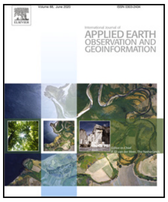
International Journal of Applied Earth Observation and Geoinformation 123 (2023) 103481

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



A hyperspectral image denoising method based on land cover spectral autocorrelation

Shuheng Zhao [a], Xiaolin Zhu [a,*], Denghong Liu [a], Fei Xu [a], Yan Wang [a], Liupeng Lin [b], Xuehong Chen [c], Qiangqiang Yuan [d]

[a] *Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China*
[b] *School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China*
[c] *State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Science and Engineering, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*
[d] *School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China*

ARTICLE INFO

Keywords:
Hyperspectral remote sensing
Image restoration
Convolutional neural network
Transformer
Spectral unmixing analysis
Noise removal

ABSTRACT

Developing denoising algorithms for hyperspectral remote sensing images (HSIs) can alleviate noise problem, improve data utilization as well as the accuracy of subsequent applications. However, existing denoising techniques are usually unstable due to the variations of landscapes, resulting in local distortion of HSIs, especially in heterogeneous areas. To tackle this issue, we propose a spatial–spectral interactive restoration (SSIR) framework by exploiting the complementarity of model-based and data-driven methods. Specifically, a deep learning-based denoising module that incorporates both convolutional neural networks (CNN) and Swin Transformer (TF) blocks is designed. This denoiser can achieve local–global dependencies modeling and content-based interactions to better capture global heterogeneity differences in HSIs. Moreover, we introduce an unsupervised unmixing module that utilizes spectral autocorrelation as prior information to effectively capture the differences in reflectance characteristics among different land cover components. This parameter-free module further improves the generalization ability of SSIR and enables stable denoising performance across different scenarios. Both modules are iteratively updated and fuel each other in SSIR. The proposed SSIR is shown to outperform others in preserving spatial details, maintaining spectral fidelity, and adapting to different landscapes based on simulated and real experiments conducted on various HSIs under diverse noise conditions.

1. Introduction

Hyperspectral remote sensing imagery (HSI) is a powerful tool in various fields of Earth system science (Van der Meer et al., 2012), including environmental monitoring (Kereszturi et al., 2018) and vegetation survey (Carvalho et al., 2013), etc. HSI provides detailed spectral signatures and spatial distributions of land surface components. However, the high spectral resolution can also lead to decreased radiant energy for each channel, making the imaging process susceptible to sensor depletion and environmental disturbance (Rasti et al., 2018). This often results in noise pollution and inaccurate information, which can negatively impact land cover mapping and surface parameter inversion (Yuan et al., 2012). Thus, developing robust restoration algorithms has become a cost-effective solution to improve the utilization of HSI data. Many studies have recognized the importance of developing such algorithms to mitigate noise pollution issues in HSIs (Rasti et al., 2018).

From an information utilization perspective, HSI denoising can be broadly divided into two classes (Yuan et al., 2012). The first class mainly relies on the spatial information of HSIs, considering every band as a grayscale image. It aims to remove noise by calculating the gray value of each pixel based on the correlation between pixels/patches in the original image (Li et al., 2010). Spatial filters, such as mean filtering (Gonzalez, 2009), median filtering (Pitas and Venetsanopoulos, 2013), Wiener filtering (Jain, 1981) et al. make use of low pass filtering on pixel groups with the statement that the noise occupies a higher region of the frequency spectrum. Normally, spatial filters eliminate noise to a reasonable extent with high efficiency but at the cost of image blurring, which in turn loses sharp edges. This approach employs various techniques such as Fourier transform (Wang et al., 2006), wavelet transform (Chen and Qian, 2010), hybrid Fourier-Wavelet filter (Pande-Chhetri and Abd-Elrahman, 2011), and weighted nuclear

---

* Corresponding author.
E-mail address: xiaolin.zhu@polyu.edu.hk (X. Zhu).

https://doi.org/10.1016/j.jag.2023.103481
Received 15 May 2023; Received in revised form 21 August 2023; Accepted 2 September 2023

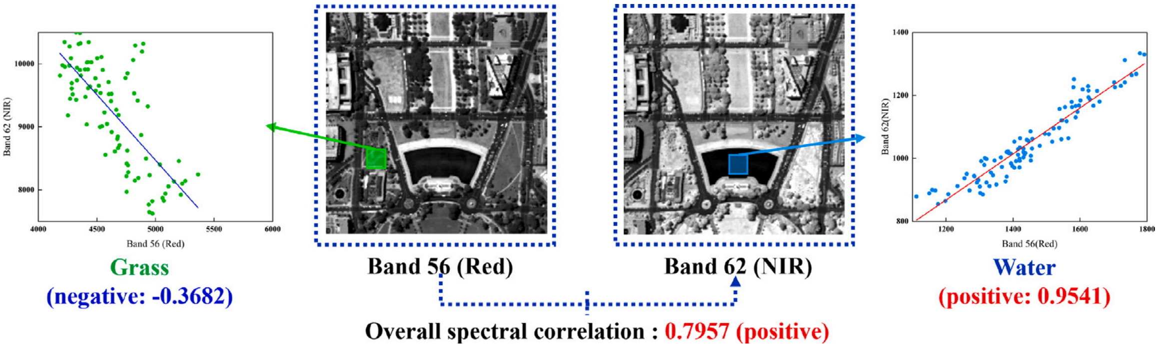


Fig. 1. Different land cover types have different spectral autocorrelation (Data example: Washington DCMall).

norm minimization (Gu et al., 2014) to handle HSIs band by band. Despite their simplicity and effectiveness, these methods overlook the spectral information present in the HSI.

Currently, the second category of HSI denoising methods that aim to simultaneously exploit both spatial and spectral correlation have become increasingly popular. For instance, Yuan et al. (2012) developed an adaptive total variation (TV) model that considered the dissimilarities of spatial structure and spectral noise during denoising. Zhang et al. (2013) explored the spatial and spectral low-rank property to suppress mixed noise in HSIs. Li et al. (2016) designed a hybrid spectral–spatial distributed sparse representation framework that fully exploited the spectral correlation across bands for HSI recovery. Chen et al. (2017) proposed the non-independent and identically mixture of Gaussians (NMoG) noise assumption to remove noise on HSIs in real scenarios. Given that HSI is a high-dimensional data, tensor-based methods such as CANDECOMP/PARAFAC (CP) decomposition (Zheng et al., 2020), Tucker decomposition (Zhang et al., 2019), and tensor-ring decomposition (Chen et al., 2019) are prevalent in HSI denoising task. He et al. (2019) integrated the spatial non-local similarity with the global spectral low-rank property of HSI with the non-local meets global (NGmeet) restoration paradigm to achieve promising restoration performance. The double-factor-regularized low-rank tensor factorization (LRTFDRF) designed by Zheng et al. (2020) can preserve spatial structural integrity and fully utilize spectral correlations of HSI.

Additionally, with the rapid advancement of deep learning, coupling spatial and spectral information processing through an end-to-end approach has become a popular method. Chang et al. (2018) used convolutional neural network (CNN) to extract the spatial information and the spectral correlation, it can adapt to various kinds of noise in HSIs with training on simulated datasets. Yuan et al. (2018) adopted a sliding window strategy and chose the residual learning mode to eliminate the noise in HSIs (HSID). Wei et al. (2020) embedded a quasi-recurrent pooling function into 3D CNN (QRNN3D) to obtain holistic features across the spectrum, thereby enhancing the spectral fidelity. Zhang et al. (2020b) established a deep spatial–spectral Bayesian posterior (DSSBP) to eliminate the non-independent and identically (i.i.d.) noise. Zhuang and Ng (2021) combined the deep image prior extracted from CNN and low rankness denoted by subspace representation to design a fast and parameterfree HSI mixed noise removal method (FastHyMix).

Although exploiting spectral correlation can boost the HSI denoising performance, it does not fully consider the differences in reflectance between different land cover components. For example, the spectral correlation between the red band and the near-infrared band is highly positive, as shown in Fig. 1. However, the correlation between grass and water is completely opposite. This can result in local distortion in areas with diverse land cover types. To validate this point, we tested six state-of-the-art HSI restoration algorithms (He et al., 2019; Wei et al., 2020; Zheng et al., 2020; Zhuang and Ng, 2021; Yuan et al., 2018; Chen et al., 2017), and the results are shown in Fig. 2. We selected three regions with different heterogeneities and found that the denoising error increases with the increase of surface heterogeneity. Therefore, to address this issue, we need to consider the differences in spectral correlation between different land cover types in a HSI, i.e., the spectral autocorrelation of the HSI. From this point of view, we propose a spatial–spectral interactive restoration (SSIR) framework, the main contributions can be concluded in three aspects:

1. We propose a data-driven and model-driven denoising framework to handle the challenge of surface heterogeneity. The spectral autocorrelation revealed by the unmixing is used as the prior for a deep learning-based denoiser, which can fully exploit the differences in reflectance characteristics of different land covers.
2. A deep learning-based denoiser combining CNN and Swin Transformer (TF) block is proposed, which enables local–global dependency modeling and content-based interaction to better characterize global heterogeneity differences in HSIs.
3. An unsupervised unmixing module is introduced in the denoising framework to describe the spectral autocorrelation of different land covers, which can enhance the generalization ability and achieve stable denoising performance under diverse scenarios.

2. Methodology

2.1. Problem formulation of HSI restoration

In this study we treat HSIs as 3D tensors, the used symbolisms and definitions can be referred to Zhuang and Ng (2021). Simply assuming the noise is additive, the degraded HSI observation (Yuan et al., 2012) can be described as:

$$\mathcal{Y} = \mathcal{X} + \mathcal{N}, \tag{1}$$

where $\mathcal{Y}$ is the degraded HSI, $\mathcal{X}$ is the clean HSI to be estimated, and $\mathcal{N}$ denotes the noise, which primarily includes Gaussian noise, stripe noise, and deadlines. The three variables share the same size $h \times w \times c$, representing the height, width and channels of HSI, respectively. Estimating $\mathcal{X}$ from $\mathcal{Y}$ is a typical ill-posed question, which requires regularization, also known as a prior, to constrain the solution space (Zoran and Weiss, 2011). Using a Bayesian approach, it can be estimated via solving a Maximum-a-Posteriori (MAP) estimation (Greig et al., 1989) problem by minimizing the energy function presented below:

$$E(\mathcal{X}) = \frac{1}{2}\|\mathcal{Y} - \mathcal{X}\|_F + \lambda \Phi(\mathcal{X}), \tag{2}$$

where the $\frac{1}{2}\|\mathcal{Y} - \mathcal{X}\|_F$ presents a data fidelity term, λ is the regularization parameter to make the contributions among different terms balanced, and $\Phi(\mathcal{X})$ denotes the prior term.

