Unsupervised Adaptation Learning for Real Multiplatform Hyperspectral Image Denoising

Zhaozhi Luo, Xinyu Wang, *Member, IEEE*, Petri Pellikka, Janne Heiskanen, and Yanfei Zhong, *Senior Member, IEEE*

***Abstract*—Real hyperspectral images (HSIs) are ineluctably contaminated by diverse types of noise, which severely limits the image usability. Recently, transfer learning has been introduced in hyperspectral denoising networks to improve model generalizability. However, the current frameworks often rely on image priors and struggle to retain the fidelity of background information. In this article, an unsupervised adaptation learning (UAL)-based hyperspectral denoising network (UALHDN) is proposed to address these issues. The core idea is first learning a general image prior for most HSIs, and then adapting it to a real HSI by learning the deep priors and maintaining background consistency, without introducing hand-crafted priors. Following this notion, a spatial–spectral residual denoiser, a global modeling discriminator, and a hyperspectral discrete representation learning scheme are introduced in the UALHDN framework, and are employed across two learning stages. First, the denoiser and the discriminator are pretrained using synthetic noisy-clean ground-based HSI pairs. Subsequently, the denoiser is further fine-tuned on the real multiplatform HSI according to a spatial–spectral consistency constraint and a background consistency loss in an unsupervised manner. A hyperspectral discrete representation learning scheme is also designed in the fine-tuning stage to extract semantic features and estimate noise-free components, exploring the deep priors specific for real HSIs. The applicability and generalizability of the proposed UALHDN framework were verified through the experiments on real HSIs from various platforms and sensors, including unmanned aerial vehicle-borne, airborne, spaceborne, and Martian datasets. The UAL denoising scheme shows a superior denoising ability when compared with the state-of-the-art hyperspectral denoisers.**

***Index Terms*—Convolutional neural network (CNN), hybrid noise removal, hyperspectral images (HSIs), unsupervised adaptation learning (UAL).**

Manuscript received 24 February 2023; revised 21 November 2023 and 11 May 2024; accepted 31 May 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 42325105 and Grant 42211530032; in part by the National Key Research and Development Program of China under Grant 2022YFB3903502; in part by the LIESMARS Special Research Funding; in part by the EDUFI Fellowship from the Ministry of Education of Finland; in part by the ESSA Project of the European Union DG International Partnerships under the DeSIRA Program under Grant FOOD/2020/418-132; and in part by the Academy of Finland under Grant 349556. This article was recommended by Associate Editor A. H. Phan. *(Corresponding author: Xinyu Wang.)*

Zhaozhi Luo and Janne Heiskanenare are with the Department of Geosciences and Geography, the Faculty of Science, University of Helsinki, 00014 Helsinki, Finland (e-mail: zhaozhi.luo@helsinki.fi; janne.heiskanen@helsinki.fi).

Xinyu Wang is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: wangxinyu@whu.edu.cn).

Petri Pellikka is with the Department of Geosciences and Geography, the Faculty of Science, University of Helsinki, 00014 Helsinki, Finland, and also with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: petri.pellikka@helsinki.fi).

Yanfei Zhong is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: zhongyanfei@whu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2024.3412270.

Digital Object Identifier 10.1109/TCYB.2024.3412270

I. Introduction

With the discriminative power of spectra, hyperspectral data acquired from multiple platforms (e.g., unmanned aerial vehicles (UAVs), and airborne and spaceborne platforms) have extensive applications, such as target tracking [1] and anomaly detection [2]. Nevertheless, on account of the drawbacks of hyperspectral sensors, hyperspectral images (HSIs) are ineluctably corrupted by diverse types of instrument noise, including quantization noise, thermal noise, and shot noise [3]. Moreover, poor atmospheric conditions are likely to induce noise, and imperfect calibration is prone to causing stripe noise in the preprocessing [4], particularly for spaceborne hyperspectral sensors. This complex degradation severely reduces the image quality and further limits image usability in real applications. Therefore, it is vital to remove the hybrid noise in real HSIs and restore the latent noise-free HSIs from the noisy observations.

When faced with real HSI denoising, a robust HSI denoiser should have the following desirable attributes: 1) effectiveness and efficiency and 2) powerful generalizability for HSIs obtained from any platform. Such a denoiser should be able to be directly used to restore high-quality images when the noise pattern of the HSIs is known. When the noise pattern is complicated and unknown, as is often the case with real multiplatform HSIs, the denoiser should still be able to produce robust denoising results.

However, the state-of-the-art HSI denoising methods are still limited in effectiveness, efficiency, and generalizability. The HSI denoising methods can be classified as follows: filtering-based methods, optimization-based methods, and learning-based methods. The filtering-based methods, such as the wavelet-Fourier adaptive filter (WFAF) [5] and multidimensional Wiener filtering (MWF) [6], can restrain noise to some degree, but they are intrinsically restricted due to the fixed transform basis. The optimization-based methods, such as low-rank matrix recovery (LRMR) [7], are effective in real HSI denoising with unknown noise patterns, but they do suffer from several drawbacks. On the one hand, image priors

may not be compatible with complex image structure representation and cannot easily model multitype hybrid noise. On the other hand, when using the optimization-based methods, it is necessary to empirically fine-tune the hyperparameters for each HSI, due to the sensitivity of the model parameters. However, this is a waste of computational time and resources from the perspective of practical applications. Last but not least, an excellent denoising performance typically requires a time-consuming optimization process.

Deep learning has been applied to the HSI denoising problem due to its powerful nonlinear representation ability and high-computational efficiency [8]. Generally speaking, learning-based methods are aimed at learning the mapping from the input noisy observation to the matching high-quality image in a supervised manner, and the denoising performance is highly reliant on numerous noisy-clean image pairs for the model training. Unfortunately, real noisy-clean image pairs are often unavailable for airborne/spaceborne hyperspectral remote sensing platforms. An alternative solution is to train a network with synthetic noisy-clean image pairs, which are generated using ground-based/airborne high-quality HSIs, and the multitype hybrid noise in airborne/spaceborne degraded HSIs is simulated. However, on account of the innumerable imaging scenes, real HSIs tend to cover complex structures and varied terrain, resulting in an extremely complicated intrinsic prior. Moreover, the unknown degradation in the imaging procedure often induces complex hybrid noise in the observations. Therefore, most of the existing networks trained with synthetic data are limited in generalizability, and often fail to obtain a satisfactory denoising performance when migrated to real data.

Recently, several unsupervised deep denoisers have been designed for the above issues, for example, Stein's unbiased risk estimate (SURE) [9], subspace representation deep prior [10], and low-rank deep prior [11], [12]. Denoisers can be trained directly on unpaired real data in a self-supervised manner. Despite the advancement, model performance could be limited when noise patterns diverge from the predefined noise assumptions or the constructed priors of these models. Additionally, training models on each real HSI is time-consuming without a reasonable strategy in real applications. Another way to improve model generalizability is to introduce transfer learning in denoising tasks [13], [14], [15]. This also involves unsupervised learning on real HSIs, often relying on specific noise assumptions [16] or specific image priors like total variation (TV) prior [14] for model fine-tuning. Although it can improve computational efficiency compared to the self-supervised denoisers, the limitations related to noise assumptions and priors persist in the current transfer learning denoising frameworks. They are less effective in exploring the deep priors specific to real HSIs with complex noise patterns and in maintaining the fidelity of background information during denoising.

To overcome the defects of the existing transfer learning denoising approaches, an innovative unsupervised adaptation learning (UAL)-based hyperspectral transfer learning denoising network (UALHDN) is proposed in this article for multiplatform HSI denoising. The proposed model is first pretrained to learn an image prior suitable for most HSIs, and is next adapted to a certain real HSI by learning the deep priors and retaining background consistency in an unsupervised manner. Compared with most of learning-based denoising methods, the proposed model can generalize well on real multiplatform HSIs, based on easily accessible ground-based data and one or several real HSIs, without the need to introduce image priors and extra supervisory data. Following this idea, the UALHDN framework incorporates a spatial–spectral residual denoiser, a global modeling discriminator, and a hyperspectral discrete representation learning scheme, and are employed across two learning stages, that is, a synthetic image pretraining stage and a real image adaptation denoising stage. In the synthetic image pretraining stage, the model is pretrained on synthetic noisy-clean ground-based HSI pairs in a supervised manner. Therein, the denoiser is used for general HSI denoising, and the discriminator is used for determining whether a denoised HSI is clean or not. Subsequently, the denoiser is further fine-tuned on real satellite/airborne HSIs according to the designed UAL loss function. The UAL loss comprises a spatial–spectral consistency constraint coupled with a background consistency loss (BC). It aims to both minimize the loss of essential information in denoised results and maintain background consistency between real HSIs and their estimated noise-free components. Furthermore, the hyperspectral discrete representation learning scheme is designed to exclude noise components, extract semantic features, and estimate noise-free components of real HSIs. Self-construction and cross-construction constraints are employed in the unsupervised learning scheme to learn the specific priors of real HSIs. Consequently, the denoiser can maintain high fidelity of background information and learn a more image-specific deep prior, effectively increasing the generalizability of the model. To demonstrate the effectiveness of the proposed model on real multiplatform HSIs, UALHDN was compared with the state-of-the-art HSI denoising methods on HSIs obtained from various platforms and sensors, including the airborne Washington DC Mall dataset, the WHU-Hi-Baoxie UAV dataset, a Chinese SPARK hyperspectral satellite dataset, and a Martian dataset from the compact reconnaissance imaging spectrometer for Mars (CRISM). The results clearly validate the superiority of UALHDN in the denoising performance. Furthermore, the ablation experiments on various datasets prove that the strategy of UAL can effectively increase the generalizability of the model on real HSIs. Moreover, the model used to perform UAL on a given noisy HSI was found to generalize well without UAL on other HSIs acquired under similar circumstances via the same sensor. Therefore, the running time of the UALHDN model can be greatly improved by using only part of the real data in the fine-tuning stage. On the whole, the UALHDN framework shows great potential for real multiplatform HSI denoising applications.

The main contributions of this article can be outlined as follows.

1) A novel UAL-based hyperspectral transfer learning denoising framework is proposed for real multiplatform HSI denoising. The proposed framework first learns a general image prior for most HSIs in the

pretraining stage, and then adapts it to a real HSI by learning the deep priors and maintaining background consistency, without introducing hand-crafted priors. A spatial–spectral consistency constraint and a BC are designed in the fine-tuning stage to minimize the loss of essential information while keeping the background fidelity between real HSIs and their estimated noise-free components.

