An End-to-End Framework for Joint Denoising and Classification of Hyperspectral Images

Xian Li, *Member, IEEE*, Mingli Ding, Yanfeng Gu, *Senior Member, IEEE*, and Aleksandra Pižurica, *Senior Member, IEEE*

Abstract— Image denoising and classification are typically conducted separately and sequentially according to their respective objectives. In such a setup, where the two tasks are decoupled, the denoising operation does not optimally serve the classification task and sometimes even deteriorates it. We introduce here a unified deep learning framework for joint denoising and classification of high-dimensional images, and we particularly apply it in the framework of hyperspectral imaging. Earlier works on joint image denoising and classification are very scarce, and to the best of our knowledge, no deep learning models were proposed or studied yet for this type of multitask image processing. A key component in our joint learning model is a compound loss function, designed in such a way that the denoising and classification operations benefit each other iteratively during the learning process. Hyperspectral images (HSIs) are particularly challenging for both denoising and classification due to their high dimensionality and varying noise statistics across the bands. We argue that a well-designed end-to-end deep learning framework for joint denoising and classification is superior to current deep learning approaches for processing HSI data, and we substantiate this by results on real HSI images in remote sensing. We experimentally show that the proposed joint learning framework substantially improves the classification performance compared to the common deep learning approaches in HSI processing, and as a by-product, the denoising results are enhanced as well, especially in terms of the semantic content, benefiting from the classification.

Index Terms— Classification, convolutional neural network (CNN), deep learning, hyperspectral image (HSI), image denoising.

I. Introduction

RAPID progress of deep learning is pushing boundaries in Earth observation as well, with tremendous improvements in processing and interpretation of remote sensing imagery [1], [2]. In this processing pipeline, noise reduction is often employed as a preprocessing step for the subsequent classification for scene interpretation. The two tasks are typically addressed separately. Here, we take interest in the interplay between noise reduction and classification, and investigate a new perspective for jointly optimizing the denoising and classification problems in an end-to-end deep learning framework.

Hyperspectral image (HSI) acquisition has become a vital part of many Earth observation systems, among others supporting the study of ecosystem service [3], anomaly detection [4], disaster management [5], and monitoring the effects of climate change [6]. While providing rich spectral information and, thereby, unique capabilities in distinguishing between different compositions of land covers, HSI data also pose huge challenges for processing [7], [8], [9], [10], [11], [12] due to their very high dimensionality (several hundreds of spectral bands). This huge dimensionality, together with the lack of sufficient labeled data, requires also special designs of deep learning models [13], [14], [15], [16], [17], [18].

In a common HSI processing framework, denoising and classification are performed separately. In such a setup, denoising alone is likely to smooth out some semantic features that are important for classification, leading to suboptimal classification performance. Alternatively, applying classification on noisy data directly yields often unstable classification performance since deep neural networks tend to be very sensitive to noise. Even small amounts of noise can lead to unexpected misclassification [19]. In this article, we propose a unified deep learning framework with a compound loss function for simultaneous denoising and classification. In this way, the classifier helps the denoising network to preserve semantic details, while the denoiser in parallel provides a better input to the classifier to enable its more robust performance.

Few works have studied joint denoising and classification in general so far. The most related to our work are [20] and [21]. In [20], a hybrid conditional random field and multinomial logistic regression model was proposed for joint HSI denoising and classification. Recent work in [21] cascades a deep denoiser and a well-trained classifier into a hybrid model to explore the effect of semantic information on RGB image denoising. However, the problem of end-to-end learning from scratch for joint noise reduction and classification is still not studied in a deep learning framework, not only in the domain of HSI analysis. Three major open questions have yet to be addressed.

1) Although current deep denoisers [21], [22], [23], [24], [25], [26] can be trained with various noisy settings, the actual noise distribution in the image at hand may differ significantly from the examples that were included in the training, which deteriorates the denoising results [27], [28], [29]. To what extent can a feedback loop supplying information about semantic details based on classification improve this performance is not known yet.