2.2. Spatial-spectral interactive restoration framework

When observing a heterogeneous landscape, HSIs may contain mixed pixels where multiple land cover types coexist within the

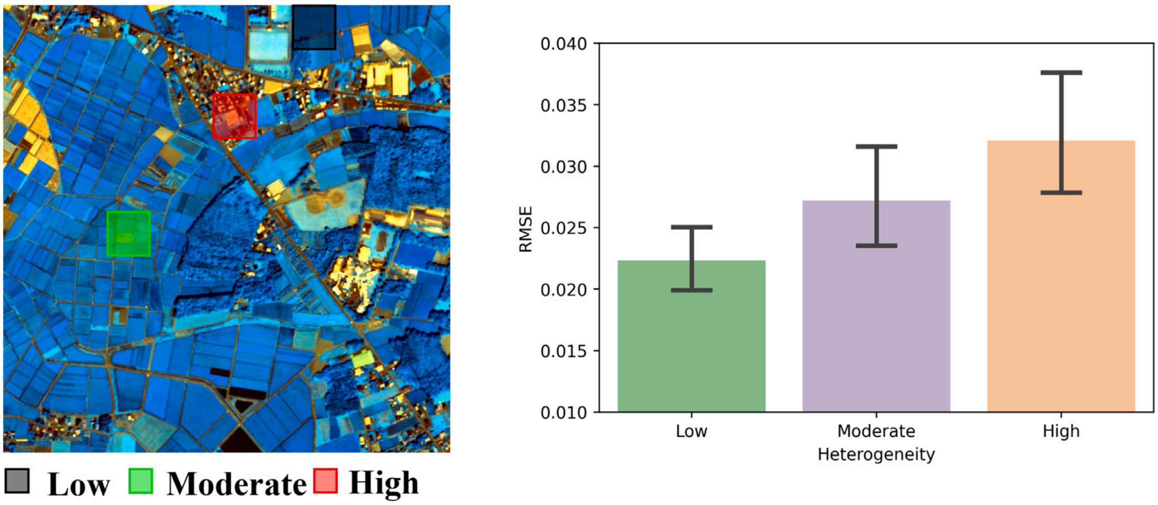


Fig. 2. Performance variation of six competing results to differences of surface heterogeneity.

instantaneous field of view of sensors. Therefore, unmixing is utilized to model the spectral autocorrelation of various land cover types. To ensure computational efficiency and account for the minor contribution of multiple scattering radiation between endmember types to the mixed pixel spectrum (Chen et al., 2011; Wang et al., 2021), a linear spectral mixing model (LMM) is applied in this study:

$$\mathcal{X} = \mathcal{A} \times_3 \mathbf{S}. \tag{3}$$

where the product of the matrix of endmember signatures $\mathbf{S} \in \mathbb{R}^{c \times r}$ and the abundance map $\mathcal{A} \in \mathbb{R}^{h \times w \times r}$ (r denotes the number of endmembers) is the clean HSI $\mathcal{X}$. Given the physical nature of endmembers and abundance fractions, the elements of $\mathbf{S}$ and $\mathcal{A}$ should be greater than zero and the abundance fractions of all endmembers within a mixed pixel should sum to one. Thus, the denoising model with the spectral autocorrelation prior is:

$$\begin{aligned} &E(\mathcal{X}, \mathcal{A}, \mathbf{S}) = \tfrac{1}{2}\|\mathcal{Y} - \mathcal{X}\|_F + \lambda \Phi(\mathcal{X}) \\ &s.t. \mathcal{X} = \mathcal{A} \times_3 \mathbf{S}, \mathbf{S} > 0, \forall i \in [1, h \times w], \sum_{j=1}^{r} \mathbf{a}_{i,j} = 1. \end{aligned} \tag{4}$$

where $\mathbf{a}_{i,j}$ denotes the fractional abundance vector of the jth endmember at the ith pixel. After integrating all the constraints into the main problem, we get:

$$\begin{aligned} E(\mathcal{X}, \mathcal{A}, \mathbf{S}) = &\frac{1}{2}\|\mathcal{Y} - \mathcal{X}\|_F + \lambda_1 \Phi_1(\mathcal{X}) + \frac{\mu}{2}\|\mathcal{X} - \mathcal{A} \times_3 \mathbf{S}\|_F \\ &+ \lambda_2 \Phi_2(\mathbf{S}) + \lambda_3 \Phi_3(\mathcal{A}), \end{aligned} \tag{5}$$

where μ is the penalty parameter. The unfolding inference of Eq. (5) is derived by the half quadratic splitting (HQS) algorithm (He et al., 2013). In this way, three subproblems can be decoupled to solve $\mathcal{X}$, $\mathbf{S}$ and $\mathcal{A}$. Since the three variables are dependent as demonstrated in Eq. (3), suggesting that once we obtain two of them, the other one can be inferred. In this study, solving the $\mathcal{X}$ that represents the clean HSI is the first problem to be addressed; and solving the $\mathbf{S}$-subproblem which denotes the spectral autocorrelation is the second problem. Consequently, Eq. (5) can be decoupled into:

$$\begin{cases} \hat{\mathbf{S}} = \arg\min\limits_{\mathbf{S}} \frac{\mu}{2}\|\mathcal{X} - \mathcal{A} \times_3 \mathbf{S}\|_F + \lambda_2 \Phi_2(\mathbf{S}). & \text{(a)} \\ \hat{\mathcal{X}} = \arg\min\limits_{\mathcal{X}} \frac{1}{2}\|\mathcal{Y} - \mathcal{X}\|_F + \lambda_1 \Phi_1(\mathcal{X}) + \frac{\mu}{2}\|\mathcal{X} - \mathcal{A} \times_3 \mathbf{S}\|_F, & \text{(b)} \end{cases} \tag{6}$$

From the Bayesian perspective, Eq. (6)(a) corresponds to an unmixing problem of HSI that characterizes the spectral autocorrelation, while Eq. (6)(b) can be regarded as a denoising problem with the constraint of $\mathbf{S}$. The clean HSI can be estimated by iteratively solving Eq. (6)(a) and (b), which can be conducted by the unfolding optimization (Zhang et al., 2020a). Based on the proposed SSIR framework, we design an unfolding HSI-denoising network with a spectral autocorrelation prior, which consists of an unmixing module U and a denoising module D alternately:

$$\begin{cases} \hat{\mathbf{S}}_k = \arg\max\limits_{\mathbf{S}} \log \mathrm{P}\left(\mathcal{X}_k | \mathbf{S}_k\right) + \log \mathrm{P}\left(\mathbf{S}_{k-1}\right), & \text{(a)} \\ \hat{\mathcal{X}}_k = \arg\max\limits_{\mathcal{X}} \log \mathrm{P}\left(\mathcal{Y} | \mathcal{X}_k\right) + \log \mathrm{P}\left(\mathbf{S}_k | \mathcal{X}_k\right) + \log \mathrm{P}\left(\mathcal{X}_k\right), & \text{(b)} \end{cases} \tag{7}$$

where $\mathrm{P}\left(\mathbf{S}_{k-1}\right)$ and $\mathrm{P}\left(\mathcal{X}_k | \mathbf{S}_k\right)$ are the prior probability and likelihood of $\mathbf{S}$-problem, while $\mathrm{P}\left(\mathcal{X}_k\right)$ and $\mathrm{P}\left(\mathbf{S}_k | \mathcal{X}_k\right)$ are the prior term and likelihood of $\mathcal{X}_k$-problem, respectively. Given that unmixing the noisy HSI will produce visible errors, and the obtained endmembers might be inaccurate as prior knowledge for the next estimation of $\mathcal{X}$. As a result, The initial stage is to estimate the preliminary noise-reduced HSI using the Maximum Likelihood Estimation (MLE):

$$\hat{\mathcal{X}}_0 = \arg\max_{\mathcal{X}} \log \mathrm{P}\left(\mathcal{Y} | \mathcal{X}_0\right) + \log \mathrm{P}\left(\mathbf{S}_0 | \mathcal{X}_0\right). \tag{8}$$

$\hat{\mathcal{X}}_0$ is the input of the first unmixing module to get $\hat{\mathbf{S}}_1$. After getting the posterior term $\mathrm{P}\left(\mathcal{X}_1 | \mathbf{S}_1\right)$ through the first unmixing module, it can be used as the prior information of the next denoiser, i.e. $\mathrm{P}\left(\mathcal{X}_1 | \mathbf{S}_1\right) = \mathrm{P}\left(\mathcal{X}_1\right)$ since the $\hat{\mathbf{S}}_1$ is solved. Similarly, the posterior $\mathrm{P}\left(\mathbf{S}_1 | \mathcal{X}_1\right)$ can act as the prior of the second endmember estimator. In other words, Eq. (7)(a) and (b) can be expressed as:

$$\begin{cases} \hat{\mathbf{S}}_k = \arg\max\limits_{\mathbf{S}} \log \mathrm{P}\left(\mathcal{X}_k | \mathbf{S}_k\right) + \log \mathrm{P}\left(\mathbf{S}_{k-1} | \mathcal{X}_{k-1}\right), & \text{(a)} \\ \hat{\mathcal{X}}_k = \arg\max\limits_{\mathcal{X}} \log \mathrm{P}\left(\mathcal{Y} | \mathcal{X}_k\right) + \log \mathrm{P}\left(\mathbf{S}_k | \mathcal{X}_k\right) + \log \mathrm{P}\left(\mathcal{X}_k | \mathbf{S}_k\right). & \text{(b)} \end{cases} \tag{9}$$

The overall flowchart of SSIR is shown in Fig. 3.

2.3. Unmixing module

To obtain the spectral autocorrelation prior, we utilize an unmixing module to estimate the endmember signatures representing the spectral correlations of different land cover components:

$$\hat{\mathbf{S}}_k = U_k\left(\hat{\mathcal{X}}_k, r_k\right). \tag{10}$$

where r_k is the number of endmembers set in the kth unmixing module. The unsupervised vertex component analysis (VCA) is chosen to implement the unmixing module for achieving robust results on HSIs covering different land covers (Nascimento and Dias, 2005). VCA considers the feature space spanned by spectral characteristics of all HSI pixels as a geometric simplex of which the vertices represent endmember signatures, and then projects the HSI to a given number of random directions in feature space and takes the pixels that become

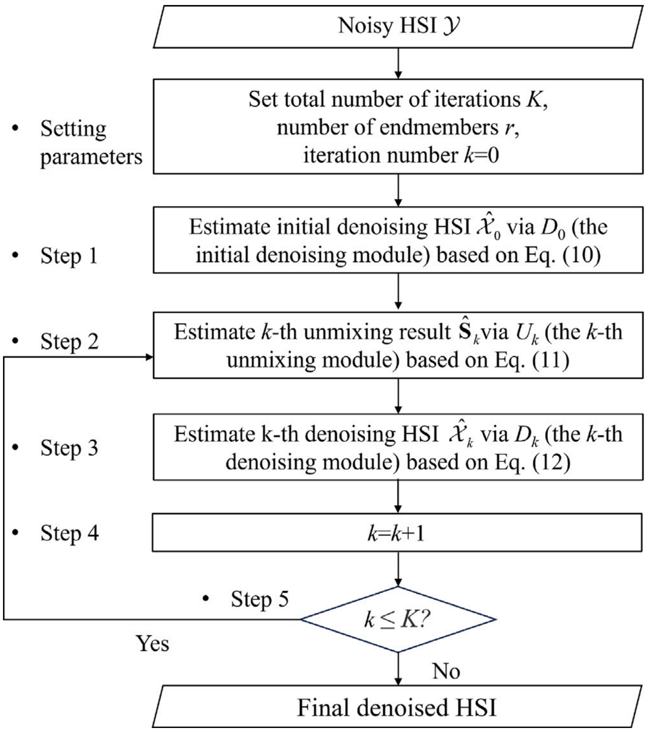


Fig. 3. The flowchart of the proposed spatial–spectral interactive restoration loop for HSI denoising.

endpoints in these projected directions as endmembers. As no trainable parameters are contained in Eq. (10), the generalizability of the model is improved by fully decoupling between the denoising module and the unmixing module.