2) An innovative hyperspectral discrete representation learning scheme is designed in the fine-tuning stage to exclude noise components, extract semantic features, and estimate noise-free components of real HSIs. The scheme incorporates self-construction and cross-construction constraints into the unsupervised learning scheme for noisy and clean bands of HSIs, which facilitates the learning of specific deep priors of real data.
3) The UALHDN framework achieves exceptional performance and surpasses the state-of-the-art denoising methods on a variety of HSIs obtained from different platforms and sensors, including UAV-borne, airborne, spaceborne, and Martian datasets, demonstrating its effectiveness and broad applicability for real multiplatform HSI denoising.

The remainder of this article is organized as follows. The related works are introduced in Section II. The proposed model is depicted in Section III. The experimental results and a discussion are provided in Section IV. Finally, our conclusions are summarized in Section V.

II. Related Work

Many learning-based approaches have been proposed over the last few decades. Therein, some unsupervised methods have been developed based on transfer learning [13], [14], [15] or self-supervised learning [9], [10], [11], [12], [17], [18], [19]. In this section, the two major categories of HSI denoising methods and transfer learning theory are briefly introduced.

A. Optimization-Based Methods

Optimization-based methods employ a reasonable prior for HSIs and evaluate the latent noise-free image from the noisy observation through an optimization framework [7]. The inverse optimization problem is well posed by specific priors, such as TV, sparse representation, and low-rank priors. By considering the rich spectral information of HSIs, the TV prior has been applied in HSI denoising [20], [21]. From the perspective of sparse representation, a noise-free HSI is supposed to be encoded by a sparse linear combination of atoms from the over-complete dictionary. Based on this prior, the spatial–spectral sparse representation model [22] was developed for HSI noise removal. From the standpoint of the low-rank prior, a noise-free high-dimensional HSI is located in a low-dimensional subspace, on account of the highly related spectral-spatial characteristics of HSIs. Based on the low-rank conception, 2-D LRMR models [7], [23], [24], [25] have been developed for HSI noise removal tasks. Viewing HSIs as tensors, 3-D low-rank tensor approximation models [26], [27], [28], [29], [30], [31] have been introduced to exploit both the spectral correlation and spatial geometrical features. In addition, multiple priors can be combined to obtain more competitive results [32], [33]. However, image priors may not be suitable for complex image structures and restrict the model spatial–spectral representation. In addition, the optimization-based methods suffer from a high-computational burden due to the time-consuming optimization process.

B. Learning-Based Methods

As an alternative to manually designing priors for HSIs, the learning-based methods are aimed at automatically learning general image priors from a number of noisy-clean training pairs in a supervised manner via an end-to-end network. They then restore the latent clean HSIs through a mapping from the noisy observation to the corresponding high-quality image [8]. By considering the spectral consistency of HSIs, the convolutional neural network (CNN)-based HSI denoising method (HSI-DeNet) [34] and the combined spatial–spectral deep CNN model [35] utilize multichannel 2-D convolution to combine neighboring spectral bands. Based on attention theory, a 3-D attention denoising network [36] was developed to consider global spatial–spectral correlation. Recurrent neural networks have also been utilized to extract the global spectral correlation of HSIs [37], [38]. To characterize the physical constraint and extract deep priors for HSIs, some methods have integrated various image priors into CNNs [39], [40], [41], [42]. In addition, Transformer models have been utilized in hyperspectral denoising [18], [43], [44] to effectively capture intrinsic global spatial–spectral correlations.

Nevertheless, due to these methods being highly reliant on lots of noisy-clean image pairs for the model training, and the absence of image pairs for real HSIs, most of the learning-based methods generate synthetic image pairs by simulating multitype hybrid noise and adding it to clean ground-based/airborne HSIs. Consequently, the structural differences and degradation discrepancies between synthetic and real HSIs cause models trained with synthetic data to have limited generalizability on real HSIs. Based on the assumption that CNNs are more likely to converge with clean images rather than noisy images, a deep hyperspectral prior (DHP) method [19] was proposed to train a CNN without paired HSIs. Unsupervised training can also be achieved based on SURE [9]. The combination of supervised learning and SURE has been shown to improve the denoising performance on real data [18]. However, the Gaussian noise distribution assumption in these methods limits their effectiveness in hybrid noise removal. By considering adjacent bands as independent noisy images with the same background information, the two-stage HSI denoising network model [15] achieves fine-tuning on real data based on noise-to-noise learning [16]. Moreover, a self-supervised denoising scheme [17] was developed to generate noisy-clean image pairs from real HSIs by extracting real noise samples from noisy spectral bands and then adding the noise to the other clean bands. Unfortunately, these models tend to consider band-dependent noise and stripe noise with

fixed patterns as the background components. A deep external-internal learning scheme [14] was proposed for noise removal of ground-based datasets, employing adversarial learning to explore an authentic deep prior and utilizes a TV prior to fine-tune the network on real data. low-rank prior [11], [12] and subspace representation [10] can also be integrated into self-supervised denoising schemes. However, the image priors might not always be suitable for cases with various noise patterns and can restrict the spatial–spectral representation of the model. The effectiveness of the model for real multiplatform HSIs with more complex noise patterns also remains to be verified. In addition, a real ground-based dataset [45] with paired noisy and clean HSIs was collected via controlling the exposure time of the camera. But the noise of hyperspectral remote sensing images is more complex in type due to the difference in the imaging processes of the various sensors and the influence of the atmosphere. As a result, this dataset provides very limited guidance for hybrid noise removal in hyperspectral remote sensing images.

C. Transfer Learning

Transfer learning is aimed at transferring knowledge learned from the source domain to the target domain [46]. For tasks with limited real labels, for example, biomedical tasks and denoising tasks, there is an urgent need for unsupervised transfer learning. Transfer learning can be conducted by reusing a portion of a pretrained network and fine-tuning it with a small set of labeled real data [47]. In addition, multiscale convolutional sparse coding [48] was proposed for biomedical classification and provides an unsupervised way to transfer learned knowledge. Another approach involves designing a domain classifier [49] to distinguish whether the extracted features originate from the source or target domain. The feature extractor can then be trained to reduce the domain discrepancy by maximizing the domain classification loss. This mode resembles the generative adversarial networks (GANs) [50], where a discriminator distinguishes the input. Following this idea, models trained on synthetic data can be adapted to real data by adversarial learning in an unsupervised manner [51]. Denoising also suffers from domain shift, due to the degradation discrepancies and intrinsic features between synthetic and real data. For this issue, adversarial learning has been utilized in time-of-flight data denoising [13] to perform unsupervised domain adaptation from synthetic to real data. It is also possible to fine-tune a denoising network on real ground-based data by a spatial–spectral constraint and TV prior [14]. In addition, when faced with independent noise, fine-tuning on real data can be achieved by the noise-to-noise strategy of adjacent bands [15]. It is worth noting that the above models tend to introduce image priors and struggle to retain the fidelity of background information.

D. Discrete Representation

Discrete representation involves transforming data into distinct and comprehensible components for specific applications. This technique is particularly useful in tasks like image restoration and enhancement, where it aids in isolating elements of degradation factors from main contents. For instance, distinct encoding methods can differentiate between the core content of blurred images and the corresponding blurring attributes [52]. Likewise, in images affected by rain, this approach can distinguish the primary background from extraneous elements like rain patterns [53]. Discrete representation also facilitates the differentiation between information specific to sensors and components inherent to scenes in image fusion [54]. Furthermore, this technique proves beneficial in separating noise elements from observations using unpaired natural images [55].

This study aims to integrate the benefits of discrete representation into the hyperspectral transfer learning denoising framework. A hyperspectral discrete representation learning scheme is designed during fine-tuning stage to extract semantic features and estimate noise-free components, exploring the specific deep priors of real HSIs in an unsupervised manner. In addition, a spatial–spectral consistency constraint and a BC are designed in the fine-tuning stage to minimize the loss of essential information while maintaining the background consistency between real HSIs and their estimated noise-free components.

III. Unsupervised Adaptation Learning Hyperspectral Denoising Network

The UALHDN framework is designed to retain the fidelity of background information and learn deep priors of real HSIs. As shown in Fig. 1, the UALHDN framework is employed across two learning stages: 1) a synthetic image pretraining stage and 2) a real image adaptation denoising stage. Full details are presented below.

A. Problem Formulation

HSIs are inevitably contaminated by diverse types of noise, including common ones, such as Gaussian noise and stripe noise [7]. In this article, all the kinds of noise are empirically assumed to be additive components [22]. The basic model of HSI degradation is then defined as

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

where Y denotes a detected low-quality HSI composed of B spectral bands with the height of H and the width of W, X represents the latent high-quality HSI, and N represents all the additive noise. The ultimate target of the HSI denoising task is to recover the latent noise-free HSI X from the low-quality HSI Y. The optimization-based denoising methods are aimed at solving the following problem:

$$\widehat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}}\ \Phi(\mathbf{Y} - \mathbf{X}) + \lambda * \mathcal{R}(\mathbf{X}) \tag{2}$$

where $\Phi(Y-X)$ represents the data fidelity term, $\mathcal{R}(X)$ denotes the regularization term related to the image priors on X, and λ controls the balance between the data fidelity term and the regularization term.

B. Unsupervised Adaptation Learning

Instead of introducing image priors, as in (2), a deep CNN is employed to directly model the noise pattern of X and learn

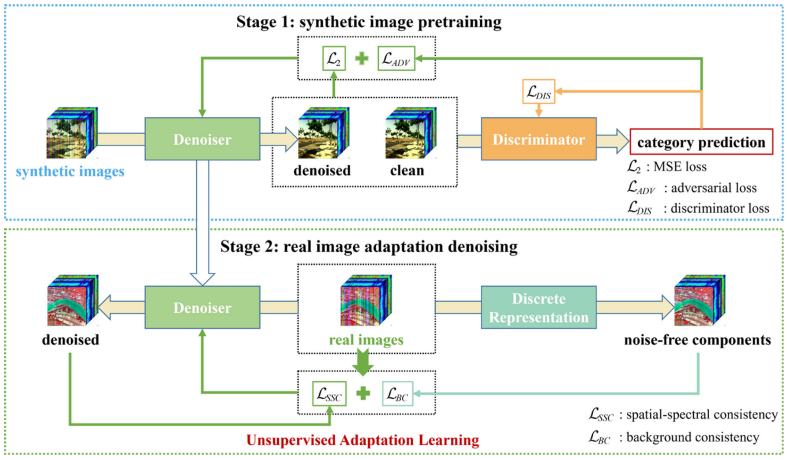


Fig. 1. Overview of the UALHDN framework.

the prior during the network training. To learn a general prior for most HSIs and adapt to a real noisy HSI, The UALHDN framework is achieved in two stages: 1) a synthetic image pretraining stage and 2) a real image adaptation denoising stage. Detailed descriptions of these stages are presented in the following sections.