2) Deep denoisers often operate in a band-by-band fashion to cope with the disparate dimensionality of the data, e.g., in remote sensing [23], [24], [25], [26]. This makes their direct cascading with the classifier that relies largely on complete spectral information less appropriate. How to design a deep denoiser that can process simultaneously an arbitrary number of spectral bands and cascade with the classifier directly is not evident.

3) Although the work in [21] studied the benefit of image semantics for both denoising and classification, it uses a well-trained classifier to train the denoiser. How to train the denoiser and classifier jointly from scratch was not studied yet.

Motivated by these research questions, we develop a fundamental deep learning framework for joint denoising and classification. In this framework, we design an all-band denoising network, which can be applied to arbitrary HSIs with disparate spectral dimensionality and enables it to connect with the classifier in an end-to-end fashion. This denoising network is connected to a spectral classification network, which captures discriminative spectral features and reduces sharply the high spectral dimensionality by processing jointly groups of spectral bands. We formulate a joint learning scheme, which enables denoising and classification to benefit each other. In particular, we define a novel compound loss function to train the proposed framework from scratch. Iteratively, the denoiser outputs directly the denoised result to the classifier for facilitating classification, and the classifier provides class information to the denoiser, which helps in preserving semantic details. This way, the denoising and classification processes improve their performance in parallel. We define the compound loss function to adapt well to the actual, nonuniform noise distribution across the bands. To train the denoiser, we use a pseudoreference as an alternative. Our framework proves to be effective to improve the denoising result for different qualities of pseudoreferences and adapts well to various noise types.

The main contributions of this article are given as follows.

1) We propose a unified deep learning framework for joint HSI denoising and classification, where the two tasks are trained from scratch with a compound loss function. To the best of our knowledge, this is the first attempt to study end-to-end learning from scratch for joint denoising and classification in a deep learning framework in general, not only in HSI analysis.

2) We demonstrate that the denoising and classification operations benefit each other, yielding better denoising results and robust classification performance simultaneously both on various types of simulated noise and real noisy images.

3) We define an effective compound loss function, which incorporates a nonuniform denoising loss component based on a logarithmic function. This novel loss function takes into account the nonlinear characteristic of spectra and nonuniform noise statistics across the bands.

The rest of this article is organized as follows. Section II reviews related work on the interaction between denoising and classification. The proposed method is presented in detail in Section III. Section IV evaluates the effectiveness of the proposed approach on simulated and real HSIs. Section V draws the conclusion of this work.

II. RELATED WORK

HSIs in practice inevitably suffer from various degradations during the collection and transmission processes. These adverse effects limit the performance of the subsequent processing tasks [19]. Therefore, denoising is often applied to HSI data prior to image analysis steps, such as classification and segmentation. HSI denoising approaches often employ spectral low-rank property [30] together with spatial priors in the form of total variation [31] and spatial nonlocal similarity [32], [33]. Recent denoising methods based on deep learning often employ convolutional neural network (CNN) architectures [23], [24], and [25]. A thorough analysis in [34] showed that denoising indeed improves significantly the classification performance on multichannel remote sensing images. Although they performed concrete experiments on three-channel images, the main conclusions of the study are valid for more general multichannel data. Lukin et al. [34] observed that the classification performance improves when the denoising is more effective but also that the true benefit of the denoising step is higher when the classifier is trained on noise-free or denoised images.

We are interested in a different approach, assuming that we do not have access to noise-free data, and we want to optimize the denoising and classification processes jointly. Little work has been done in this direction, even with traditional approaches, with exception of a reported work [20], and none within the deep learning context yet. The closest work to ours is [20], where the authors jointly trained a conditional random field denoiser and a multinomial logistic regression classifier. They showed that the denoising step can reduce unexpected misclassification caused by noise, while the class information, in turn, benefits from denoising. As opposed to [20], which extracts handcrafted features via the shallow structure, we aim to develop a unified framework with deep network structures to learn deep features from HSIs.