2.4. Denoising module

The denoising module aims to obtain a noise-reduced HSI. It is well-known that the unmixing module may face challenges in obtaining reliable spectral signatures when the noise level is high, which can lead to slower convergence and unsatisfactory performance. To overcome this issue, the initial denoising module is designed to remove noise based solely on learning the noise distribution characteristics of training data, without any prior information. The input of the initial denoising module is the noisy observation:

$$\hat{\mathcal{X}}_0 = D_0(\mathcal{Y}). \tag{11}$$

Through the noise-reduced $\hat{\mathcal{X}}_0$, the preliminary endmember signatures can be obtained. These signatures can serve as a prior for subsequent denoisers, and can be expressed as follows:

$$\hat{\mathcal{X}}_k = D_k\left(\hat{\mathcal{X}}_{k-1} \oplus \hat{\mathbf{S}}_k\right), \tag{12}$$

where $\hat{\mathcal{X}}_{k-1} \oplus \hat{\mathbf{S}}_k$ denotes the coupling operation of the spatial characteristics and the spectral autocorrelation of the HSI, by which the input HSI is projected into a high-dimensional feature space:

$$\hat{\mathcal{X}}_{k-1} \oplus \hat{\mathbf{S}}_k = \left[\hat{\mathcal{X}}_{k-1}, \hat{\mathcal{X}}_{k-1} \times_3 \hat{\mathbf{S}}_k\right], \tag{13}$$

where [... , ...] denotes the concatenation operation. The proposed denoiser effectively combines the spatial information and spectral autocorrelation of HSIs, resulting in a dynamic fusion of the two. As such, we refer to the denoising module as the spatial–spectral interactive restoration block (SSIRB).

SSIRB consists of four convolutional layers and six Swin Transformer (TF) layers (each layer consists of two successive Swin TF blocks) (Liu et al., 2021) as shown in Fig. 4. The first convolutional layer is crafted to capture shallow features from the noisy HSI:

$$f_{sf} = h_{conv}\left(\hat{\mathcal{X}}_{k-1} \oplus \hat{\mathbf{S}}_k\right), \tag{14}$$

where f_{sf} is the extracted shallow features, h_{conv} is the first 3 × 3 convolutional layer. Next, the extracted shallow features are sent to six consecutive Swin TF layers, resulting in the intermediate features of HSI:

$$f_{i+1} = h_{swin}\left(f_i\right)(i = 1, 2, \ldots, L-1), \tag{15}$$

where f_i $(i = 1, 2, \ldots, L-1)$ is the intermediate features obtained by the Swin TF layer h_{swin}. Notably, we set L to 6 in consideration of a trade-off between the efficiency and accuracy. The Swin TF is a window-based self-attention, allowing higher efficiency and cross-window connections. It can be viewed as a spatially varying convolution with long-range dependency modeling capability. A given pixel's response can be estimated by weightily summing values of all surrounding positions based on the dynamic attention mechanism of Swin TF layer, thereby flexibly interacting with the input. Following that, the intermediate feature acquired by the last Swin TF layer is conveyed to another convolution layer to form the deep feature f_{df}:

$$f_{df} = h_{conv}\left(f_6\right). \tag{16}$$

Next, both shallow and deep features are aggregated by the residual connection and fused in the following convolution layer:

$$f_{fusion} = h_{conv}\left(f_{sf} + f_{df}\right), \tag{17}$$

where f_{fusion} represents the aggregated features for reconstructing the residual image f_{rec} via a skip connection and a convolution layer. The obtained residual image is added to the input HSI through another residual connection to produce the final output:

$$\hat{\mathcal{X}}_k = f_{rec} + \hat{\mathcal{X}}_{k-1}. \tag{18}$$

Multi-level residual connections are used in our approach. These connections, in which skip connections are added following a convolutional operation, aim to enforce the sharing of redundant information among network layers, facilitate gradient flow, and enhance the optimization capability of the model (Zhang et al., 2017).

2.5. End-to-end training

The end-to-end learning strategy is adopted to train the proposed SSIR for the sake of performance and the training objective is to acquire the trainable parameters of SSIR by minimizing a loss function on a large training dataset:

$$Loss = \sum_{k=0}^{K} \left\|\hat{\mathcal{X}}_k - \mathcal{X}\right\|_1, \tag{19}$$

K the defined maximum number of iterations. All stages are simultaneously supervised by commonly-used $L1$ loss during training, not only to give full play to the VCA as it performs robustly on the clean HSI, but also to ease the difficulty of training and find better network parameters (Kim et al., 2016).

During the training process, the batch size is configured to 16. For the gradient descent in the backpropagation, the ADAM optimizer is selected for its efficiency. The base learning rate is 0.002, and decayed at epochs. SSIR is viewed as convergence when the validation performance cannot be increased to surpass the setting threshold anymore. The sliding window inference strategy in the spectral dimension is applied to handle HSIs with any number of spectra. The K denoisers are trained together with different parameters. Although parameter sharing can reduce the number of parameters, it would increase the training difficulty. Specifically, in the case that the number of input bands of the network is L, if the number of test image bands is larger than L, the restoration is performed by sliding over the spectral dimension with the stride of L; if it is less than L, the replicate padding of spectral bands is conducted before inputting to the network. In this study, L is set to 32.

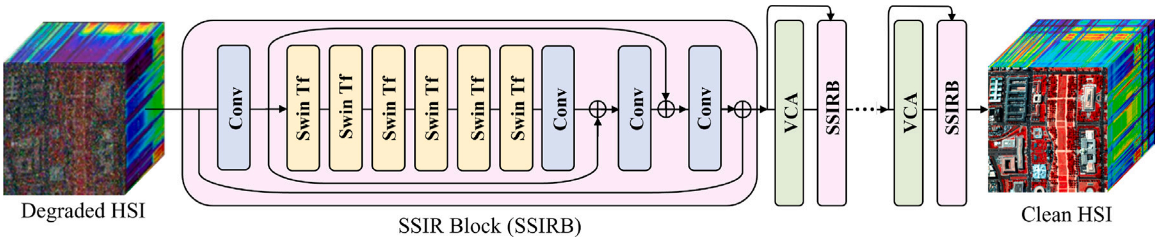
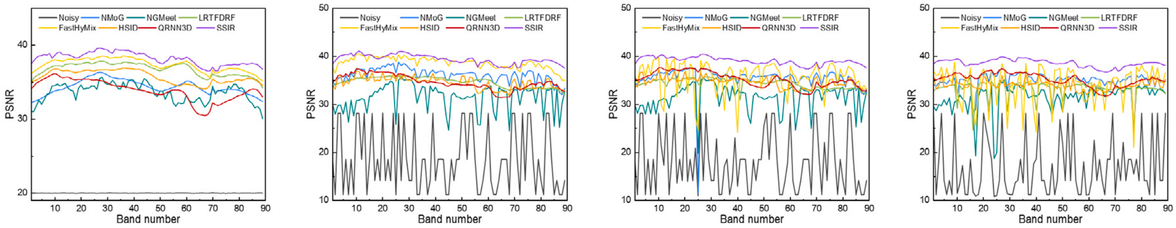


Fig. 4. The detailed architecture of SSIR.



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

Fig. 5. PSNRs of Pavia University data.

3. Experiments

The proposed SSIR was tested on both synthetic and real datasets. Six advanced spatial–spectral joint denoising algorithms were selected for comparison, including three model-based methods, namely NMoG, NGMeet, LRTFDRF. We also included two deep learning network HSID and QRNN3D, and a hybrid method FastHyMix. All relevant parameters were optimized and trained using the same dataset with the proposed framework according to the instructions from the literature.

Training samples are generated from the high-quality airborne Pavia Center dataset (Gege et al., 2009) in Pavia, Italy. Its spatial resolution is 1.3 meters and has 102 spectral channels. After eliminating bands without information, multiple overlapped volumes are cropped to enlarge the training set. Samples have a spatial scale of 64 × 64 and 32 spectral bands to maintain spectral continuity. Data augmentation techniques, such as flipping and rotation, are applied to increase the training samples.

All the experiments were conducted on MATLAB R2021a, Pytorch and a Windows 10 desktop equipped with an Intel i7-12700K CPU, 64 GB RAM and NVIDIA GTX 3090 GPU.

3.1. Simulation experiments

3.1.1. Experimental configurations

All the datasets are normalized to [0, 1]. The validation set, a different scene of Pavia, is the Pavia University data, covering 430 to 860 nm with a size of 340 × 340 × 89. Headwall Hyperspec-VNIR-C imaging system captured Chikusei airborne HSI datasets in agricultural and urban regions in Chikusei, Japan, is used as the synthetic test set, with 128 bands from 363 nm to 1018 nm. A 512 × 512 × 128 subcube was selected as the testing data. The Chikusei data covers rural farmland, while the Pavia University data covers urban areas.

Four types of degradation cases were added to quantitatively assess the effectiveness of SSIR:

- Case 1: I.i.d. Gaussian noise: All the bands were contaminated by zero-mean Gaussian noise with identical intensity 0.1.
- Case 2: Non-i.i.d. Gaussian noise: All the bands were affected by zero-mean Gaussian noise with different intensities varying from 0.03 to 0.3.
- Case 3: Non-i.i.d. Gaussian noise + stripe noise: The original HSI was degraded by non-i.i.d. Gaussian noise as Case 2, and 20% bands in the original HSI were randomly picked to add stripes with widths from [1,5]. The percentage of stripes on these bands is ranging from 10% to 50%.
- Case 4: Mixed noise: The original HSI was contaminated by non-i.i.d. Gaussian noise as Case 2. The percentage of stripes on these bands is ranging from 5% to 20%. Also, 10% of bands in the original HSI were randomly selected to add deadlines with widths from [1,5]. The percentage of stripes on these bands is ranging from 5% to 20%.

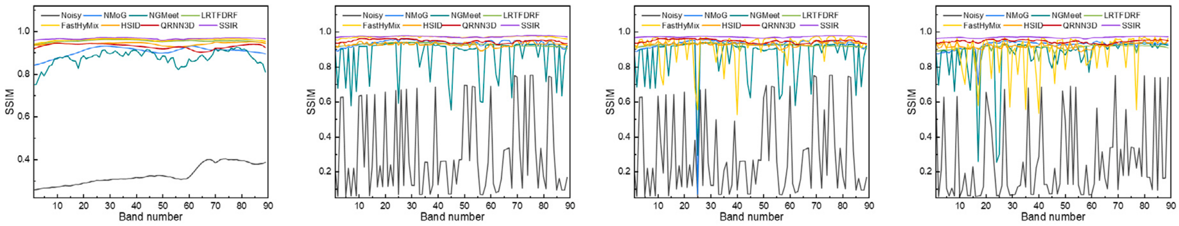
To increase the adaptability of our model to different types and levels of noise, the first model (model 1) was trained on non-i.i.d. Gaussian noise samples, then it was tuned via the training set with mixed noise to get model 2. Model 1 was tested on the data of Case 1 and 2, while Model 2 for mixed noise was tested on the data of Case 3 and 4.