1) Synthetic Image Pretraining: The UALHDN model is first pretrained on synthetic image pairs in a supervised manner to obtain a general image prior. In this stage, the discriminator is trained to determine whether the denoised HSI appears clean and natural [14]. It aims to solve the following problem:

$$\widehat{\Theta}_{\mathcal{D}} = \underset{\Theta_{\mathcal{D}}}{\operatorname{argmin}}\ \mathcal{L}_{+}(\mathbf{X}) + \mathcal{L}_{-}(\mathbf{Y}) \tag{3}$$

where $\Theta_{\mathcal{D}}$ represents the trainable parameters of the discriminator, $\mathcal{L}_{+}(X)$ denotes the clean HSI loss, and $\mathcal{L}_{-}(Y)$ denotes the denoised HSI loss. When a clean HSI is input into the discriminator, the result should be clean. $\mathcal{L}_{+}(X)$ is then defined as follows:

$$\mathcal{L}_{+}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}\|\mathcal{D}(\mathbf{X}_i;\Theta_{\mathcal{D}}) - 1\|_2^2 \tag{4}$$

where X_i denotes the target clean HSI, $\mathcal{D}(\cdot;\Theta_{\mathcal{D}})$ denotes the discriminator network parameterized by $\Theta_{\mathcal{D}}$, and n represents the number of batches. Similarly, the result can be expected not to be clean if the input is the denoised HSI from the denoiser during discriminator training. Therefore, $\mathcal{L}_{-}(Y)$ is designed as follows:

$$\mathcal{L}_{-}(\mathbf{Y}) = \frac{1}{n}\sum_{i=1}^{n}\|\mathcal{D}(\sigma(\mathbf{Y}_i;\Theta_{\sigma});\Theta_{\mathcal{D}})\|_2^2 \tag{5}$$

where Y_i denotes the observed HSI in the synthetic training pairs and $\sigma(\cdot;\Theta_{\sigma})$ denotes the denoiser module parameterized by $\Theta_{\mathcal{G}}$. On the other hand, the denoiser is trained to achieve general HSI denoising and aims to solve the following problem:

$$\widehat{\Theta}_{\sigma} = \underset{\Theta_{\sigma}}{\operatorname{argmin}}\ \mathcal{L}_2(\mathbf{X},\mathbf{Y}) + \lambda_1 * \mathcal{L}_{\mathrm{ADV}}(\mathbf{Y}) \tag{6}$$

where $\mathcal{L}_2(X,Y)$ denotes the mean-squared error (MSE) between the output of the denoiser and the target clean HSI, $\mathcal{L}_{\mathrm{ADV}}(Y)$ represents the adversarial loss, and λ_1 controls the balance between the two terms. The pixelwise MSE loss $\mathcal{L}_2(X,Y)$ is as follows:

$$\mathcal{L}_2(\mathbf{X},\mathbf{Y}) = \frac{1}{n}\sum_{i=1}^{n}\|\sigma(\mathbf{Y}_i;\Theta_{\sigma}) - \mathbf{X}_i\|_2^2. \tag{7}$$

The output denoising results of the denoiser are expected to be clean HSIs when they are input into the discriminator. Therefore, the adversarial loss $\mathcal{L}_{\mathrm{ADV}}(Y)$ is as follows:

$$\mathcal{L}_{\mathrm{ADV}}(\mathbf{Y}) = \frac{1}{n}\sum_{i=1}^{n}\|\mathcal{D}(\sigma(\mathbf{Y}_i;\Theta_{\sigma});\Theta_{\mathcal{D}}) - 1\|_2^2. \tag{8}$$

2) Real Image Adaptation Denoising: To learn the deep prior and retain background consistency for a specific real HSI, it is crucial to estimate the background information and fine-tune the denoiser parameters in an unsupervised manner. To achieve this, the pretrained parameters of the denoiser $\Theta_{\mathcal{G}}$ are further updated according to the designed UAL loss on the real HSI. The UAL loss function consists of a spatial–spectral consistency constraint and a BC, as follows:

$$\widehat{\Theta}_{\sigma} = \underset{\Theta_{\sigma}}{\operatorname{argmin}}\ \mathcal{L}_{\mathrm{SSC}}(\mathbf{Y}') + \lambda_2 * \mathcal{L}_{\mathrm{BC}}(\mathbf{Y}') \tag{9}$$

where $\mathcal{L}_{\mathrm{SSC}}(Y')$ denotes the spatial–spectral consistency constraint, $\mathcal{L}_{\mathrm{BC}}(Y')$ denotes the BC on the real HSIs, and λ_2 controls the balance between the two terms. $\mathcal{L}_{\mathrm{SSC}}(Y')$ is calculated as follows:

$$\mathcal{L}_{\mathrm{SSC}}(\mathbf{Y}') = \frac{1}{n}\sum_{i=1}^{n}\|\sigma(\mathbf{Y}'_i;\Theta_{\sigma}) - \mathbf{Y}''_i\|_1 \tag{10}$$

where Y'_i denotes the real HSI. The constraint of spatial–spectral consistency is employed to minimize the loss of critical information throughout the denoising process. It is important to note that the choice of loss function for the spatial–spectral consistency constraint differs from the reconstruction term used in the pretraining stage. On the one hand, compared to l_1-norm loss, the gradient of l_2-norm loss decreases as the error decreases and large errors are heavily penalized, promoting better convergence during the general prior training. On the other hand, l_2-norm loss suffers from the

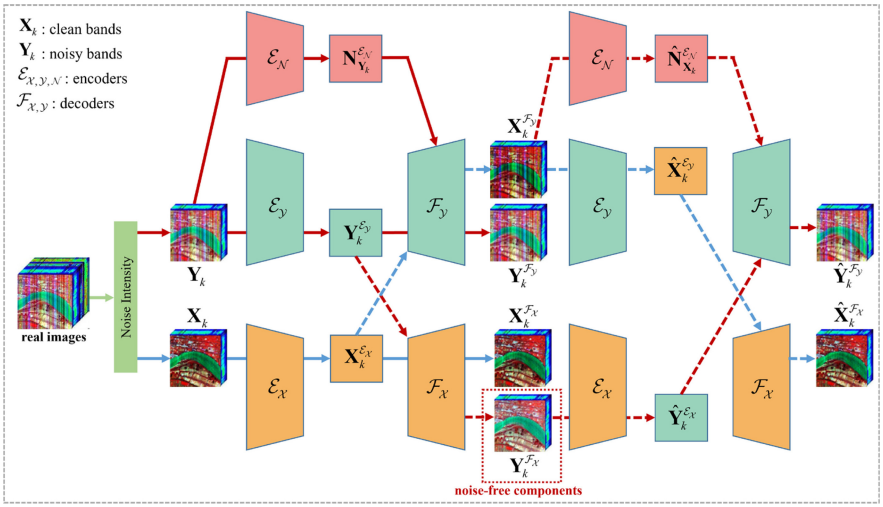


Fig. 2. Structure of the hyperspectral discrete representation learning scheme.

Gaussian noise assumption and the signal-independent noise assumption. In comparison, l_1-norm loss may be more suitable for hybrid noise removal and perform better in suppressing possible spotty artifacts in flat regions [56]. Therefore, l_2-norm loss is employed in the reconstruction term of the synthetic image pretraining stage, and l_1-norm is utilized in the spatial–spectral consistency constraint of the real image adaptation denoising stage, instead of using l_2-norm loss in both stages [14]. Additionally, $\mathcal{L}_{\mathrm{BC}}(Y')$ is calculated as follows:

$$\mathcal{L}_{\mathrm{BC}}(\mathbf{Y}') = \frac{1}{n}\sum_{i=1}^{n}\left\| \begin{matrix} \mathcal{B}_\sigma\left(\mathcal{F}_{\mathbf{x}}\left(\mathcal{E}_{\mathbf{y}}\left(\mathbf{Y}'_i;\Theta_{\mathcal{E}_{\mathbf{y}}}\right);\Theta_{\mathcal{F}_{\mathbf{x}}}\right)\right) \\ -\mathcal{B}_\sigma\left(\mathbf{Y}'_i\right) \end{matrix} \right\|_1 \tag{11}$$

where $\mathcal{B}_\sigma(\cdot)$ represents operators associated with Gaussian blurring. σ denotes the standard deviation of the convolution kernel, empirically determined to be 15. $\mathcal{E}_{\mathbf{y}}(\cdot;\Theta_{\mathcal{E}_{\mathbf{y}}})$ and $\mathcal{F}_{\mathbf{x}}(\cdot;\Theta_{\mathcal{F}_{\mathbf{x}}})$ refer to an encoder parameterized by $\Theta_{\mathcal{E}_{\mathbf{y}}}$ for noisy images and a decoder parameterized by $\Theta_{\mathcal{F}_{\mathbf{x}}}$ for clean images. The noisy encoder and clean decoder, respectively, aim to extract semantic features from noisy images and restore noise-free components. The noise-free components and original observation, after being processed by Gaussian blurring operators, are assumed to be background information and should have similar attributes [57]. The design of $\mathcal{L}_{\mathrm{BC}}(Y')$ aims to maintain consistency in the background between real images and their respective noise-free components.