Deep learning methods for HSI denoising, including [23], [24], and [25], build on the ideas originally developed for traditional grayscale or color image denoising [22]. Commonly, a deep learning model for image denoising is trained on pairs of ideal, noise-free image patches and their noisy versions, with simulated noise. Applying this approach directly to HSI denoising is difficult due to the huge and disparate dimensionality of the data. Hence, current deep denoisers for
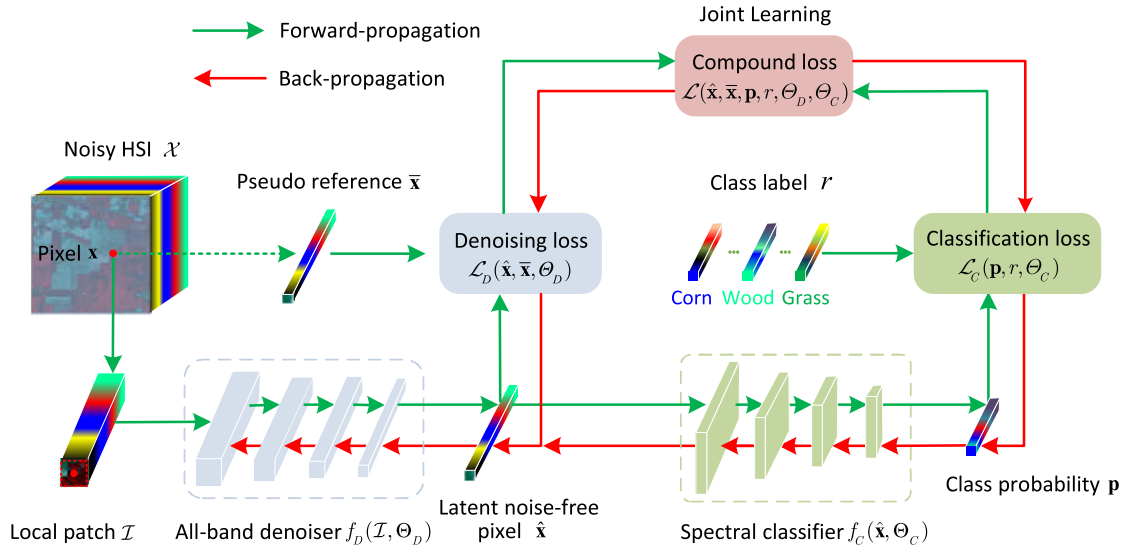
Fig. 1. Overall framework of the proposed IDCN. It comprises an all-band denoiser, a spectral classifier, and a compound loss that connects two stages in a unified framework. The pseudoreference can be derived from the noisy HSI directly or by employing the existing available denoisers.

HSI often operate in a band-by-band fashion, taking the noisy band together with a predefined number of adjacent bands as the inputs to spectral–spatial denoising. These CNN-based denoisers perform well when trained on the same type of signal structures and similar levels of noise as in the test data. However, the actual noise distribution and the type of signal structures are often unknown and may differ significantly from the examples that were included in the training, which deteriorates the denoising performance [27], [28], [29]. Here, we shall exploit the classification to improve denoising by supplying information about semantic details.

Another major problem for the subsequent classification task is that current denoising methods are often optimized in terms of mean squared error and visual perception metrics rather than for the classification performance. Although some works, including [24], [25], and [31], have reported their classification results on classical classifiers [such as support vector machine (SVM)], the denoising and classification were performed independently, each optimized according to their individual objectives. In contrast, the proposed method develops a unified deep framework, where these two processing stages are optimized jointly.

