed method shows the high efficiency due to the superiority of data-driven strategy.

5. Conclusion

In this work, a novel deep spatio-spectral Bayesian posterior (DSSBP) framework in HSI is proposed for non-i.i.d. noise removal. The proposed method unites intrinsic merits of both model-driven and data-driven methods through collaboratively models the non-i.i.d. noise embedding in HSI and removals them under the spatio-spectral Bayesian posterior structure. Specifically, the noise modeling and removing procedure are both executed with the spatio-spectral convolutional neural network. Experiments on different HSI non-i.i.d. noise scenarios testify that the presented approach outperforms other contrastive algorithms for hybrid noise removal.

In our future work, we will exploit the low-rank constraint of 3-D tensor, to better model the complicated noise structure in HSIs. Besides, the efficiency of the presented framework needs to be further enhanced under complex noise scenario.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

Asaari, M., et al., 2018. Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping platform. ISPRS J. Photogramm. Remote Sens. 138, 121–138.

Brell, M., Segl, K., Guanter, L., Bookhagen, B., 2019. 3D hyperspectral point cloud generation: fusing airborne laser scanning and hyperspectral imaging sensors for improved object-based information extraction. ISPRS J. Photogramm. Remote Sens. 149, 200–214.

Chang, Y., Yan, L., Fang, H., Zhong, S., Liao, W., 2019. HSI-DeNet: Hyperspectral image restoration via convolutional neural network. IEEE Trans. Geosci. Remote Sens. 57, 667–682.

Chen, Y., Cao, X., Zhao, Q., Meng, D., Xu, Z., 2018. Denoising hyperspectral image with non-iid noise structure. IEEE Trans. Cybern. 48, 1054–1066.

Chen, Y., Guo, Y., Wang, Y., Wang, D., Peng, C., He, G., 2017. Denoising of hyperspectral images using nonconvex low rank matrix approximation. IEEE Trans. Geosci. Remote Sens. 55, 5366–5380.

Chen, Y., He, W., Yokoya, N., Huang, T., 2019a. Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition. IEEE Trans. on Cybern. DOI: 10.1109/TCYB.2019.2936042.

Chen, Y., He, W., Yokoya, N., Huang, T., 2019b. Blind cloud and cloud shadow removal of multitemporal images based on total variation regularized low-rank sparsity decomposition. ISPRS J. Photogramm. Remote Sens. 157, 93–107.

Dong, W., Wang, H., Wu, F., Shi, G., Li, X., 2019. Deep spatial-spectral representation learning for hyperspectral image denoising. IEEE Trans. Comput. Imaging. 5, 635-648.

Fan, H., Chen, Y., Guo, Y., Zhang, H., Kuang, G., 2017. Hyperspectral image restoration using low-rank tensor recovery. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 10, 4589–4604.

Fan, H., Li, C., Guo, Y., Kuang, G., Ma, J., 2018. Spatial-spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising. IEEE Trans. Geosci. Remote Sens. 56, 6196–6213.

Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L., 2019. Toward convolutional blind denoising of real photographs. Proc. Conf. Comput. Vis. Pattern Recognit. (CVPR) 1712–1722.

Guo, X., Huang, X., Zhang, L., Zhang, L., 2013. Hyperspectral image noise reduction based on rank-1 tensor decomposition. ISPRS J. Photogramm. Remote Sens. 83, 50–63.

He, W., Zhang, H., Zhang, L., Shen, H., 2016. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. IEEE Trans. Geosci. Remote Sens. 54, 178–188.

Hong, D., Yokoya, N., Chanussot, J., Xu, J., Zhu, X., 2019. Learning to propagate labels on graphs: an iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction. ISPRS J. Photogramm. Remote Sens. 158, 35–49.

Huang, H., Duan, Y., He, H., Shi, G., Luo, F., 2019. Spatial-spectral local discriminant projection for dimensionality reduction of hyperspectral image. ISPRS J. Photogramm. Remote Sens. 156, 77–93.