To obtain such $\mathcal{E}_{\mathbf{y}}(\cdot;\Theta_{\mathcal{E}_{\mathbf{y}}})$ and $\mathcal{F}_{\mathbf{x}}(\cdot;\Theta_{\mathcal{F}_{\mathbf{x}}})$, a hyperspectral discrete representation learning scheme, inspired by [55], is formulated for real HSIs, as shown in Fig. 2. The proposed scheme aims at separating noise attributes and extracting semantic features from real HSIs based on self-reconstruction and cross-reconstruction constraints. Subsequently, rough noise-free components can be restored from extracted features. To achieve it, the noisy autoencoder $\mathcal{E}_{\mathbf{y}}(\cdot;\Theta_{\mathcal{E}_{\mathbf{y}}})$ and $\mathcal{F}_{\mathbf{y}}(\cdot;\Theta_{\mathcal{F}_{\mathbf{y}}})$, and the clean autoencoder $\mathcal{E}_{\mathbf{x}}(\cdot;\Theta_{\mathcal{E}_{\mathbf{x}}})$ and $\mathcal{F}_{\mathbf{x}}(\cdot;\Theta_{\mathcal{F}_{\mathbf{x}}})$ are introduced to model noisy and clean spectral bands, respectively. An extra noise encoder $\mathcal{E}_{\mathcal{N}}(\cdot;\Theta_{\mathcal{E}_{\mathcal{N}}})$ is employed to extract noise attributes. HSIs are categorized as noisy bands Y_k and clean bands X_k based on noise intensity estimation [58]. The self-reconstruction constraint for two autoencoders is defined as follows:

$$\begin{aligned}&\mathcal{L}_{\mathrm{SR}}(\mathbf{X}_k,\mathbf{Y}_k)\\ &=\frac{1}{n}\sum_{k=1}^{n}\left\|\mathcal{F}_{\mathbf{x}}\left(\mathcal{E}_{\mathbf{x}}\left(\mathbf{X}_k;\Theta_{\mathcal{E}_{\mathbf{x}}}\right);\Theta_{\mathcal{F}_{\mathbf{x}}}\right)-\mathbf{X}_k\right\|_1+\\ &\quad\left\|\mathcal{F}_{\mathbf{y}}\left(\mathcal{E}_{\mathbf{y}}\left(\mathbf{Y}_k;\Theta_{\mathcal{E}_{\mathbf{y}}}\right)+\mathcal{E}_{\mathcal{N}}\left(\mathbf{Y}_k;\Theta_{\mathcal{E}_{\mathcal{N}}}\right);\Theta_{\mathcal{F}_{\mathbf{y}}}\right)-\mathbf{Y}_k\right\|_1.\end{aligned} \tag{12}$$

Semantic features $Y_k^{\mathcal{E}_{\mathbf{y}}}$ and noise attributes $N_{Y_k}^{\mathcal{E}_{\mathcal{N}}}$ are derived from noisy bands Y_k by $\mathcal{E}_{\mathbf{y}}(\cdot;\Theta_{\mathcal{E}_{\mathbf{y}}})$ and $\mathcal{E}_{\mathcal{N}}(\cdot;\Theta_{\mathcal{E}_{\mathcal{N}}})$, respectively. The features are subsequently integrated into $\mathcal{F}_{\mathbf{y}}(\cdot;\Theta_{\mathcal{F}_{\mathbf{y}}})$, yielding the reconstructed noisy bands $Y_k^{\mathcal{F}_{\mathbf{y}}}$.Likewise, $\mathcal{E}_{\mathbf{x}}(\cdot;\Theta_{\mathcal{E}_{\mathbf{x}}})$ take clean bands Y_k as input and generate semantic features $X_k^{\mathcal{E}_{\mathbf{x}}}$.The restored results $X_k^{\mathcal{F}_{\mathbf{y}}}$ are then obtained from $\mathcal{F}_{\mathbf{x}}(\cdot;\Theta_{\mathcal{F}_{\mathbf{x}}})$. Under the self-construction constraints, both autoencoders are able to extract semantic features and restore their respective input. To eliminate noise components from $Y_k^{\mathcal{E}_{\mathbf{y}}}$ while retaining as much noise as possible in $N_{Y_k}^{\mathcal{E}_{\mathcal{N}}}$, the cross-reconstruction constraint is introduced as below

$$\begin{aligned}&\mathcal{L}_{\mathrm{CR}}\left(\mathbf{X}_k^{\mathcal{F}_{\mathbf{y}}},\mathbf{Y}_k^{\mathcal{F}_{\mathbf{x}}}\right)\\ &\quad\left\|\mathcal{F}_{\mathbf{x}}\left(\mathcal{E}_{\mathbf{y}}\left(\mathbf{X}_k^{\mathcal{F}_{\mathbf{y}}};\Theta_{\mathcal{E}_{\mathbf{y}}}\right);\Theta_{\mathcal{F}_{\mathbf{x}}}\right)-\mathbf{Y}_k\right\|_1+\\ &=\frac{1}{n}\sum_{k=1}^{n}\left\|\mathcal{F}_{\mathbf{y}}\left(\begin{matrix}\mathcal{E}_{\mathbf{x}}\left(\mathbf{Y}_k^{\mathcal{F}_{\mathbf{x}}};\Theta_{\mathcal{E}_{\mathbf{x}}}\right)\\ +\mathcal{E}_{\mathcal{N}}\left(\mathbf{X}_k^{\mathcal{F}_{\mathbf{y}}};\Theta_{\mathcal{E}_{\mathcal{N}}}\right);\Theta_{\mathcal{F}_{\mathbf{y}}}\end{matrix}\right)-\mathbf{Y}_k\right\|_1\end{aligned} \tag{13}$$

where the cross restored images $X_k^{\mathcal{F}_{\mathbf{y}}} = \mathcal{F}_{\mathbf{y}}(X_k^{\mathcal{E}_{\mathbf{x}}} + N_{Y_k}^{\mathcal{E}_{\mathcal{N}}};\Theta_{\mathcal{F}_{\mathbf{y}}})$ are generated by incorporated $X_k^{\mathcal{E}_{\mathbf{x}}}$ and $N_{Y_k}^{\mathcal{E}_{\mathcal{N}}}$ into $\mathcal{F}_{\mathbf{y}}(\cdot;\Theta_{\mathcal{F}_{\mathbf{y}}})$. Subsequently, semantic features $\widehat{X}_k^{\mathcal{E}_{\mathbf{y}}}$ are derived from $\mathcal{E}_{\mathbf{y}}(\cdot;\Theta_{\mathcal{E}_{\mathbf{y}}})$ and fed into $\mathcal{F}_{\mathbf{x}}(\cdot;\Theta_{\mathcal{F}_{\mathbf{x}}})$, generating the final cross restored images $\widehat{X}_k^{\mathcal{F}_{\mathbf{x}}}$. Similarly, the cross restored images $Y_k^{\mathcal{F}_{\mathbf{x}}} = \mathcal{F}_{\mathbf{x}}(Y_k^{\mathcal{E}_{\mathbf{y}}};\Theta_{\mathcal{F}_{\mathbf{x}}})$ are obtained by inputting $Y_k^{\mathcal{E}_{\mathbf{y}}}$ into $\mathcal{F}_{\mathbf{x}}(\cdot;\Theta_{\mathcal{F}_{\mathbf{x}}})$. Afterwards, the noise attributes $\widehat{N}_{Y_k}^{\mathcal{E}_{\mathcal{N}}} = \mathcal{E}_{\mathcal{N}}(X_k^{\mathcal{F}_{\mathbf{y}}};\Theta_{\mathcal{E}_{\mathcal{N}}})$ and the semantic features $\widehat{Y}_k^{\mathcal{E}_{\mathbf{x}}} =$

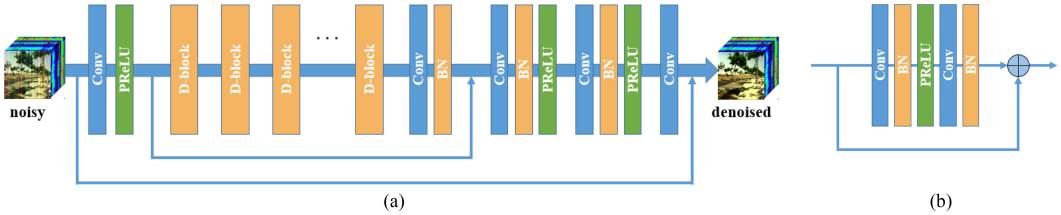


Fig. 3. Structure of the proposed spatial–spectral residual denoiser. (a) Spatial–spectral residual denoiser. (b) Denoising block (D-block).

$\mathcal{E}_{\mathbf{x}}(Y_k^{\mathcal{F}_{\mathbf{x}}}; \Theta_{\mathcal{E}_{\mathbf{x}}})$ are integrated into $\mathcal{F}_{\mathbf{y}}(\cdot; \Theta_{\mathcal{F}_{\mathbf{y}}})$, yielding the restored images $\widehat{Y}_k^{\mathcal{F}_{\mathbf{y}}}$.

The optimization process for the hyperspectral discrete representation learning scheme can be formulated as follows:

$$\widehat{\Theta}_{\mathcal{E}_*}, \widehat{\Theta}_{\mathcal{F}_*} = \underset{\Theta_{\mathcal{E}_*}, \Theta_{\mathcal{F}_*}}{\operatorname{argmin}} \ \lambda_3 * \mathcal{L}_{\mathrm{SR}}(\mathbf{X}_k, \mathbf{Y}_k) + \lambda_4 * \mathcal{L}_{\mathrm{CR}}\left(\mathbf{X}_k^{\mathcal{F}_{\mathbf{y}}}, \mathbf{Y}_k^{\mathcal{F}_{\mathbf{x}}}\right) \tag{14}$$

where $\Theta_{\mathcal{E}_*}$ and $\Theta_{\mathcal{F}_*}$ symbolize the trainable parameters associated with all encoders and decoders, respectively. λ_3 and λ_4 control the balance between two constraints.

C. Hyperspectral Denoising Network Architecture

The UALHDN framework consists of a spatial–spectral residual denoiser module, a global modeling discriminator module, and a hyperspectral discrete representation learning scheme. A residual block and skip connection are incorporated into the denoiser to retain the spatial–spectral information, while global pooling is employed in the discriminator for global contextual information modeling. Autoencoders are utilized in the hyperspectral discrete representation learning scheme. All modules are discussed in detail below.

1) Spatial–Spectral Residual Denoiser: Unlike in other applications (e.g., classification [59], anomaly detection [60], and recognition [61]), low-level features, such as the edge properties, are critical for HSI restoration. Therefore, a spatial–spectral residual denoiser is incorporated, based on the basic architecture of SRGAN [62], to fully utilize the low-level features in the network. As shown in Fig. 3(a), 12 D-blocks with identical layouts are designed to retain the low-level features and accelerate the convergence. Each D-block, illustrated in Fig. 3(b), comprises two convolutional layers followed by batch normalization layers, with a parametric ReLU (PReLU) activation function [63] succeeding the first batch normalization layer. The output of D-block is generated by adding the extracted features and the input. Moreover, skip connection with addition strategy is also employed in the denoiser to accelerate the convergence. To utilize the spectral correlation of HSIs, multiple bands (10 bands are chosen) are concatenated as the input of the denoiser. Specifically, each spectral band is treated as a channel in the convolutional layer to mine the intraband properties and the interband relevance. Finally, with the strategy of residual learning, the denoiser learns the noise pattern and outputs the denoising result. In the denoiser, the filter size and filter number are set to 3×3 and 64, respectively. The filter number of the last convolutional layer is set to be the same as the number of input bands.

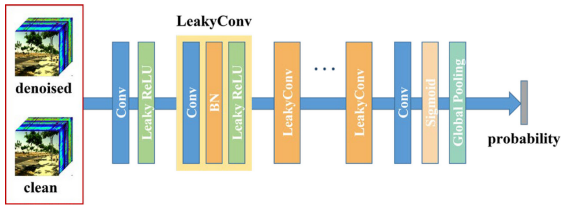


Fig. 4. Structure of the proposed global modeling discriminator.