III. METHODOLOGY

Let $\mathcal{H} \in \mathbb{R}^{M \times N \times B}$ denote a 3-D hyperspectral cube with the spatial size of $M \times N$ and with B spectral bands. Suppose that T pixels are labeled in $\mathcal{H}$. Denote by $\mathcal{T} = \{\mathcal{I}_i, r_i\}_{i=1}^{T}$ the set of labeled input–output pairs, where $\mathcal{I}_i \in \mathbb{R}^{P \times P \times B}$ is the local patch with window size $P \times P$ centered at pixel $\mathbf{x}_i \in \mathbb{R}^B$, and r_i denotes the class label from the set $\mathcal{C} = \{1, \ldots, C\}$, where C is the number of classes. We assume the following additive noise model $\mathcal{H} = \mathcal{X} + \mathcal{N}$, where $\{\mathcal{X}, \mathcal{N}\} \in \mathbb{R}^{M \times N \times B}$ are the ideal noise-free data and noise, respectively. Most existing deep denoising models [23], [24], [25], [26] are trained on pairs of noisy and ideal image patches from $\mathcal{H}$ and $\mathcal{X}$, respectively. However, these trained denoisers often show limitations in generalizing to challenging cases since the actual noise is often unknown and may differ significantly from the predefined $\mathcal{N}$.

A. Overall Framework and Joint Learning

Here, we propose a unified framework based on CNN to study the interplay between the denoising and classification of HSIs. In this framework, the denoiser and the spectral classifier are combined into a unified network and trained in an end-to-end manner with a compound loss function. We refer to this integrated denoising and classification network as IDCN. The overall framework of the proposed IDCN is shown in Fig. 1. Specifically, given a local patch $\mathcal{I}$ centered at pixel $\mathbf{x}$, the all-band denoiser f_D with trainable parameters Θ_D outputs a latent noise-free pixel $\hat{x} = f_D(\mathcal{I}, \Theta_D)$, which is directly fed into the subsequent spectral classifier f_C with trainable parameters Θ_C to estimate the class probability $\mathbf{p} = f_C(\hat{\mathbf{x}}, \Theta_C)$ for final classification. In this way, the proposed framework can be formulated as an end-to-end deep network learning problem, which is trained with the compound loss $\mathcal{L}$ and supervised by the pseudoreference $\bar{\mathbf{x}}$ and the class label r via the back-propagation algorithm. The approach of generating $\bar{\mathbf{x}}$ will be given in Section III-B2.

1) Joint Learning: Having constructed the proposed framework, we need to optimize it. As opposed to the common denoising and classification methods, which train the two tasks independently with their individual loss functions, we propose a joint learning approach to establish the interplay between the two tasks. With an appropriate compound loss, we shall enable end-to-end training from scratch. This way, the denoiser and classifier will iteratively interact with each other.

Specifically, in the forward-propagation stage, the denoiser makes use of full-band spectral–spatial information and provides the denoised data with full spectral bands to the subsequent classifier directly, where a better denoised result yields a more accurate classification result. In the back-propagation stage, the gradients of the compound loss are propagated back to update simultaneously the classifier and denoiser, and the

gradients of the classifier (as the intermediate gradients of the whole network) are also used to further update the denoiser, where better classification result provides more accurate class guidance to the denoiser. Thus, the established interplay between denoising and classification, as illustrated in Fig. 1, benefits both denoising and classification performance.

2) Compound Loss: Now, we turn to define the compound loss function explicitly. For the classification task, we adopt the common cross-entropy loss. For the denoising task, current works in HSI denoising, including [23], [24], [25], and [26], typically use mean square error (mse) same as in RGB image denoising. The mse treats all the bands uniformly. However, the noise level in HSI images differs strongly from one band to another due to the nonlinear spectral reflectance [28], and thus, a nonuniform loss function is more appropriate. Inspired by the cross-entropy loss, we also adopt the logarithmic function to formulate the denoising loss nonuniformly. We, thus, propose a logarithmic absolute error (LAE) loss function as

$$\mathcal{L}_D = -\frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{B} \log(1 - \phi(\Delta_{i,j})) \quad (1)$$