Karami, A., Yazdi, M., Asli, A., 2011. Noise reduction of hyperspectral images using kernel non-negative tucker decomposition. IEEE J. Sel. Topics Signal Process. 5, 487–493.

Kingma, D., Ba, J., 2014. Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., Schindler, K., 2018. Super-resolution of Sentinel-2 images: learning a globally applicable deep neural network. ISPRS J. Photogramm. Remote Sens. 146, 305–319.

LeCun, Y., et al., 1990. Handwritten digit recognition with a back-propagation network. Proc. Adv. Neural Inf. Process. Syst. (NeurIPS) 396–404.

Li, C., Ma, Y., Huang, J., Mei, X., Ma, J., 2015. Hyperspectral image denoising using the robust low-rank tensor recovery. JOSA A 32, 1604–1612.

Liu, W., Lee, J., 2019. A 3-D Atrous convolution neural network for hyperspectral image denoising. IEEE Trans. Geosci. Remote Sens. 57, 5701–5715.

Lu, T., Li, S., Fang, L., Ma, Y., Benediktsson, J., 2016. Spectral-spatial adaptive sparse representation for hyperspectral image denoising. IEEE Trans. Geosci. Remote Sens. 54, 373–385.

Maggioni, M., Katkovnik, V., Egiazarian, K., Foi, A., 2012. A nonlocal transform-domain filter for volumetric data denoising and reconstruction. IEEE Trans. Image Process. 22, 119–133.

Othman, H., Qian, S., 2006. Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage. IEEE Trans. Geosci. Remote Sens. 44, 397–408.

Paoletti, M., Haut, J., Plaza, J., Plaza, A., 2018. A new deep convolutional neural network for fast hyperspectral image classification. ISPRS J. Photogramm. Remote Sens. 145, 120–147.

Qian, Y., Ye, M., 2013. Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 6, 499–515.

Rasti, B., Scheunders, P., Ghamisi, P., Licciardi, G., Chanussot, J., 2018. Noise reduction in hyperspectral imagery: overview and application. Remote Sens. 10, 482.

Roshan, Pande-Chhetri, Amr, Abd-Elrahman, 2011. De-striping hyperspectral imagery using wavelet transform and adaptive frequency domain filtering. ISPRS J. Photogram. Remote Sensing 66, 620–636. http://www.sciencedirect.com/science/article/pii/S0924271611000530.

Sidike, P., Asari, V., Sagan, V., 2018. Progressively expanded neural network (PEN Net) for hyperspectral image classification: a new neural network paradigm for remote sensing image analysis. ISPRS J. Photogramm. Remote Sens. 146, 161–181.

Sun, W., Yang, G., Wu, K., Li, W., Zhang, D., 2017a. Pure endmember extraction using

robust kernel archetypoid analysis for hyperspectral imagery. ISPRS J. Photogramm. Remote Sens. 131, 147–159.

Sun, L., Jeon, B., Zheng, Y., Wu, Z., 2017b. Hyperspectral image restoration using low-rank representation on spectral difference image. IEEE Geosci. Remote Sens. Lett. 14, 1151–1155.

Sun, L., Zhan, T., Wu, Z., Xiao, L., Jeon, B., 2018. Hyperspectral mixed denoising via spectral difference-induced total variation and low-rank approximation. Remote Sens. 10, 1956.

Wang, Y., Yuan, Q., Li, T., Shen, H., Zheng, L., Zhang, L., 2019. Large-scale MODIS AOD products recovery: spatial-temporal hybrid fusion considering aerosol variation mitigation. ISPRS J. Photogramm. Remote Sens. 157, 1–12.

Wu, C., Du, B., Zhang, L., 2018. Hyperspectral anomalous change detection based on joint sparse representation. ISPRS J. Photogramm. Remote Sens. 146, 137–150.

Xie, Q., Zhao, Q., Meng, D., Xu, Z., 2018. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. IEEE Trans. Pattern Anal. Mach. Intell. 40, 1888–1902.