2) Global Modeling Discriminator: Exploiting global contextual information is essential for learning image priors. Therefore, the global modeling discriminator is devised to fully extract the large-scale information. Therein, the LeakyConv module consists of multichannel 2-D convolution, batch normalization, and LeakyReLU [64]. As shown in Fig. 4, the discriminator is comprised of seven LeakyConv modules with small 3×3 kernels, whose number increases by a factor of 2 from 64 to 512. Following the last convolutional layer, global pooling is employed to model the global contextual information. Similar to the denoiser, multiple bands (10 bands are chosen) are concatenated as the input for exploiting the interband correlations. In general, the discriminator unites the spatial global contextual information and spectral correlation to establish whether the input is a clean and natural HSI or not.

3) Autoencoders for Discrete Representation: The structure of the encoders resembles that of the proposed discriminator. It differs by omitting the final sigmoid activation function and the global pooling layer. Skip connections are added between the input and output of the LeakyConv module. Additionally, the number of filters in each convolutional layer is set to 64. The filter number of the final layer is configured to match the quantity of input bands. The design of the decoders mirrors the architecture of the encoder in a symmetrical style.

IV. Experiments and Analysis

The experiments conducted in this study are presented in this section, including the multiplatform hyperspectral datasets, used in the experiments, the primary experimental settings, the denoising results for both synthetic noisy and real HSI denoising experiments, the ablation experiments with various modules, a effectiveness analysis of and l_1-norm and l_2-norm loss and a generalization analysis with homologous data.

TABLE I
MAIN PARAMETERS OF THE MULTIPLATFORM HYPERSPECTRAL DATASETS CONSIDERED IN THIS STUDY

Dataset	Platform	Sensor	Spectral range (nm)	Band number	Spectral resolution (nm)	Spatial resolution (m)	Noise intensity
Washington DC Mall	CV-580 Aircraft	HYDICE Airborne	400–2500	210	10	2.8	Noise-free
WHU-Hi-Baoxie	M600 Pro UAV	Nano-hyperspec	400–1000	270	6	0.12	Medium
SPARK	Spectral Micro-nano Satellite	SPARK Imager	420–1000	160	4.5	50	Severe
CRISM	Mars Reconnaissance Orbiter	CRISM	362–3920	544	6.55	18.4	Slight

A. Multiplatform Experimental Hyperspectral Datasets

To demonstrate the superiority of the proposed model, four multiplatform hyperspectral datasets were adopted in the experiments, that is, the Washington DC Mall airborne dataset, the WHU-Hi-Baoxie UAV dataset,[1] a Chinese SPARK hyperspectral satellite dataset, and a CRISM hyperspectral satellite dataset. These datasets differed in both the spatial and spectral resolutions, and were contaminated by hybrid noise with different intensities. Therein, the Washington DC Mall dataset, considered as a noise-free HSI, was adopted in the simulated experiment. The main parameters of these datasets are listed in Table I.

B. Experimental Settings

To test the effectiveness of the proposed model on nontraining data from a quantitative perspective, the Washington DC Mall dataset was not used in the pretraining stage. In addition, the paired training dataset was generated by simulating the multitype hybrid noise in real HSIs and adding this to 100 clean HSIs selected from the interdisciplinary computational vision laboratory (ICVL) [65]. Therein, the HSIs were processed into patches of size $40 \times 40 \times 10$ with a stride of 40. Moreover, the model was initialized according to [63]. During the training with 500 epochs, the Adam optimizer was employed with $\beta = (0.5, 0.999)$ and the learning rate was initialized to 0.0002. A mini-batch strategy of size 256 was adopted to accelerate the convergence. In the fine-tuning stage, every 10 continuous bands of the real HSI were split into the same size as the pretraining data with a stride of 40. The hyperspectral discrete representation learning scheme is first trained on the real HSI with 50 epochs. The noise intensity threshold for distinguishing between noisy and clean bands, was empirically set to 0.02. The mini-batch size was set to 64 and the learning rate was initialized as 0.0001. Afterwards, the denoiser was fine-tuned using the noise-free components estimated through discrete representation. The settings were similar to those of the discrete representation learning. Regularization parameters λ_1, λ_2, λ_3, and λ_4 were both empirically set to 0.001, 0.5, 1, and 1, respectively. In the final denoising process, the real HSIs are split into their original sizes with intervals of every 10 continuous bands.

[1] https://www.rsidea.whu.edu.cn

To demonstrate the superiority of the UALHDN framework, it was compared with several state-of-the-art hyperspectral denoising methods, that is, WFAF [5], spatial–spectral TV regularized local LRMR (LLRSSTV) [23], double-factor-regularized low-rank tensor factorization (LRTF-DFR) [26], nonlocal meets global (NGmeet) [30], HSI-DeNet [34], DHP [19], the 3-D quasi-recurrent hyperspectral denoising network (QRNN3D) [37], and a nonlocal spatial–spectral neural network (NSSNN) [38]. Notably, in order to ensure the fair comparison, the training datasets and the training settings for all the CNN-based models were consistent with the UALHDN framework. The codes of all the algorithms were provided by the authors, and the parameters were set by default or fine-tuned following the advice of the authors. The environment was MATLAB R2018a and Python 3.7 with the PyTorch package. The methods were implemented using Intel Xeon E5-2690 CPUs (2.60 GHz) and two NVIDIA Tesla P100 GPUs.

C. Noise Removal Experiments With Synthetic Hyperspectral Data

The Washington DC Mall dataset, measuring $1208 \times 307 \times 191$, is of high quality. A subimage with a size of $200 \times 200 \times 180$ was adopted as the simulated experimental data. To simulated real degradation in HSI imaging systems, noise simulation targeted the initial bands (bands 1–10, 400–500 nm) and those most susceptible to water vapor absorption (bands 91–120 and 1300–1600 nm, and bands 141–160 and 1800–2000 nm), incorporating three common noise types, that is, stripe noise, Gaussian noise, and impulse noise. Five synthetic cases were employed as follows:

Case 1: Stripe noise with varying noise intensities (with standard deviation $\sigma_S \in [0, 0.05]$).

Case 2: Based on case 1, the HSI was degraded by slight Gaussian noise $N(0, \sigma_{G_1}^2)$ ($\sigma_S \in [0, 0.05], \sigma_{G_1} \in [0, 0.025]$).

Case 3: Based on case 1, the HSI was degraded by Gaussian noise $N(0, \sigma_{G_2}^2)$ ($\sigma_S \in [0, 0.05], \sigma_{G_2} \in [0, 0.05]$).

Case 4: Based on case 2, the HSI was degraded by slight impulse noise ($\sigma_S \in [0, 0.05], \sigma_{G_1} \in [0, 0.025], P_1 \in [0, 0.025]$).

Case 5: Based on case 2, the HSI was degraded by impulse noise ($\sigma_S \in [0, 0.05], \sigma_{G_1} \in [0, 0.025], P_2 \in [0, 0.05]$).

Both a qualitative visual evaluation and a quantitative assessment were adopted in the experiments. Three common

TABLE II
QUANTITATIVE EVALUATION FOR THE SIMULATED NOISE REMOVAL

	Noise intensity	Index	Noisy	WFAF	LLRSSTV	LRTF-DFR	NGMeet	HSI-DeNet	DHP	QRNN3D	NSSNN	UALHDN
Case 1	$\sigma_S \in [0, 0.05]$	MPSNR	37.56	32.37	41.08	39.17	42.01	39.20	29.69	39.37	41.59	**46.03**
		MSSIM	0.9178	0.9455	0.9756	0.9824	0.9905	0.9821	0.8650	0.9826	0.9881	**0.9960**
		MSAM	5.843	6.685	3.435	3.670	2.999	4.481	6.216	3.741	3.495	**2.366**
Case 2	*Case 1* + $\sigma_{G_1} \in [0, 0.025]$	MPSNR	33.38	31.11	40.50	38.83	39.81	38.22	30.43	38.63	40.29	**41.91**
		MSSIM	0.8885	0.9120	0.9722	0.9808	0.9837	0.9781	0.8822	0.9795	0.9845	**0.9886**
		MSAM	6.644	7.501	3.639	3.773	3.426	4.584	5.886	3.856	3.610	**3.312**
Case 3	*Case 1* + $\sigma_{G_2} \in [0, 0.05]$	MPSNR	30.70	29.69	39.69	37.63	39.32	37.49	29.56	37.49	39.07	**39.92**
		MSSIM	0.8438	0.8637	0.9671	0.9757	0.9816	0.9741	0.8679	0.9736	0.9805	**0.9831**
		MSAM	7.808	8.618	3.884	3.925	3.610	4.658	6.409	4.062	3.701	**3.541**
Case 4	*Case 2* + $P_1 \in [0, 0.025]$	MPSNR	23.60	23.35	39.52	38.57	38.70	37.19	30.04	37.41	37.68	**39.73**
		MSSIM	0.7406	0.7538	0.9622	0.9785	0.9787	0.9724	0.8766	0.9752	0.9762	**0.9842**
		MSAM	11.526	12.229	4.154	3.859	3.959	4.813	6.275	4.171	4.429	**3.787**
Case 5	*Case 2* + $P_2 \in [0, 0.05]$	MPSNR	20.96	20.90	**39.12**	37.75	35.90	36.24	29.50	37.26	37.41	38.84
		MSSIM	0.6375	0.6448	0.9576	0.9767	0.9614	0.9660	0.8609	0.9720	0.9741	**0.9809**
		MSAM	15.326	15.859	4.382	4.089	4.893	5.094	6.772	4.561	4.481	**3.942**

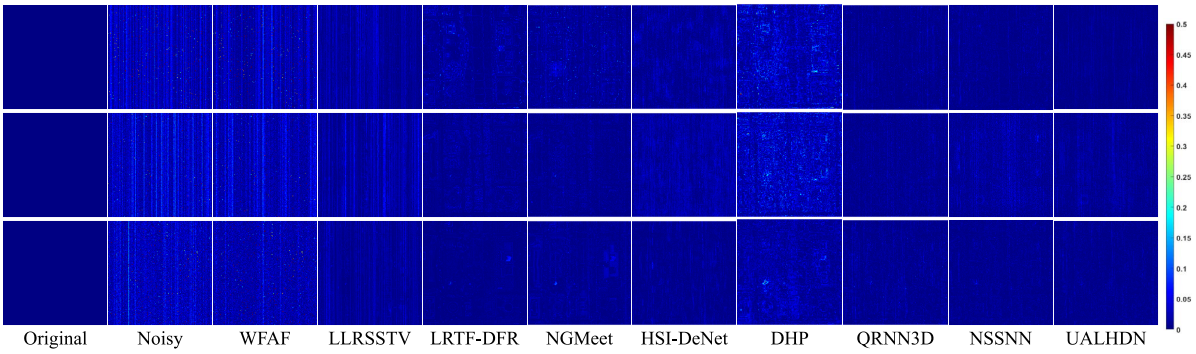


Fig. 5. Error maps of the synthetic noise removal results for the Washington DC Mall dataset under case 4 (band 5: 440 nm; band 106: 1450 nm; and band 160: 1990 nm, top to bottom) (original image, noisy image, WFAF, LLRSSTV, LRTF-DFR, NGMeet, HSI-DeNet, DHP, QRNN3D, NSSNN, and UALHDN, left to right).

metrics were employed in the quantitative assessment. The peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) indices [66] were adopted to assess the spatial fidelity of every band, and the mean values of all the bands are denoted as MPSNR and MSSIM, respectively. Moreover, the spectral angle mapper (SAM) was utilized to assess the spectral consistency. The average value for all the spatial pixels is denoted as MSAM and is displayed in angle. Generally speaking, higher-MSSIM and MPSNR values and a lower-MSAM value manifest a preferable result.