where $\Delta_i = |\hat{x}_i - \overline{\mathbf{x}}_i|$ is the absolute error between the denoised pixel $\hat{x}_i \in \mathbb{R}^B$ and its pseudoreference $\overline{\mathbf{x}}_i \in \mathbb{R}^B$ for the ith training data. $\Delta_{i,j}$ indicates the error in the jth spectral band. ϕ is a clipping function, ensuring that the logarithm is well-defined, i.e., $\phi(x) \in [0, 1 - \epsilon)$ for $\forall x$, where ϵ is a small positive constant. The reason for adopting such a logarithm loss function is to put more emphasis on suppressing the noisier bands since it more heavily penalizes an increased absolute error compared to traditional mse, as shown in Fig. 2(a). One clear advantage of our loss is that it can produce a stable denoising result over entire bands, as illustrated in Fig. 2(b). More evidence will be given in Section IV-D.

In machine learning problems involving multicomponent objectives, the loss function is often defined as a weighted sum of the losses associated with each objective [24], [35]. The weighting factors are then predefined or tuned on a case-by-case basis. To avoid manually tuning the loss factors, we shall automatically determine the contributions of the loss components based on the model uncertainty [36]. We define the compound loss function

$$\mathcal{L} = \mathcal{L}_C + \frac{1}{v_D}\mathcal{L}_D + \log v_D \quad (2)$$

where $\mathcal{L}_C$ denotes the classification loss. v_D is the noise variance representing the uncertainty of the denoising model. The last term $\log v_D$ acts as a regularizer to avoid overfitting the overall objective to the denoising aspect alone because, when v_D becomes very small, the item $\mathcal{L}_D/v_D$ becomes dominant. During the training process, $\log v_D$ is set to a learning parameter (like Θ_D and Θ_C) and updates automatically. In particular, we set $w = \log v_D$, where w is the learning parameter, and calculate v_D as $\exp(w)$. This is to avoid any division by zero for $\mathcal{L}_D/v_D$ and squash v_D to the positive domain. Unlike [36], where dynamic weights are assigned to each component of the loss functions, we use only one parameter. We experimentally find that this makes the output noise efficiently suppressed.
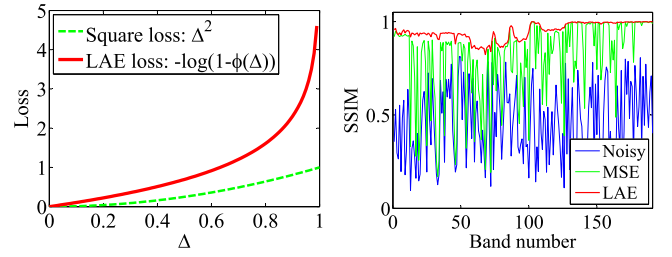

Fig. 2. Comparison between mse and the proposed LAE loss. Left: square loss and the proposed LAE loss as a function of the absolute error Δ. Right: denoising results across the bands of Washington DC with mixture noise using mse and LAE.

The whole optimization procedure of the proposed method is summarized in Algorithm 1.

B. All-Band Denoising

While deep denoisers [24], [25], [26] typically operate in a band-by-band fashion, which extracts local spectral correlations and limits their direct connecting with classification, we, thus, propose an all-band denoising method that makes use of spectral information and improves its performance via the information interplay with classification.