3.1.2. Evaluations on synthetic datasets

Three quantitative indicators were used to assess the denoising effectiveness: mean peak signal-to-noise ratio (MPSNR), mean structural similarity index (MSSIM), and mean spectral angle (MSA). MPSNR and MSSIM are positive indices, with larger values indicating greater spatial structural recovery ability. Conversely, MSA is a negative indicator, with lower values indicating less spectral distortion.

Experiments on the pavia university data. Quantitative indicators of Pavia University data are shown in Table 1, optimal results are marked in bold, followed by the underlined ones. Table 1 indicates SSIR has superior MPSNR, MSSIM, and MSA values compared to other methods. Figs. 5 and 6 show that SSIR acquired higher PSNRs and SSIMs across all the bands compared with other approaches, proving its robustness to different intensities and patterns of corruption.

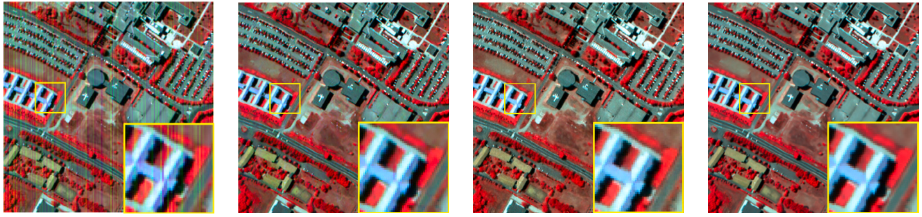
Fig. 7 displays the false color composite results under mixed noise conditions. Traditional models, such as LRTFDRF, exhibit effective noise suppression and yield favorable quantitative metrics. However, they have longer computational times compared to learning-based methods. The hybrid approach, FastHyMix, removes noise in the shortest time, but struggles to eliminate stripes. Regarding data-driven approaches, HSID and QRNN3D generate satisfactory results, while the proposed SSIR demonstrates improved performance in preserving local details.



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

Fig. 6. SSIMs of Pavia University data.

(a) Original (b) Degraded (c) NMoG (d) NGMeet (e) LRTFDRF



(f) FastHyMix (g) HSID (h) QRNN3D (i) SSIR

Fig. 7. Visual results of Pavia University data in Case 4 (R: Band 80, G: Band 32, B: Band 21).

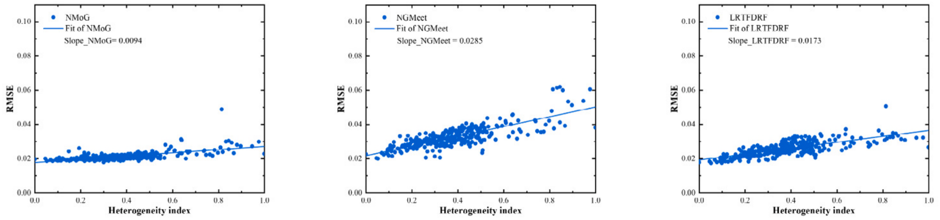
Table 1
Quantitative indicators on Pavia University data.

Index	Noisy	NMoG	NGMeet	LRTFDFR	FastHyMix	HSID	QRNN3D	SSIR
Case 1								
MPSNR	20.0010	34.3385	33.6020	36.7482	37.3715	35.7309	33.9159	**38.2059**
MSSIM	0.3263	0.9050	0.8750	0.9540	0.9597	0.9416	0.9294	**0.9660**
MSA	28.8359	5.0148	6.4667	3.2161	3.2893	3.7835	4.7494	**2.8916**
Time/s	–	133.19	94.79	133.51	0.59	**0.45**	2.73	4.13
Case 2								
MPSNR	18.4955	36.1176	31.4679	34.6636	38.6934	34.5308	34.6496	**39.4084**
MSSIM	0.3087	0.9365	0.8658	0.9270	0.9707	0.9279	0.9355	**0.9988**
MSA	42.4371	4.2909	3.9694	3.0937	8.0392	4.6960	4.7991	**2.5982**
Time/s	–	74.83	39.14	121.89	0.59	**0.31**	1.31	4.22
Case 3								
MPSNR	18.0589	35.4422	31.4818	34.2970	35.8007	34.6857	35.2206	**39.0730**
MSSIM	0.2934	0.9229	0.8673	0.9239	0.9216	0.9303	0.9471	**0.9717**
MSA	42.7760	9.3518	8.3067	4.3607	5.9809	4.6498	4.1529	**2.6777**
Time/s	–	59.87	67.31	115.49	0.52	**0.33**	1.30	4.26
Case 4								
MPSNR	16.8384	34.9453	31.4389	33.2185	34.0531	34.0476	35.1971	**38.5747**
MSSIM	0.2567	0.9240	0.8687	0.9088	0.8885	0.9208	0.9448	**0.9684**
MSA	45.4986	5.2081	9.2651	5.8052	7.6038	4.9904	4.1642	**2.7835**
Time/s	–	66.35	66.35	130.45	0.51	**0.37**	1.32	4.28

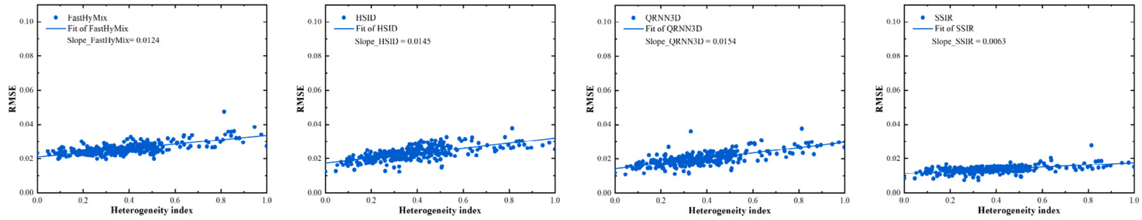
Previous research has not investigated the impact of heterogeneity variations on overall recovery outcomes. To address this knowledge gap, we introduce a regional heterogeneity index based on the ratio of local to global standard deviation. Then we divide the 340 × 340 Pavia University data into 400 non-overlapping 17 × 17 sub-blocks and calculate the heterogeneity value of each sub-block:

$$HI_{cube} = normalization\left(\frac{std_{cube}}{std}, [0,1]\right), \tag{20}$$

where std_{cube} and std are the standard deviations of the local image cube and the entire image, respectively. Then HI_{cube} values of all the image cubes are rescaled with Min-Max normalization. The denoising performance is measured by the root mean square error (RMSE) between the denoised cube and the ground truth. In Fig. 8, each dot represents a sub-block, and the fit line obtained by linearly fitting the dots illustrates the sensitivity of the denoising method to surface heterogeneity. The RMSE values tend to increase as the regional heterogeneity increases, and the slope and intercept of the linear fitting

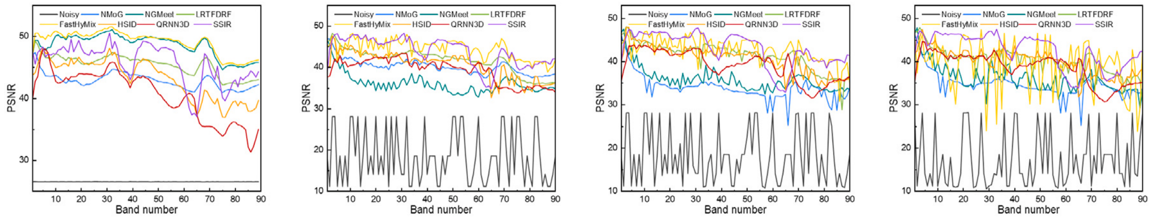


(a) NMoG (b) NGMeet (c) LRTFDRF



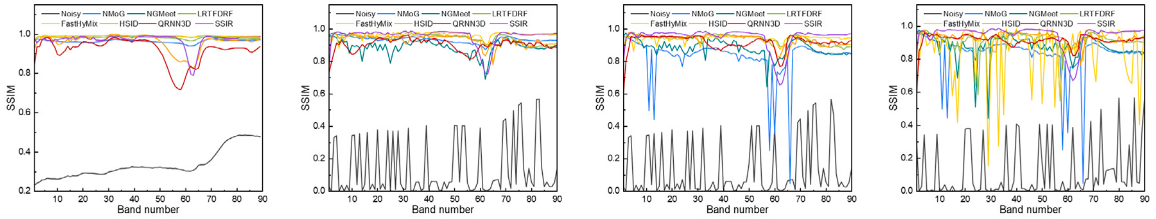
(d) FastHyMix (e) HSID (f) QRNN3D (g) SSIR

Fig. 8. Restoration performance of different local heterogeneity: Pavia University data (Case 4).



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

Fig. 9. PSNRs of Chikusei data.



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

Fig. 10. SSIMs of Chikusei data.

line indicate the sensitivity of the denoising method. NGMeet, LRTFDRF, FastHyMix, HSID, and QRNN3D can achieve good restoration results in homogeneous areas, but their performance suffers as the local heterogeneity increases. NMoG and SSIR exhibit the least susceptibility to changes in the heterogeneity index, with SSIR demonstrating the lowest RMSE values.

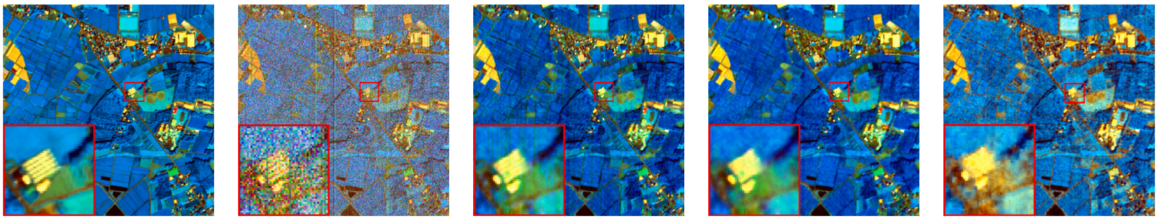
Experiments on the chikusei data. The quantitative results of Chikusei data in four cases are listed in Table 2, and the PSNR values and SSIMs of different bands are displayed in Figs. 9 and 10. Without any fine-tuning, SSIR obtained sub-optimal performance in Gaussian noise removal (Case 1 and 2), but it still outperforms other methods when noise condition is more complex.

Fig. 11 shows the qualitative false-color results of Case 4. It illustrates that FastHyMix and HSID are not effective in removing stripes in cases of complex degradation. Spectral distortions can be observed through the results of NMoG, NGMeet, and QRNN3D. The performance of HSID and QRNN3D on Chikusei data is inferior to that on Pavia University data, which can be attributed to their dependence on training samples, a common limitation of learning-based models. Compared to SSIR, all comparison methods cannot reconstruct spatial details satisfactorily, as demonstrated by the enlarged boxes in Fig. 11. With model-based unmixing module, SSIR can better suppress complex noise while preserving both local structural features and overall image contrast.