The quantitative assessments for all the synthetic cases are listed in Table II, where the best results are highlighted in bold, and the second-best results are underlined. It can be seen that the UALHDN framework obtains the best-denoising results under most circumstances.

From the aspect of a visual comparison, the error maps in case 4 and false-color images in case 5 with several typical bands are shown in Figs. 5 and 6. Fig. 6(b) shows the false-color image contaminated by hybrid noise. In the denoising result of WFAF, the stripe noise and the Gaussian noise are somewhat alleviated. However, WFAF is limited in the impulse noise removal and introduces tone distortion, which is obvious in Fig. 6(c). LLRSSTV and NSSNN remove the Gaussian noise and impulse noise, and suppress the stripe noise partly. Nevertheless, there are still some residual stripes in the images. For the denoising results of LRTF-DFR, NGMeet, HSI-DeNet, DHP, and QRNN3D, all the noise is removed, but at the cost of unexpected blurring. In comparison, the proposed UALHDN framework obtains a superior hybrid noise removal result with excellent spatial–spectral detail conservation.

D. Noise Removal Experiments With Real Hyperspectral Data

Three hyperspectral datasets acquired via diverse platforms, that is, UAV, Earth satellite, and Mars satellite platforms, were employed in the real HSI denoising experiments, to further demonstrate the superiority of the UALHDN framework on real HSIs. Due to the absence of reference images, both visual appearance and the no-reference HSI quality assessment metric [67] were employed to assess the denoising performance. Generally speaking, a lower score indicates a better-denoising result. The running time and the no-reference HSI quality assessment scores are detailed in Tables III and IV. It is worth noting that the time for the fine-tuning and discrete representation learning of UALHDN is not included in the running time but is shown in brackets. The reason for this is that only part of the real HSIs will be used in the UAL for homologous data in practical denoising tasks. More details are provided in the generalization analysis with homologous data.

1) WHU-Hi-Baoxie UAV Dataset: The first experiment was conducted on the WHU-Hi-Baoxie UAV dataset. A subimage

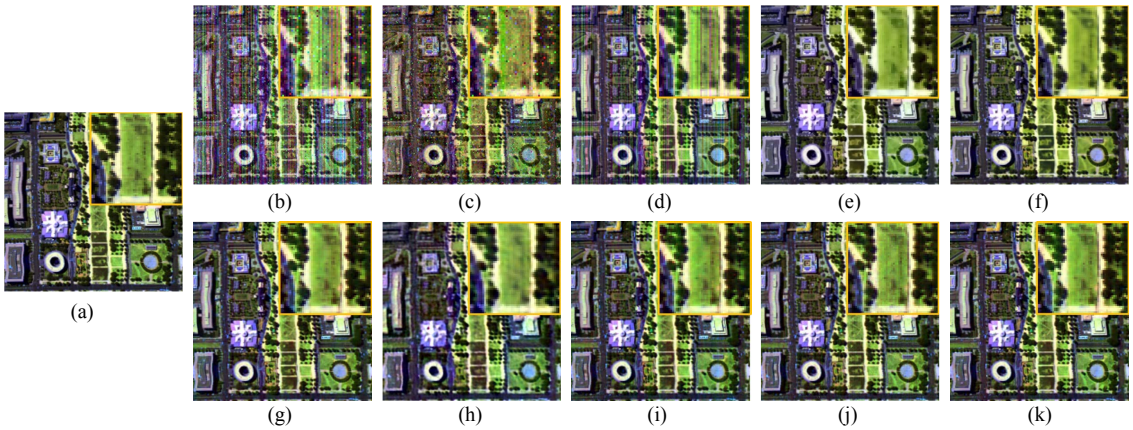


Fig. 6. Synthetic noise removal results for the Washington DC Mall dataset under case 5 (false-color image (R: band 160: 1990 nm; G: band 94: 1330 nm; and B: band 6: 450 nm)).

TABLE III
RUNNING TIME (SECONDS) FOR THE REAL HYPERSPECTRAL DATASET NOISE REMOVAL

Dataset	WFAF	LLRSSTV	LRTF-DFR	NGMeet	HSI-DeNet	DHP	QRNN3D	NSSNN	UALHDN
WHU-Hi-Baoxie	9.64	3472.43	291.22	146.90	3.63	2510.16	9.96	28.16	2.34 (1478.54)
SPARK HSI	1.13	47.16	30.17	35.98	0.32	140.22	2.29	5.85	0.19 (103.71)
CRISM HSI	17.59	3924.79	365.20	164.87	6.54	4410.09	18.43	48.76	3.56 (2192.26)

TABLE IV
NO-REFERENCE HSI QUALITY ASSESSMENT SCORES FOR THE REAL HYPERSPECTRAL DATASET DENOISING EXPERIMENTS

Dataset	Noisy	WFAF	LLRSSTV	LRTF-DFR	NGMeet	HSI-DeNet	DHP	QRNN3D	NSSNN	UALHDN
WHU-Hi-Baoxie	17.20	**15.91**	15.95	15.97	15.39	16.21	16.44	16.01	16.20	17.08
SPARK HSI	16.85	14.24	14.50	14.86	16.41	**13.41**	14.20	14.88	14.92	13.86
CRISM HSI	20.48	20.49	13.51	15.88	16.08	16.80	28.86	15.04	14.67	**13.02**

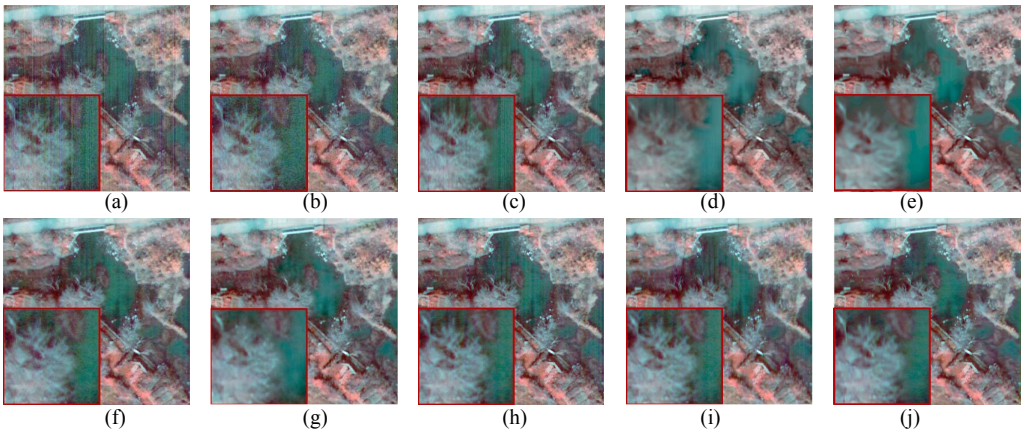


Fig. 7. Denoising results for the WHU-Hi-Baoxie UAV dataset (R: band 270: 1000 nm; G: band 15: 431 nm; and B: band 5: 408 nm). (a) Original image. (b) WFAF. (c) LLRSSTV. (d) LRTF-DFR. (e) NGMeet. (f) HSI-DeNet. (g) DHP. (h) QRNN3D. (i) NSSNN. (j) UALHDN.

of size 400 × 400 × 270 was employed in the experiment. Fig. 7 presents the denoising results obtained in multiple bands for visualization. For the denoising results of WFAF and HSI-DeNet, there is still residual random noise. The hybrid noise models of LLRSSTV are limited in destriping, which can be ascribed to the unconformity between the noise modeling in the regularization and the realistic noise distribution in real HSIs. There is also residual stripe noise in the result of QRNN3D and NSSNN. While LRTF-DFR, NGMeet, and DHP are effective in hybrid noise removal, they fail to retain the local details and tend to blur the terrain boundaries. In comparison, the proposed UALHDN framework acquires the best-denoising performance, further demonstrating the robustness and effectiveness of the proposed model for UAV datasets with both high spectral and spatial resolutions.

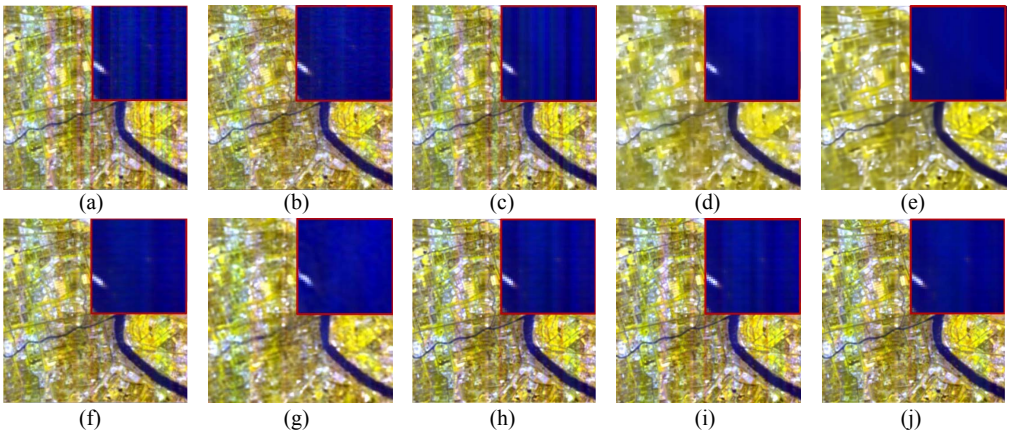


Fig. 8. Denoising results for the SPARK hyperspectral dataset (R: band 80: 539 nm; G: band 67: 518 nm; and B: band 30: 458 nm). (a) Original image. (b) WFAF. (c) LLRSSTV. (d) LRTF-DFR. (e) NGMeet. (f) HSI-DeNet. (g) DHP. (h) QRNN3D. (i) NSSNN. (j) UALHDN.