Xie, W., Li, Y., 2017. Hyperspectral imagery denoising by deep learning with trainable nonlinearity function. IEEE Geosci. Remote Sens. Lett. 14, 1963–1967.

Xing, Y., Wang, M., Yang, S., Jiao, L., 2018. Pan-sharpening via deep metric learning. ISPRS J. Photogramm. Remote Sens. 145, 165–183.

Xiong, F., Zhou, J., Qian, Y., 2019. Hyperspectral restoration via L0 gradient regularized low-rank tensor factorization, IEEE Trans. Geosci. Remote Sens. 57, 10410-10425.

Xu, Z., et al., 2019. Deep gradient prior network for DEM super-resolution: Transfer learning from image to DEM. ISPRS J. Photogramm. Remote Sens. 150, 80–90.

Xue, J., Zhao, Y., Liao, W., Kong, S., 2018. Joint spatial and spectral low-rank regularization for hyperspectral image denoising. IEEE Trans. Geosci. Remote Sens. 56, 1940–1958.

Yuan, Q., Zhang, L., Shen, H., 2012. Hyperspectral image denoising employing a spectral-spatial adaptive total variation model. IEEE Trans. Geosci. Remote Sens. 50, 3660–3677.

Yuan, Q., Zhang, Q., Li, J., Shen, H., Zhang, L., 2019. Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network. IEEE Trans. Geosci. Remote Sens. 57, 1205–1218.

Yue, Z., Meng, D., Sun, Y., Zhao, Q., 2018. Hyperspectral image restoration under

complex multi-band noises. Remote Sens. 10, 1631.

Yue, Z., Yong, H., Meng, D., Zhao, Q., Leung, Y., Zhang, L., 2020. Robust multiview subspace learning with nonindependently and nonidentically distributed complex noise, IEEE Trans. Neural Netw. Learn. Syst. 31, 1070-1083.

Yue, Z., Yong, H., Zhao, Q., Zhang, L., Meng, D., 2019b. Variational denoising network: Toward blind noise modeling and removal. Proc Adv. Neural Inf. Process. Syst. (NeurIPS).

Zhang, H., He, W., Zhang, L., Shen, H., 2014. Hyperspectral image restoration using low-rank matrix recovery. IEEE Trans. Geosci. Remote Sens. 52, 4729–4743.

Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L., 2017. Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process. 26, 3142–3155.

Zhang, Q., Yuan, Q., Zeng, C., Li, X., Wei, Y., 2018a. Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network. IEEE Trans. Geosci. Remote Sens. 56, 4274–4288.

Zhang, Q., Yuan, Q., Li, J., Yang, Z., Ma, X., 2018b. Learning a dilated residual network for SAR image despeckling. Remote Sens. 10, 196.

Zhang, K., Zuo, W., Zhang, L., 2018c. FFDNet: toward a fast and flexible solution for CNN-based image denoising. IEEE Trans. Image Process. 27, 4608–4622.

Zhang, Q., Yuan, Q., Li, J., Liu, X., Shen, H., Zhang, L., 2019. Hybrid noise removal in hyperspectral imagery with a spatial-spectral gradient network. IEEE Trans. Geosci. Remote Sens. 57, 7317–7329.

Zhang, Q., Yuan, Q., Li, J., Li, Z., Shen, H., Zhang, L., 2020. Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning. ISPRS J. Photogramm. Remote Sens. 162, 148–160.

Zhao, Y., Yang, J., 2015. Hyperspectral image denoising via sparse representation and low-rank constraint. IEEE Trans. Geosci. Remote Sens. 53, 296–308.

Zheng, X., Yuan, Y., Lu, X., 2019. Hyperspectral image denoising by fusing the selected related bands. IEEE Trans. Geosci. Remote Sens. 57, 2596–2609.

Zhu, X., et al., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5, 8–36.

Zhuang, L., Bioucas-Dias, J., 2018. Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 11, 730–742.