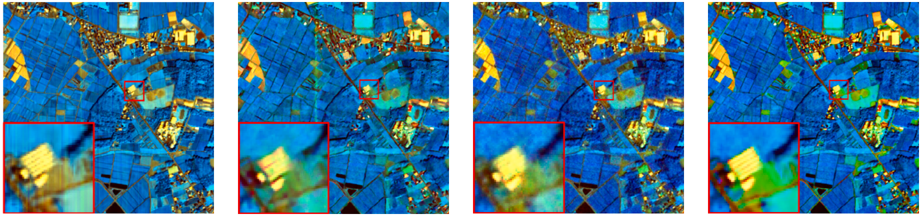
Likewise, we divide the 512 × 512 Chikusei data into 256 non-overlapping 32 × 32 sub-blocks, and compute the heterogeneity value of each sub-block using Eq. (20). The relationship between the RMSE values and local heterogeneity is depicted in Fig. 12. The stability of NMoG is greatly reduced for Chikusei data compared to Pavia University data, implying that the performance of NMoG is still affected by the content of the HSI, i.e., the geographical information present in the image. In this series of experiments, SSIR outperforms other methods with the lowest slope and intercept and produces reliable results in heterogeneous landscapes.

Table 2
Quantitative indicators on Chikusei data.

Index	Noisy	NMoG	NGMeet	LRTFDRF	FastHyMix	HSID	QRNN3D	SSIR
Case 1								
MPSNR	26.5680	42.4664	44.4147	44.5912	**44.8734**	44.3304	39.9263	44.7345
MSSIM	0.3699	0.9652	0.9767	**0.9787**	0.9780	0.9675	0.9314	0.9736
MSA	24.1990	3.2924	2.5399	2.7332	**2.4748**	4.2239	6.0446	2.7331
Time/s	–	191.76	155.60	253.64	**1.56**	2.93	6.78	15.62
Case 2								
MPSNR	18.0391	39.6889	35.9795	40.3633	**43.3415**	38.8454	38.2170	42.8172
MSSIM	0.1440	0.9345	0.8897	0.9318	**0.9638**	0.9226	0.9064	0.9579
MSA	58.8000	4.1644	6.2480	3.3903	**2.8528**	6.3320	6.5328	3.7339
Time/s	–	112.08	247.50	212.71	**1.41**	2.17	5.89	15.12
Case 3								
MPSNR	17.7649	33.7939	35.5781	39.4813	39.1742	39.0805	38.2475	**41.9260**
MSSIM	0.1356	0.8361	0.8793	0.9114	0.8660	0.9242	0.9195	**0.9506**
MSA	59.0860	8.2362	6.1979	**4.3545**	8.9999	5.7291	5.7288	4.3989
Time/s	–	192.66	295.40	319.82	**1.16**	1.97	5.46	16.25
Case 4								
MPSNR	17.4721	34.4739	35.7048	39.2325	38.8929	38.8438	37.9175	**41.8244**
MSSIM	0.1354	0.8385	0.8716	0.9146	0.8678	0.9236	0.9195	**0.9524**
MSA	59.5142	7.4402	6.5502	5.0186	9.0067	5.8500	5.8559	**4.4789**
Time/s	–	182.78	157.08	372.79	**1.23**	2.35	5.72	15.25

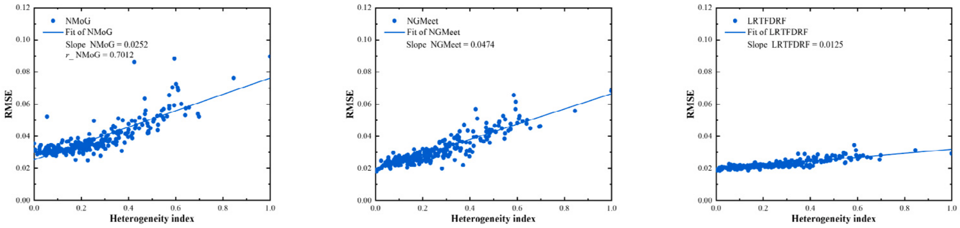


(a) Original (b) Degraded (c) NMoG (d) NGMeet (e) LRTFDRF

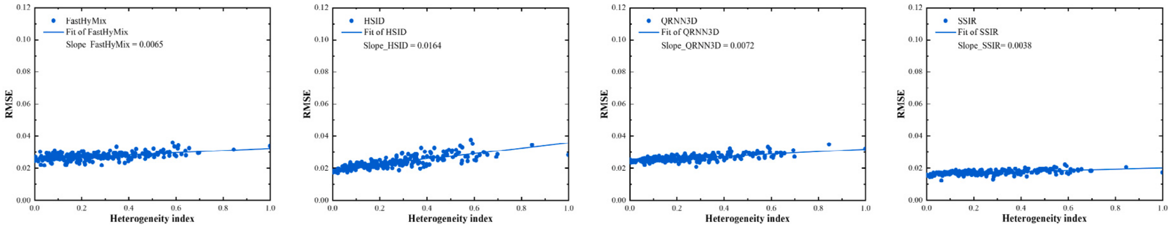


(f) FastHyMix (g) HSID (h) QRNN3D (i) SSIR

Fig. 11. Visual results of Chikusei data in Case 4(R: Band 49, G: Band 69, B: Band 90).



(a) NMoG (b) NGMeet (c) LRTFDRF



(d) FastHyMix (e) HSID (f) QRNN3D (g) SSIR

Fig. 12. Restoration performance of different local heterogeneity: Chikusei dataset (Case 4).

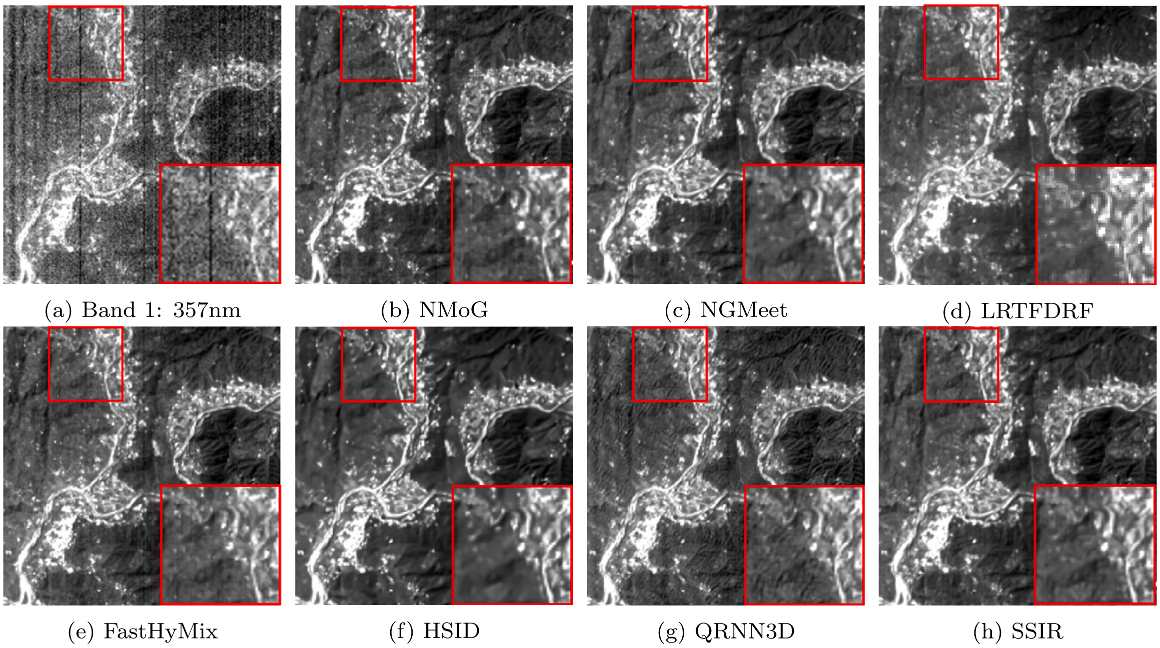

(a) Band 1: 357nm
(b) NMoG
(c) NGMeet
(d) LRTFDRF
(e) FastHyMix
(f) HSID
(g) QRNN3D
(h) SSIR

Fig. 13. Visual results of EO-1 Hyperion, locations of diagonal pixels: (30,1)-(79,50).

Table 3
Real degraded HSI datasets used for experiments.

Data	Selected image size	Spectral range
EO-1 Hyperion	200 × 200 × 166	357–2567 nm
Indian Pines	145 × 145 × 206	400–2500 nm
HYDICE Urban	307 × 307 × 188	400–2500 nm
OHS-1	400 × 400 × 32	400–1000 nm
Gaofen-5	400 × 400 × 330	390–2500 nm

3.2. Real experiments

3.2.1. Experimental configurations

To further demonstrate the effectiveness of SSIR in real-world situations, we selected five commonly used real HSI datasets (EO-1 Hyperion, Indian Pines, HYDICE Urban, OHS-1, and Gaofen-5) with mixed degraded conditions for comparison purposes. Test data can be found at https://qzhang95.github.io/#Resources (HSI Denoising Dataset V2.0). To minimize experimental error, bands contaminated by cirrus and water absorption were manually removed. A brief introduction of these datasets is presented in Table 3. Notably, for the real experiments, SSIR and the other two deep-learning methods utilized the same models as those used in the simulated experiment, which were trained on mixed-noise data.

Due to the lack of ground for real experiments, two non-reference evaluation indices, the mean Qmetric (Zhu and Milanfar, 2010) and the mean inverse coefficient of variation (MICV) (Nichol and Vohora, 2004) can serve as the performance assessment of denosing algorithms. The mean Q-metric was calculated on the global image, while MICV was computed on five randomly selected 10 × 10 regions on each test image. Both indicators demonstrate a positive correlation with image quality and are indicative of lower noise.

3.2.2. Evaluations on real datasets

The blind indices of five real HSI datasets are listed in Table 4, and Figs. 13–17 show the visual results. In terms of indicators and visual results, SSIR can obtain the highest values in most cases. By contrast, the performance of the other two deep learning methods has slightly declined compared with that in the simulated experiment. This advantage may be attributed to the integration of model-based and data-driven approaches. While NMoG, NGMeet, and LRTFDRF were effective at reducing noise, NGMeet and LRTFRDRF produced better visual results than NMoG. NMoG led to noticeable spectral distortion in the Indian Pines data. FastHyMix shows promise, but there are still stripe residuals on the results of EO-1, Indian Pines and HYDICE Urban data. In heavily contaminated datasets like EO-1 and Indian Pines, HSID produced overly-smooth results, and QRNN3D generated artifacts. The OHS-1 data shows low contrast between the noise and image, which makes it challenging for NMoG to effectively identify and remove noise. Similar low contrast is also observed in Gaofen-5 data, while all the methods are able to suppress the noise to a certain degree, only the proposed SSIR method can reconstruct clear spatial details of the solar panel, as shown in the enlarged window of Fig. 17. This suggests that SSIR can better balance the spatial and spectral information, resulting in improved restoration performance.