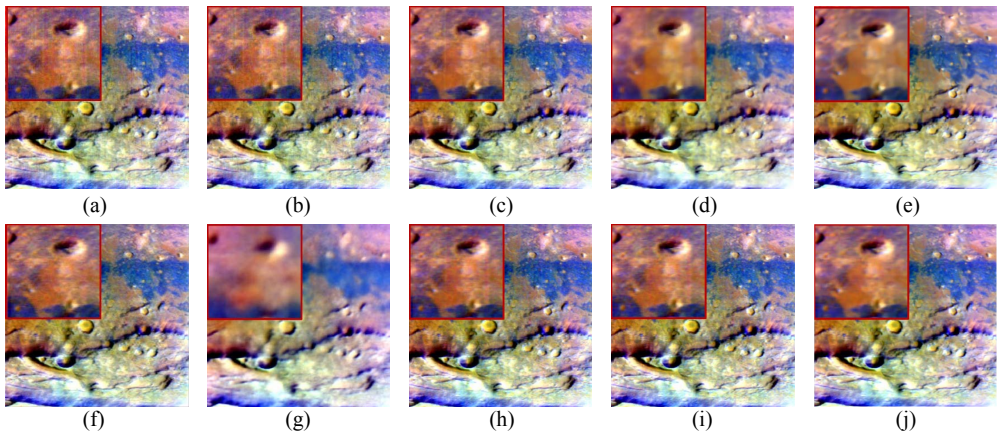


Fig. 9. Denoising results for the CRISM dataset (R: band 480: 3510 nm; G: band 440: 3244 nm; and B: band 10: 423 nm). (a) Original image. (b) WFAF. (c) LLRSSTV. (d) LRTF-DFR. (e) NGMeet. (f) HSI-DeNet. (g) DHP. (h) QRNN3D. (i) NSSNN. (j) UALHDN.

2) Chinese SPARK Hyperspectral Satellite Dataset: The second experiment adopted a subimage of size $200 \times 200 \times 80$ from the SPARK-02 satellite product. As a result of the corruption with severe stripe noise and random noise, the quality of the SPARK dataset is severely downgraded, as shown in Fig. 8(a). The WFAF method is adept at stripe noise removal, but it tends to become less effective in real hybrid noise removal. On the other hand, LLRSSTV, QRNN3D, and NSSNN can effectively suppress the random noise. However, there is severe residual stripe noise in the results. For the denoising results of LRTF-DFR, NGMeet, and DHP, even though almost all the noise is suppressed, severe spatial distortion can be observed in Fig. 8(d), (e), and (f). HSI-DeNet is effective in hybrid noise removal. Nevertheless, residual stripes and slight random noise are evident in the water region of the upper-right corner, as presented in Fig. 8(e). In contrast, UALHDN yields the best-denoising result since it suppresses almost all the noise while simultaneously retaining most of the spatial information.

3) CRISM Dataset: A subimage of size $400 \times 400 \times 490$ from the CRISM dataset was employed in the third real denoising experiment. The original calibrated images suffer from stripe noise and random noise, as presented in Fig. 9(a). The filtering-based WFAF method is less effective on the CRISM dataset. For the results of LLRSSTV, the random noise appears to be effectively suppressed. However, residual stripe noise is obvious in the results. While LRTF-DFR, NGMeet, and DHP remove all the noise, excessive blurring is introduced in the results. Although HSI-DeNet, QRNN3D, and NSSNN show good denoising performance, slight noise corruption can be still observed in the results. In contrast, the proposed UALHDN framework obtain convincing denoising performance with the CRISM dataset. Moreover, the UALHDN framework obtains excellent results in the no-reference HSI quality assessment scores, as shown in Table IV, indicating that it is practical for real applications.

E. Ablation Experiments With Various Modules

To demonstrate the effectiveness of UAL and the other modules, a comparative analysis was conducted using twelve models. These models were selected from various combinations of the following components: the discriminator (DIS), l_2-norm spatial-special consistency constrain (SSC-L2), the proposed l_1-norm spatial-special consistency constrain (SSC-L1), TV regularization, and the proposed BC. Therein, DIS, SSC-L2, and TV are the modules proposed in [14]. The performance of these models was evaluated on the datasets used in the denoising experiments. The quantitative results

TABLE V
EFFECTIVENESS ANALYSIS OF VARIOUS MODULES FOR THE WASHINGTON DC MALL DATASET WITH SYNTHETIC NOISE UNDER CASE 5

Module	1	2	3	4	5	6	7	8	9 [10]	10	11	12 (Proposed)
DIS [10]	×	×	×	×	√	√	√	√	√	√	√	√
SSC-L2 [10]	×	√	×	×	×	√	×	×	√	×	√	×
SSC-L1 (Proposed)	×	×	√	×	×	×	√	×	×	√	×	√
TV [10]	×	×	×	×	×	×	×	×	√	√	×	×
BC (Proposed)	×	×	×	√	×	×	×	√	×	×	√	√
MPSNR	35.46	37.21	37.96	37.37	36.25	36.98	38.51	38.51	36.52	38.45	38.52	**38.84**
MSSIM	0.9631	0.9750	0.9773	0.9736	0.9760	0.9776	0.9803	0.9799	0.9758	0.9794	0.9800	**0.9809**
MSAM	6.715	4.354	4.206	4.727	4.607	4.372	4.079	4.406	4.584	4.005	4.372	**3.942**

TABLE VI
EFFECTIVENESS ANALYSIS OF THE DESIGNED LOSS FUNCTIONS FOR THE NO-REFERENCE HSI QUALITY ASSESSMENT SCORES ON THE REAL DATASETS

Dataset	1	2	3	4	5	6	7	8	9	10	11	12
Baoxie	18.18	18.29	17.27	18.27	17.70	17.65	17.33	17.71	18.14	17.83	17.41	**17.08**
SPARK	13.48	13.49	13.50	**13.47**	13.89	13.88	13.95	13.90	13.91	13.91	13.88	13.86
CRISM	15.43	15.55	14.71	14.89	14.66	14.98	14.86	13.47	15.65	14.81	13.85	**13.02**

TABLE VII
EFFECTIVENESS OF L1 LOSS AND L2 LOSS IN DIFFERENT STAGES

Dataset	Index	L1	L2	L1/L1	L1/L2	L2/L2	L2/L1 (Proposed)
DC Case 5	MPSNR	36.90	36.25	38.26	37.89	38.52	**38.84**
	MSSIM	0.9749	0.9760	0.9791	0.9779	0.9800	**0.9809**
	MSAM	5.128	4.607	3.994	4.389	4.372	**3.942**

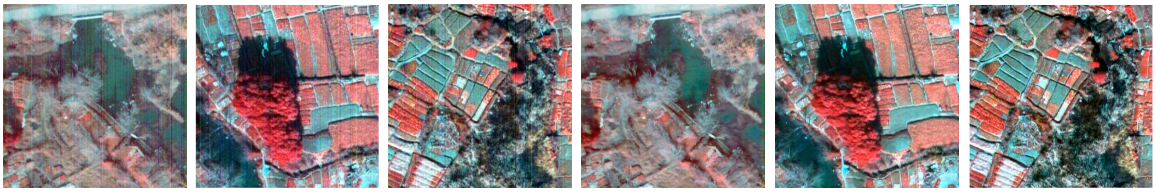
for case 5 of the Washington DC Mall dataset are listed in Table V. Improvement was observed across all three quantitative assessment indicators with either the spatial–spectral constraint or the BC. The proposed SSC-L1 achieves higher accuracy among them. In addition, better performance is obtained with the combinations of discriminator and SSC-L1/BC, while the combination with SSC-L2 has no obvious improvement. The full model proposed in [14] adopts the combination of DIS, SSC-L2, and TV, which yields relatively poor accuracy. The improvement of denoising performance is achieved after substituting SSC-L2 with the proposed SSC-L1 or swapping out TV with the proposed BC. The fully integrated proposed UALHDN framework, which combines DIS, SSC-L1, and BC, demonstrates superior accuracy in comparison to the other configurations. The no-reference HSI quality assessment scores obtained on all the real datasets are listed in Table VI, where the scores for the WHU-Hi-Baoxie dataset and the CRISM dataset are improved significantly with the DIS, SSC-L1, and BC. In contrast, the models employing SSC-L2 and TV show limited improvement in denoising performance. Therefore, this indicates that the proposed UAL, consisting of SSC-L1 and BC, outperforms the previous modules used for fine-tuning. It effectively improves the generalizability of the model on real HSIs.

F. Analysis of the Selection of l_1-Norm and l_2-Norm Loss

In this section, four strategies employing l_1-norm and l_2-norm loss in different stages, that is, l_1-norm form of the reconstruction term in the pretraining stage and l_1-norm form of the spatial–spectral consistency in the fine-tuning stage (l_1/l_2), l_1/l_2, l_2/l_2, and l_2/l_1, were adopted to perform denoising on case 5 of the Washington DC Mall dataset. As shown in Table VII, the model employing l_2-norm loss in the pretraining stage outperformed those using l_1-norm loss after the fine-tuning. In addition, all the models using the l_1-norm spatial–spectral consistency obtain better performances than the corresponding models employing the l_2-norm form. Compared with the other strategies, l_2/l_1 obtains the best-denoising result, which demonstrates the superiority of this strategy, and also provides guidance on how to combine l_2-norm and l_1-norm loss functions to improve the model performance in other HSI restoration tasks.

G. Generalization Analysis With Homologous Data

In this analysis, UAL was performed on a single noisy HSI, based on the UALHDN framework. To verify the generalizability of UALHDN, the model learned the image-specific deep prior on Image 1 from the WHU-Hi-Baoxie dataset and performed denoising without UAL on two subscenes obtained over different geographical locations. As presented in Fig. 10, the random noise and stripe noise can be effectively suppressed, and the terrain boundaries are clearly legible. This demonstrates that the UALHDN framework possesses good generalizability, in that the model used to perform UAL on a given noisy HSI can also obtain an excellent denoising performance without UAL on other HSIs acquired under similar circumstances via the same sensor. Therefore, to effectively improve the running time of the UALHDN model



(a) (b) (c) (d) (e) (f)

Fig. 10. Generalization analysis for the UALHDN framework on various subscenes of the WHU-Hi-Baoxie dataset (R: band 270; G: band 15; and B: band 5). (a) Image 1. (b) Image 2. (c) Image 3. (d) UALHDN (Image 1). (e) SSRD-DIS (Image 2). (f) SSRD-DIS (Image 3).

when performing denoising on homologous data, only some of the real HSIs need to be utilized in the fine-tuning stage.

V. Conclusion

In this article, an innovative UALHDN has been proposed for real multiplatform HSI denoising. A spatial–spectral consistency constraint and a BC are integrated into a transfer learning denoising framework to retain fidelity of background information and learn deep priors of real HSIs. A novel hyperspectral discrete representation learning scheme is designed in the fine-tuning stage to extract semantic features and estimate noise-free components, exploring the deep priors specific for real HSIs in an unsupervised manner. To our knowledge, the UALHDN framework is the first to explore the deep priors of real data based on discrete representation learning in multiplatform HSI noise removal. The extensive experiments conducted on the airborne Washington DC Mall dataset, the WHU-Hi-Baoxie UAV dataset, a Chinese SPARK hyperspectral satellite dataset, and a Martian dataset from CRISM confirmed the excellent denoising performance of the UALHDN framework. Moreover, the UALHDN framework represents a promising approach for other CNN-based image restoration methods to increase their generalizability on real HSIs. The limitation of UALHDN framework is that signal-dependent noise tends to be considered as signal components. In the future, new discrete representation learning methods should be explored to handle all types of noise corruption.

Acknowledgment

The authors would like to thank the authors of the WFAF, LLRSSTV, LRTF-DFR, NGMeet, HSI-DeNet, DHP, QRNN3D, and NSSNN methods for supplying the algorithm codes.

References

[1] Z. Liu, X. Wang, Y. Zhong, M. Shu, and C. Sun, "SiamHYPER: Learning a hyperspectral object tracker from an RGB-based tracker," *IEEE Trans. Image Process.*, vol. 31, pp. 7116–7129, 2022.

[2] M. Wang, Q. Wang, D. Hong, S. K. Roy, and J. Chanussot, "Learning tensor low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 679–691, Jan. 2023.

[3] P. Ghamisi et al., "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017, doi: 10.1109/MGRS.2017.2762087.

[4] Y. Zhong, W. Li, X. Wang, S. Jin, and L. Zhang, "Satellite-ground integrated destriping network: A new perspective for EO-1 Hyperion and Chinese hyperspectral satellite datasets," *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111416.

[5] R. Pande-Chhetri and A. Abd-Elrahman, "De-striping hyperspectral imagery using wavelet transform and adaptive frequency domain filtering," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 5, pp. 620–636, 2011.

[6] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2069, Jul. 2008.

[7] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[8] W. Xie and Y. Li, "Hyperspectral imagery denoising by deep learning with trainable nonlinearity function," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1963–1967, Nov. 2017.

[9] H. V. Nguyen, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral image denoising using SURE-based unsupervised convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3369–3382, Apr. 2021.

[10] L. Zhuang, M. K. Ng, L. Gao, J. Michalski, and Z. Wang, "Eigenimage2Eigenimage (E2E): A self-supervised deep learning network for hyperspectral image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 19, 2023, doi: 10.1109/TNNLS.2023.3293328.

[11] Q. Zhang, Q. Yuan, M. Song, H. Yu, and L. Zhang, "Cooperated spectral low-rankness prior and deep spatial prior for HSI unsupervised denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 6356–6368, 2022.

[12] Y. Chen, H. Zhang, Y. Wang, Y. Yang, and J. Wu, "Flex-DLD: Deep low-rank decomposition model with flexible priors for hyperspectral image denoising and restoration," *IEEE Trans. Image Process.*, vol. 33, pp. 1211–1226, 2024.

[13] G. Agresti, H. Schaefer, P. Sartor, and P. Zanuttigh, "Unsupervised domain adaptation for ToF data denoising with adversarial learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5584–5593.

[14] Y. Fu, T. Zhang, L. Wang, and H. Huang, "Coded hyperspectral image reconstruction using deep external and internal learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3404–3420, Jul. 2022.

[15] Y. Qian, H. Zhu, L. Chen, and J. Zhou, "Hyperspectral image restoration with self-supervised learning: A two-stage training approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5520917.

[16] J. Lehtinen et al., "Noise2Noise: Learning image restoration without clean data," 2018, *arXiv:1803.04189*.

[17] X. Wang, Z. Luo, W. Li, X. Hu, L. Zhang, and Y. Zhong, "A self-supervised denoising network for satellite-airborne-ground hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art no. 5503716, doi: 10.1109/TGRS.2021.3064429.

[18] C. Wang et al., "Translution-SNet: A semisupervised hyperspectral image stripe noise removal based on transformer and CNN," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5533114.

[19] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 1–8.

[20] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.

[21] J. Peng, Q. Xie, Q. Zhao, Y. Wang, L. Yee, and D. Meng, "Enhanced 3DTV regularization and its applications on HSI denoising and compressed sensing," *IEEE Trans. Image Process.*, vol. 29, pp. 7889–7903, 2020.

[22] J. Li, Q. Yuan, H. Shen, and L. Zhang, "Noise removal from hyperspectral image with joint spectral–spatial distributed sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5425–5439, Sep. 2016.

[23] W. He, H. Zhang, H. Shen, and L. Zhang, "Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 11, no. 3, pp. 713–729, Mar. 2018.

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

[25] T. Xie, S. Li, and J. Lai, "Adaptive rank and structured sparsity corrections for hyperspectral image restoration," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 8729–8740, Sep. 2022.

[26] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, Y. Chen, and W. He, "Double-factor-regularized low-rank tensor factorization for mixed noise removal in hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8450–8464, Dec. 2020.

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

[28] Y. Chang, L. Yan, X.-L. Zhao, H. Fang, Z. Zhang, and S. Zhong, "Weighted low-rank tensor recovery for hyperspectral image restoration," *IEEE Trans. Cybern.*, vol. 50, no. 11, pp. 4558–4572, Nov. 2020.

[29] Y. Su, H. Zhu, K.-C. Wong, Y. Chang, and X. Li, "Hyperspectral image denoising via weighted multidirectional low-rank tensor recovery," *IEEE Trans. Cybern.*, vol. 53, no. 5, pp. 2753–2766, May 2023.

[30] W. He et al., "Non-local meets global: An iterative paradigm for hyperspectral image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 2089–2107, Apr. 2022.

[31] H. Zhang, T.-Z. Huang, X.-L. Zhao, W. He, J. K. Choi, and Y.-B. Zheng, "Hyperspectral image denoising: Reconciling sparse and low-tensor-ring-rank priors in the transformed domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5502313.

[32] Z. Zha, B. Wen, X. Yuan, J. Zhou, C. Zhu, and A. C. Kot, "Low-rankness guided group sparse representation for image restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 7593–7607, Oct. 2023.

[33] L. Zhuang and M. K. Ng, "FastHyMix: Fast and parameter-free hyperspectral image mixed noise removal," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4702–4716, Aug. 2023.

[34] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "HSI-DeNet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, Feb. 2019.

[35] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, Feb. 2019, doi: 10.1109/TGRS.2018.2865197.

[36] Q. Shi, X. Tang, T. Yang, R. Liu, and L. Zhang, "Hyperspectral image denoising using a 3-D attention denoising network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10348–10363, Dec. 2021.

[37] K. Wei, Y. Fu, and H. Huang, "3-D quasi-recurrent neural network for hyperspectral image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 363–375, Jan. 2021.

[38] G. Fu, F. Xiong, J. Lu, J. Zhou, and Y. Qian, "Nonlocal spatial–spectral neural network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5541916.

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

[40] F. Xiong, J. Zhou, S. Tao, J. Lu, J. Zhou, and Y. Qian, "SMDS-net: Model guided spectral-spatial network for hyperspectral image denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 5469–5483, 2022.

[41] F. Xiong, J. Zhou, J. Zhou, J. Lu, and Y. Qian, "Deep parameterized neural networks for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5525015.

[42] L. Zhuang, M. K. Ng, L. Gao, and Z. Wang, "Eigen-CNN: Eigenimages plus eigennoise level maps guided network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5512018.

[43] M. Li, Y. Fu, and Y. Zhang, "Spatial–spectral transformer for hyperspectral image denoising," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 1368–1376.

[44] G. Fu, F. Xiong, J. Lu, J. Zhou, J. Zhou, and Y. Qian, "Hyperspectral image denoising via spatial–spectral recurrent transformer," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5511214.

[45] T. Zhang, Y. Fu, and C. Li, "Hyperspectral image denoising with realistic data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2248–2257.

[46] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Proc. 27th Int. Conf. Artif. Neural Netw.*, 2018, pp. 270–279.

[47] Y. Kim, J. W. Soh, G. Y. Park, and N. I. Cho, "Transfer learning from synthetic to real-noise denoising with adaptive instance normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3482–3492.

[48] H. Chang, J. Han, C. Zhong, A. M. Snijders, and J.-H. Mao, "Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1182–1194, May 2018.

[49] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.

[50] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[51] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3722–3731.

[52] B. Lu, J.-C. Chen, and R. Chellappa, "Unsupervised domain-specific deblurring via disentangled representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10225–10234.

[53] G. Wang, C. Sun, X. Xu, J. Li, Z. Wang, and Z. Ma, "Disentangled representation learning and enhancement network for single image de-raining," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 3015–3023.

[54] H. Xu, X. Wang, and J. Ma, "DRF: Disentangled representation for visible and infrared image fusion," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, Feb. 2021.

[55] W. Du, H. Chen, and H. Yang, "Learning invariant representation for unsupervised image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 14483–14492.

[56] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[57] X. Jin, Z. Chen, J. Lin, Z. Chen, and W. Zhou, "Unsupervised single image deraining with self-supervised constraints," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2019, pp. 2761–2765.

[58] G. Chen, F. Zhu, and P. A. Heng, "An efficient statistical method for image noise level estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 477–485.

[59] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, "Automatically designing CNN architectures using the genetic algorithm for image classification," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3840–3854, Sep. 2020.

[60] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of gaussian for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4363–4372, Sep. 2021.

[61] F. Wu et al., "Intraspectrum discrimination and interspectrum correlation analysis deep network for multispectral face recognition," *IEEE Trans. Cybern.*, vol. 50, no. 3, pp. 1009–1022, Mar. 2020.

[62] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4681–4690.

[63] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.

[64] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2016, pp. 1–16.

[65] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. 14th Eur. Conf. Comput. Vis.*, 2016, pp. 19–34.

[66] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, pp. 600–612, 2004.

[67] J. Yang, Y. Zhao, C. Yi, and J. C.-W. Chan, "No-reference hyperspectral image quality assessment via quality-sensitive features learning," *Remote Sens.*, vol. 9, no. 4, p. 305, 2017.