Fig. 18 shows the mean vertical profiles, which reflect the gray level of the band. The profiles before and after denoising should have the same trend but different smooth degree. Fig. 18 illustrates that the conventional methods and SSIR maintain the original gray level, indicating good spectral fidelity. HSID and QRNN3D, while successful in eliminating noise, cannot maintain mean vertical profiles due to limited generalization capabilities. Comparatively, SSIR shows great robustness owing to the combination of model-based and data-driven methods.

3.2.3. Further evaluation via classification

To assess whether denoising models could improve the accuracy of HSI applications, classification was conducted. The Indian Pines test site was selected for classification since the ground truth was available via in-situ measurement by the AVIRIS team. Classification was performed by support vector machine (SVM) and the Kappa coefficient and over accuracy (OA) were quantitative benchmarks. Ten feature classes were selected, and the SVM process was repeated ten times with 10% of samples in each class randomly chosen as training samples. The classification excluded 16 bands absorbed by the atmosphere and water.

Table 5 presents the quantitative indicators, and Fig. 19 shows the classification maps. They demonstrate denoising algorithms can improve classification accuracy, and SSIR achieved the best performance among all methods. Fig. 19(b) reveals the existence of noise causes

Table 4
Quantitative indicators on real datasets.

Index	Original	NMoG	NGMeet	LRTFDRF	FastHyMix	HSID	QRNN3D	SSIR
EO-1 Hyperion								
MQmetric	0.1529	0.1555	0.1571	0.1560	0.1549	0.1526	0.1556	**0.1627**
MICV	3.6822	3.8718	4.0261	3.9742	4.0044	4.1315	4.8481	**4.9356**
Time/s	–	91.08	66.75	197.54	**0.19**	0.37	1.48	3.98
Indian Pines								
MQmetric	0.0243	0.0255	0.0252	0.0210	**0.0257**	0.0247	0.0190	0.0247
MICV	38.8757	**53.2514**	51.5039	41.2938	44.7701	49.9731	37.7116	44.4573
Time/s	–	53.92	37.89	141.66	**0.11**	0.59	0.84	2.53
HYDICE Urban								
MQmetric	0.0960	0.0968	0.0986	0.0982	0.0967	0.0968	0.0943	**0.1032**
MICV	2.2278	2.2621	2.2859	2.7990	2.2782	2.5134	**2.9211**	2.8239
Time/s	–	255.27	187.46	403.48	**0.63**	1.19	2.41	7.54
OHS-1								
MQmetric	0.0455	0.0527	**0.0601**	0.0511	0.0531	0.0513	0.0489	0.0515
MICV	11.7790	13.7520	14.5193	13.9992	13.6290	13.7165	**15.6448**	14.7449
Time/s	–	165.25	226.49	299.47	0.54	**0.26**	0.54	0.86
Gaofen-5								
MQmetric	0.0241	0.0248	**0.0250**	0.0241	0.0239	0.0193	0.0191	0.0232
MICV	142.1228	224.7845	252.8445	253.1413	261.9608	262.8368	201.0801	**267.2695**
Time/s	–	941.84	489.21	1365.20	**4.69**	6.27	13.71	28.63

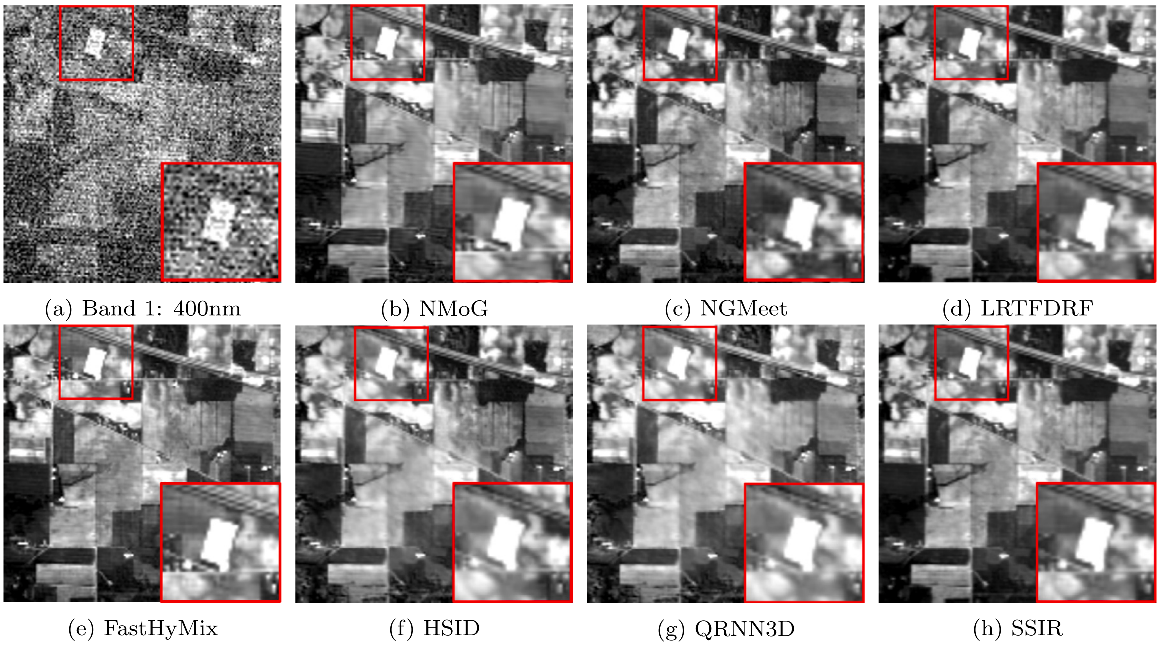

(a) Band 1: 400nm (b) NMoG (c) NGMeet (d) LRTFDRF
(e) FastHyMix (f) HSID (g) QRNN3D (h) SSIR

Fig. 14. Visual results of Indian Pines, locations of diagonal pixels: (40,1)-(89,50).

Table 5
Classification results of Indian Pines data.

Index	Original	NMoG	NGMeet	LRTFDRF	FastHyMix	HSID	QRNN3D	SSIR
OA	0.7923	0.8543	0.8527	0.8785	0.8790	0.8591	0.8718	**0.8793**
Kappa	0.7480	0.7894	0.7870	0.8143	0.8162	0.8186	0.7974	**0.8216**

misclassifications, which can be mitigated by denoising. SSIR reduces the fragmentary phenomenon in the classification maps, especially in soybean-notill, grass-trees, and woods regions.

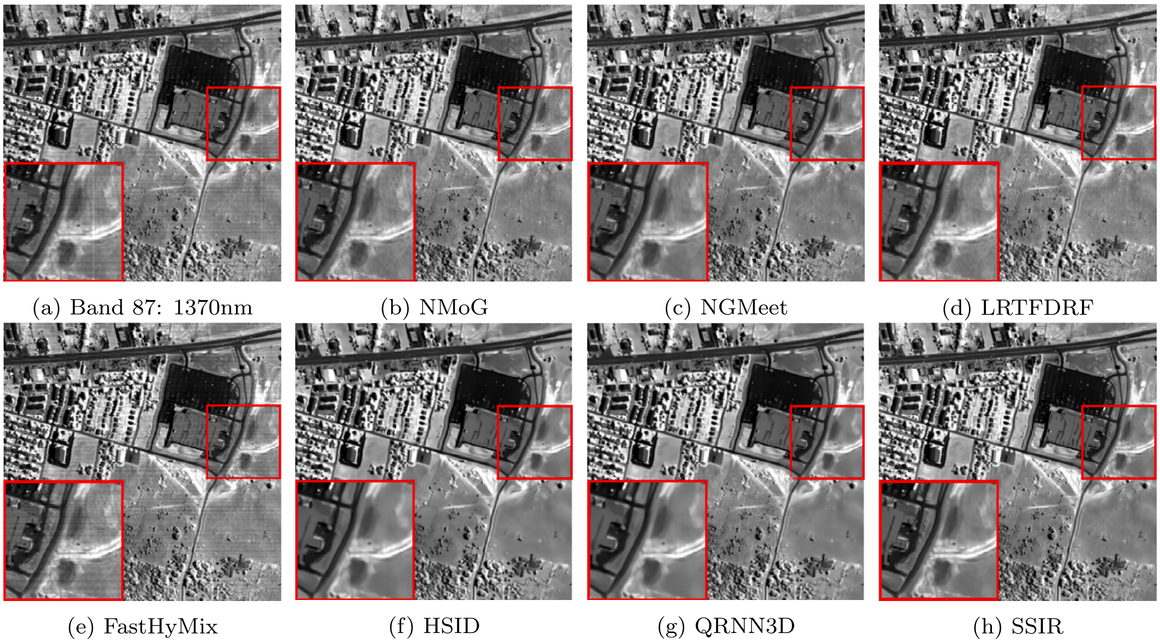
4. Discussion

The analyses about the functionality of subcomponents, parameters of the proposed SSIR model and their sensitivity, are given in this section.

4.1. Component analysis in SSIR

In this part, we analyze the contributions of SSIR components (Swin TF structure, residual connection, and unmixing module) on Chikusei data. The results are listed in Table 6.

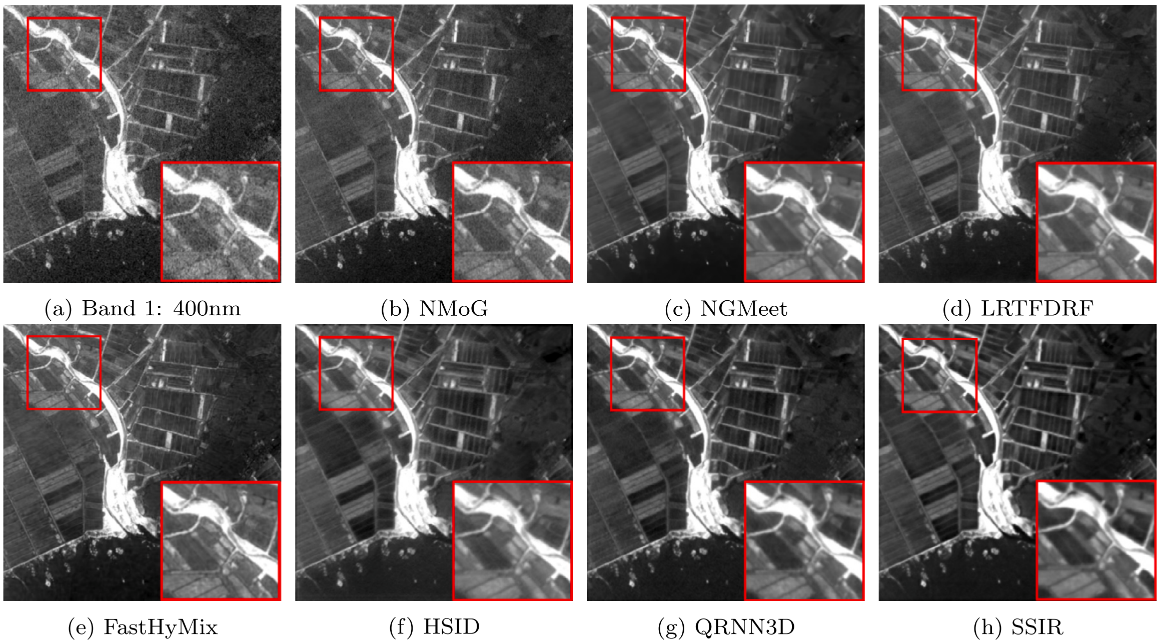
In HSIs, ground features that appear similar may vary due to differences in spatial scales, viewing angles, and aspect ratios. However, the inclusion of residual connections facilitates these differences to be fully exploited through stacked hierarchical features, resulting in improved



(a) Band 87: 1370nm (b) NMoG (c) NGMeet (d) LRTFDRF

(e) FastHyMix (f) HSID (g) QRNN3D (h) SSIR

Fig. 15. Visual results of HYDICE Urban, locations of diagonal pixels: (258,92)-(307,141).



(a) Band 1: 400nm (b) NMoG (c) NGMeet (d) LRTFDRF

(e) FastHyMix (f) HSID (g) QRNN3D (h) SSIR

Fig. 16. Visual results of OHS-1, locations of diagonal pixels: (36,14)-(85,63).

reconstruction results. Comparing the results of models (1), (2), and (3) shows that utilizing the Swin TF structure enables the integration of the long-range dependencies modeling with the local attention mechanisms of CNN, leading to better outcomes in texture-rich heterogeneous areas.

By comparing models (1) and (4), (2) and (5), and (3) and SSIR, we find that the unmixing module boosts denoising performance. The incorporation of unmixing characterizes spectral autocorrelations of land cover types and expands deep learning model generalization, resulting in improved noise suppression and spectral fidelity in heterogeneous areas.

4.2. Parameter sensitivity analysis

Parameter guidelines for SSIR, including iteration and endmember numbers, are provided in this section based on the denoising performance-computational burden tradeoff.

4.2.1. The number of iterations

The impact of increasing network depth on performance is assessed while setting the number of endmembers to 6. Table 7 presents the results of Chikusei data with varying unfolding stages. Our findings indicate that deepening network layers results in a linear growth in parameters and computational complexity presented by the floating

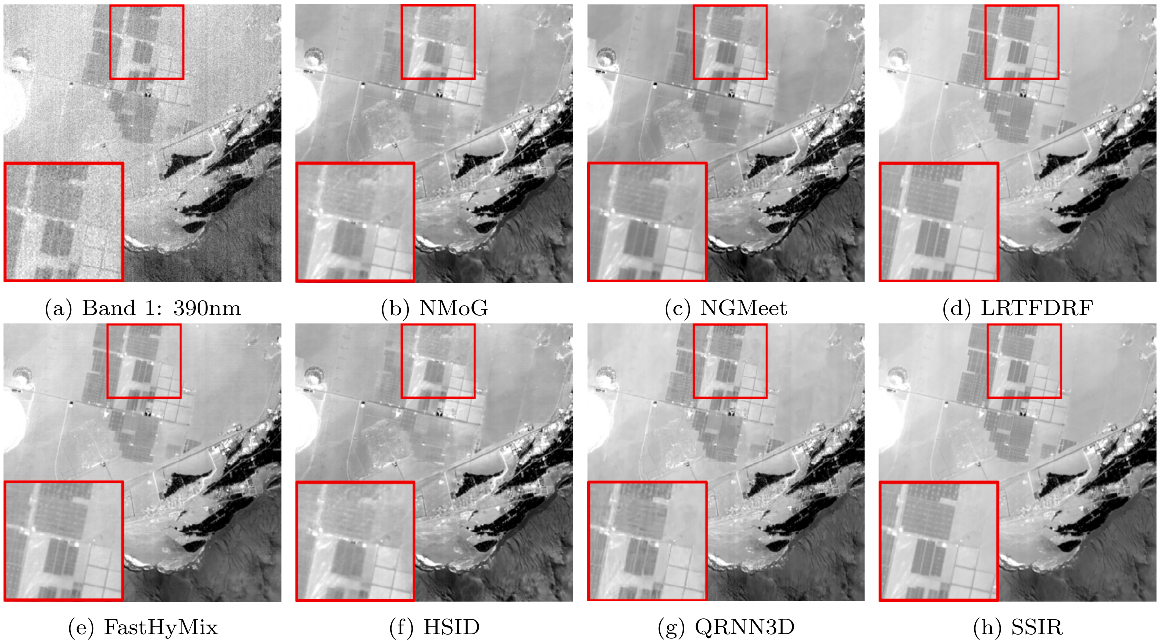
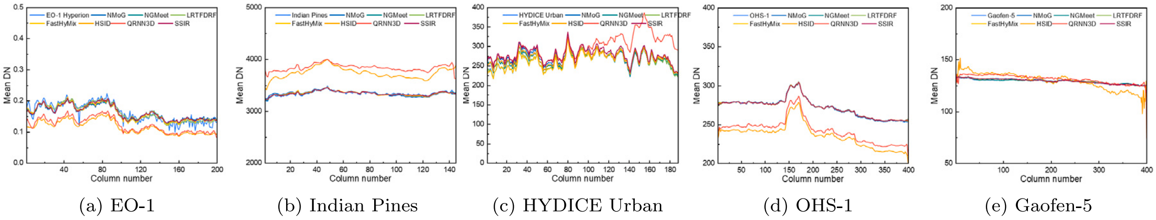

(a) Band 1: 390nm (b) NMoG (c) NGMeet (d) LRTFDRF
(e) FastHyMix (f) HSID (g) QRNN3D (h) SSIR

Fig. 17. Visual results of Gaofen-5, locations of diagonal pixels: (114,1)-(163,50).



(a) EO-1 (b) Indian Pines (c) HYDICE Urban (d) OHS-1 (e) Gaofen-5

Fig. 18. Mean vertical profiles of real datasets.

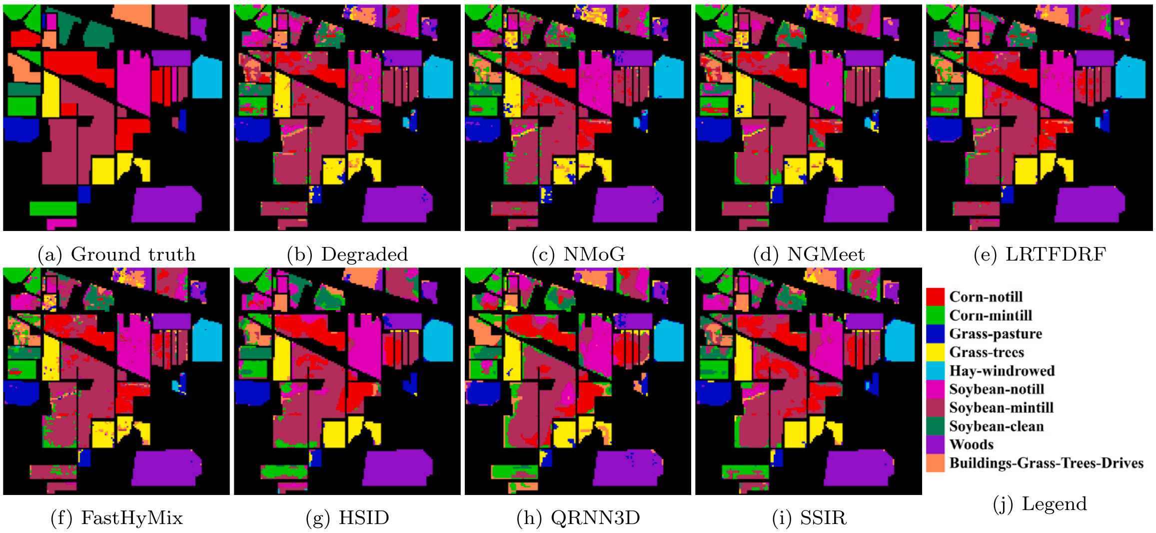
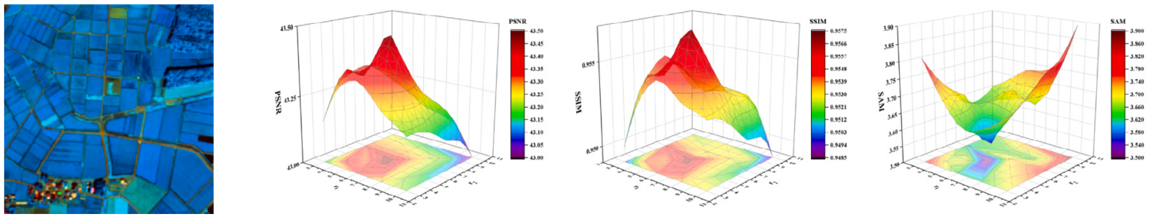
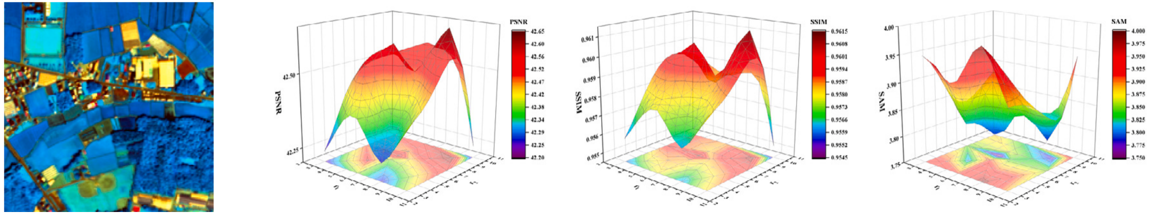

(a) Ground truth (b) Degraded (c) NMoG (d) NGMeet (e) LRTFDRF
(f) FastHyMix (g) HSID (h) QRNN3D (i) SSIR (j) Legend

Fig. 19. Classification results on Indian Pines test site.

point of operations (FLOPs), but does not necessarily improve performance proportionally. The 4D+3U (4 denoising modules and 3 unmixing modules) model has a 50% increase in computational burden compared to the 3D+2U model, yet only achieves a marginal improvement of 0.0001 in MPSNR and 0.0009 in MSA, while SSIM remains constant. Moreover, the overall performance decreases when the number of denoisers exceeds 4 (5D+4U). Although SSIR achieves rapid convergence with only a few iterations, increasing the number of iterations does not yield significant improvements due to the difficulty in training a very deep network. Therefore, we recommend using the 3D+2U structure to achieve a balance between efficiency and accuracy.



(a) Cube1 (HI_1=0.3840) (b) MPSNR of Cube1 (c) MSSIM of Cube1 (d) MSA of Cube1



(e) Cube2 (HI_2=0.8192) (f) MPSNR of Cube2 (g) MSSIM of Cube2 (h) MSA of Cube2

Fig. 20. Denoising performance with different endmember numbers r_1 and r_2 ranging from 3 to 10.

Table 6
Ablations on the Chikusei dataset (Case 2).

Model	Swin TF	RC[a]	EE[a]	MPSNR	MSSIM	MSA
(1)	×	×	×	38.8770	0.9163	6.4023
(2)	×	✓	×	39.9701	0.9307	4.7566
(3)	✓	✓	×	42.5787	0.9555	3.8649
(4)	×	×	✓	39.2343	0.9211	5.0114
(5)	×	✓	✓	40.1373	0.9332	4.5348
SSIR	✓	✓	✓	42.8172	0.9579	3.7339

[a] RC: Residual connection, EE: Endmember estimator.

Table 7
Performance of different numbers of iterations.

Index	2D+1U	3D+2U	4D+3U	5D+4U
Parameters	4.4M	6.7M	9.0M	11.2M
FLOPs	18.40G	27.65G	36.90G	46.10G
MPSNR	41.3930	42.8172	42.8173	42.3391
MSSIM	0.9412	0.9579	0.9579	0.9572
MSA	3.9813	3.7339	3.7330	4.2674

4.2.2. The number of endmembers

To study the effect of endmember numbers on performance, we test different numbers of endmembers in each unmixing module using the 3D + 2U structure. The test data includes two non-overlapping image cubes (256 × 256) from the Chikusei dataset (Case 2) with varying heterogeneity, as shown in Fig. 20(a) and (e). The heterogeneity is calculated using Eq. (20), and the number of endmembers ranges from 3 to 10.

Fig. 20 illustrates the endmember number sensitivity. Our findings indicate the endmember number has low sensitivity on overall performance, but selecting appropriate parameters based on HSI heterogeneity can improve results. The optimal results are achieved with 4–7 endmembers for low heterogeneity areas, and 7–10 endmembers for high heterogeneity areas.

5. Conclusion

We developed a novel spatial–spectral interactive restoration (SSIR) method for removing noise in HSIs. Firstly, we decouple the HSI restoration problem into two sub-problems: denoising and unmixing, by unfolding the MAP inference. The deep learning-based denoiser and an unsupervised unmixing module are integrated and trained in an end-to-end manner to obtain both good denoising results and strong generalization across different datasets. Secondly, the unmixing module characterizes the spectral autocorrelations of various land covers, which serves as a prior for the denoising module to further improve spectral fidelity. Finally, the proposed denoiser is capable of modeling both local attention and long-range dependencies by leveraging the Swin TF block and CNN, leading to further improvement in denoising performance and effectively alleviating local distortions in heterogeneous regions. Although unsupervised unmixing modules can help our method better extract spectral correlations of different land covers, the denoising performance is still limited to some extent by the training dataset. Therefore, in the future, our goal is to develop more flexible methods that have the ability to adapt more effectively to various types of data and noise.

CRediT authorship contribution statement

Shuheng Zhao: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing. **Xiaolin Zhu:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Denghong Liu:** Software, Writing – review & editing. **Fei Xu:** Writing – review & editing. **Yan Wang:** Writing – review & editing. **Liupeng Lin:** Software, Writing – review & editing. **Xuehong Chen:** Writing – review & editing. **Qiangqiang Yuan:** Writing – review & editing, Data curation.

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.

Data availability

The authors do not have permission to share data

Acknowledgments

We express our gratitude for the support provided by the National Natural Science Foundation of China (Grant Nos. 42022060), the Hong Kong PhD Fellowship Scheme, Guangdong Basic and Applied Basic Research Foundation (No. 2022B1515130001), and the Hong Kong Polytechnic University (project Nos. 4-ZZND and Q-CDBP).

References

Carvalho, S., Schlerf, M., van Der Putten, W.H., Skidmore, A.K., 2013. Hyperspectral reflectance of leaves and flowers of an outbreak species discriminates season and successional stage of vegetation. Int. J. Appl. Earth Obs. Geoinf. 24, 32–41.

Chang, Y., Yan, L., Fang, H., Zhong, S., Liao, W., 2018. HSI-DeNet: Hyperspectral image restoration via convolutional neural network. IEEE Trans. Geosci. Remote Sens. 57 (2), 667–682.

Chen, Y., Cao, X., Zhao, Q., Meng, D., Xu, Z., 2017. Denoising hyperspectral image with non-iid noise structure. IEEE Trans. Cybern. 48 (3), 1054–1066.

Chen, X., Chen, J., Jia, X., Somers, B., Wu, J., Coppin, P., 2011. A quantitative analysis of virtual endmembers' increased impact on the collinearity effect in spectral unmixing. IEEE Trans. Geosci. Remote Sens. 49 (8), 2945–2956.

Chen, Y., He, W., Yokoya, N., Huang, T.-Z., Zhao, X.-L., 2019. Nonlocal tensor-ring decomposition for hyperspectral image denoising. IEEE Trans. Geosci. Remote Sens. 58 (2), 1348–1362.

Chen, G., Qian, S.-E., 2010. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. IEEE Trans. Geosci. Remote Sens. 49 (3), 973–980.

Gege, P., Fries, J., Haschberger, P., Schötz, P., Schwarzer, H., Strobl, P., Suhr, B., Ulbrich, G., Vreeling, W.J., 2009. Calibration facility for airborne imaging spectrometers. ISPRS J. Photogramm. Remote Sens. 64 (4), 387–397.

Gonzalez, R.C., 2009. Digital Image Processing. Pearson Education India.

Greig, D.M., Porteous, B.T., Seheult, A.H., 1989. Exact maximum a posteriori estimation for binary images. J. R. Stat. Soc. Ser. B Stat. Methodol. 51 (2), 271–279.

Gu, S., Zhang, L., Zuo, W., Feng, X., 2014. Weighted nuclear norm minimization with application to image denoising. In: Proc. CVPR. pp. 2862–2869.

He, W., Yao, Q., Li, C., Yokoya, N., Zhao, Q., 2019. Non-local meets global: An integrated paradigm for hyperspectral denoising. In: Proc. CVPR. pp. 6868–6877.

He, R., Zheng, W.-S., Tan, T., Sun, Z., 2013. Half-quadratic-based iterative minimization for robust sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. 36 (2), 261–275.

Jain, A.K., 1981. Image data compression: A review. Proc. IEEE 69 (3), 349–389.

Kereszturi, G., Schaefer, L.N., Schleiffarth, W.K., Procter, J., Pullanagari, R.R., Mead, S., Kennedy, B., 2018. Integrating airborne hyperspectral imagery and LiDAR for volcano mapping and monitoring through image classification. Int. J. Appl. Earth Obs. Geoinf. 73, 323–339.

Kim, J., Lee, J.K., Lee, K.M., 2016. Deeply-recursive convolutional network for image super-resolution. In: Proc. CVPR. IEEE, pp. 1637–1645.

Li, X., Hu, Y., Gao, X., Tao, D., Ning, B., 2010. A multi-frame image super-resolution method. Signal Process. 90 (2), 405–414.

Li, J., Yuan, Q., Shen, H., Zhang, L., 2016. Noise removal from hyperspectral image with joint spectral–spatial distributed sparse representation. IEEE Trans. Geosci. Remote Sens. 54 (9), 5425–5439.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proc. ICCV. pp. 10012–10022.

Nascimento, J., Dias, J., 2005. Vertex component analysis: a fast algorithm to unmix hyperspectral data. IEEE Trans. Geosci. Remote Sens. 43 (4), 898–910.

Nichol, J., Vohora, V., 2004. Noise over water surfaces in landsat TM images. Int. J. Remote Sens. 25 (11), 2087–2093.

Pande-Chhetri, R., Abd-Elrahman, A., 2011. De-striping hyperspectral imagery using wavelet transform and adaptive frequency domain filtering. ISPRS J. Photogramm. Remote Sens. 66 (5), 620–636.

Pitas, I., Venetsanopoulos, A.N., 2013. Nonlinear Digital Filters: Principles and Applications, Vol. 84. Springer Science & Business Media.

Rasti, B., Scheunders, P., Ghamisi, P., Licciardi, G., Chanussot, J., 2018. Noise reduction in hyperspectral imagery: Overview and application. Remote Sens. 10 (3), 482.

Van der Meer, F.D., Van der Werff, H.M., Van Ruitenbeek, F.J., Hecker, C.A., Bakker, W.H., Noomen, M.F., Van Der Meijde, M., Carranza, E.J.M., De Smeth, J.B., Woldai, T., 2012. Multi-and hyperspectral geologic remote sensing: A review. Int. J. Appl. Earth Obs. Geoinf. 14 (1), 112–128.

Wang, Q., Ding, X., Tong, X., Atkinson, P.M., 2021. Spatio-temporal spectral unmixing of time-series images. Remote Sens. Environ. 259, 112407.

Wang, J., Guo, Y., Ying, Y., Liu, Y., Peng, Q., 2006. Fast non-local algorithm for image denoising. In: Proc. ICIP. IEEE, pp. 1429–1432.

Wei, K., Fu, Y., Huang, H., 2020. 3-d quasi-recurrent neural network for hyperspectral image denoising. IEEE Trans. Neural Netw. Learn. Syst. 32 (1), 363–375.

Yuan, Q., Zhang, Q., Li, J., Shen, H., Zhang, L., 2018. Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network. IEEE Trans. Geosci. Remote Sens. 57 (2), 1205–1218.

Yuan, Q., Zhang, L., Shen, H., 2012. Hyperspectral image denoising employing a spectral–spatial adaptive total variation model. IEEE Trans. Geosci. Remote Sens. 50 (10), 3660–3677.

Zhang, K., Gool, L.V., Timofte, R., 2020a. Deep unfolding network for image super-resolution. In: Proc. CVPR. pp. 3217–3226.

Zhang, H., He, W., Zhang, L., Shen, H., Yuan, Q., 2013. Hyperspectral image restoration using low-rank matrix recovery. IEEE Trans. Geosci. Remote Sens. 52 (8), 4729–4743.

Zhang, H., Liu, L., He, W., Zhang, L., 2019. Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition. IEEE Trans. Geosci. Remote Sens. 58 (5), 3071–3084.

Zhang, K., Sun, M., Han, T.X., Yuan, X., Guo, L., Liu, T., 2017. Residual networks of residual networks: Multilevel residual networks. IEEE Trans. Circuits Syst. Video Technol. 28 (6), 1303–1314.

Zhang, Q., Yuan, Q., Li, J., Sun, F., Zhang, L., 2020b. Deep spatio-spectral Bayesian posterior for hyperspectral image non-iid noise removal. ISPRS J. Photogramm. Remote Sens. 164, 125–137.

Zheng, Y.-B., Huang, T.-Z., Zhao, X.-L., Chen, Y., He, W., 2020. Double-factor-regularized low-rank tensor factorization for mixed noise removal in hyperspectral image. IEEE Trans. Geosci. Remote Sens. 58 (12), 8450–8464.

Zhu, X., Milanfar, P., 2010. Automatic parameter selection for denoising algorithms using a no-reference measure of image content. IEEE Trans. Image Process. 19 (12), 3116–3132.

Zhuang, L., Ng, M.K., 2021. FastHyMix: Fast and parameter-free hyperspectral image mixed noise removal. IEEE Trans. Neural Netw. Learn. Syst..

Zoran, D., Weiss, Y., 2011. From learning models of natural image patches to whole image restoration. In: Proc. ICCV. IEEE, pp. 479–486.