Algorithm 1 End-to-End Framework for Joint HSI Denoising and Classification

Input: HSI $\mathcal{X}$, training set $\mathcal{T}$, the pseudo reference $\overline{\mathcal{X}}$.

while *Model is convergent or maximum iteration is met* **do**

  1 Forward-propagation

  (a) Extract training pairs $\{\mathcal{I}_i, \overline{\mathbf{x}}_i, r_i\}_{i=1}^{T}$;

  (b) Random initialize $\log v_D$, Θ_D and Θ_C;

  (c) Calculate denoised pixels $\left\{\hat{x}_i \leftarrow f_D(\mathcal{I}_i, \Theta_D)\right\}_{i=1}^{T}$ and class probabilities $\left\{\mathbf{p}_i \leftarrow f_C(\hat{\mathbf{x}}_i, \Theta_C)\right\}_{i=1}^{T}$;

  2 Back-propagation

  (a) Calculate the compound loss based on Eq. (2) and the gradients of the whole network;

  (b) Update $\log v_D$, Θ_D and Θ_C by minimizing Eq. (2) using mini-batch Adadelta [37];

end

Output: Estimate the denoised image $\hat{X} = \left\{\hat{x}_i\right\}_{i=1}^{MN}$ and the labels $\mathbf{r} = \{r_i\}_{i=T+1}^{MN}$ of unlabelled pixels.

Specifically, the proposed denoiser consumes full spectral information and learns global nonlinear spectral correlations. Another advantage is that it is able to connect with the classifier directly. Iteratively, with the proposed joint learning, our denoiser not only offers a better input to the classifier for facilitating discrimination (see the green lines in Fig. 1) but also receives the class guidances from the compound loss and the classifier (see the red lines in Fig. 1) to reproduce sufficient class details via the back-propagation. Moreover, we avoid problems originating from the model mismatch between training and the actual data that are inherent to the models trained with simulated noise [23], [24], [25], and [26]. This approach is important in its own right as it can be applied in combination with other classification networks too and is

especially of interest when the data are heavily contaminated by noise.

1) Denoising Network: Our denoiser takes at its input a raw 3-D data cube with an arbitrary number of spectral bands. The output should match the input of the subsequent classifier, which is a spectral vector. We, thus, develop a spatial subsampling denoising network based on 2-D-CNN, which consists of several 2-D convolutional layers following a global max pooling layer. The role of the convolutional layers is to learn the spectral–spatial features and reduce the size of the feature maps. The global max pooling ensures that the output of our denoising model is a spectral vector. This allows our denoiser to input a raw 3-D cube with an arbitrary spatial size. Different from current deep denoisers [22], [24], [25], [26], [38], we use the activation in the last convolutional layer, which leads to an improved denoising and classification performance. We make the observation that ReLU [39] is likely to smooth out some spectral information by zeroing some small input reflectance values. We, thus, choose sigmoid to squash the denoised results to the range 0–1 and preserve the detail spectral information. The main network hyperparameters are given in Table I.

2) Pseudoreference: Having developed the denoising network, we need an appropriate reference to train it. Since the noise-free reference is not available in practice, we propose an effective solution that uses a pseudoreference as an alternative. A key question is how the quality of the pseudoreference affects the final denoising and classification performance of the proposed framework. By testing different qualities of pseudoreferences (including the input noisy image and the denoised results from the existing denoisers) both on various types of simulated noise and real noisy images, our framework turns out to be able to enhance the denoising results (particularly in terms of the semantic content) and yield stable classification accuracies simultaneously. Therefore, the pseudoreference in our framework is flexible to derive from the input noisy image directly or the existing available denoisers. More evidence can be found in Sections IV-B and IV-C.

C. Group Spectral Classification

Since the denoising network already takes care of the noise-induced fluctuations, we design a low-complexity classifier network based on the spectral features alone. In order to effectively extract discriminative features from all the available spectral bands, we process jointly groups of bands, using the group convolution concept [40], [41], [42]. The main idea is to first apply group convolution over the spectral dimension to reduce its huge dimensionality and then extend this operation into a multiscale representation. Specifically, on each scale, we first split the input spectral vector into several groups of subbands and apply 1-D convolutions over each group in parallel. This way, the group convolutions reduce the spectral dimensionality sharply and extract discriminative spectral features by considering the differences among different groups of subbands.

We accomplish the multiscale representation with different kernel sizes, which processes spectral information at multiple
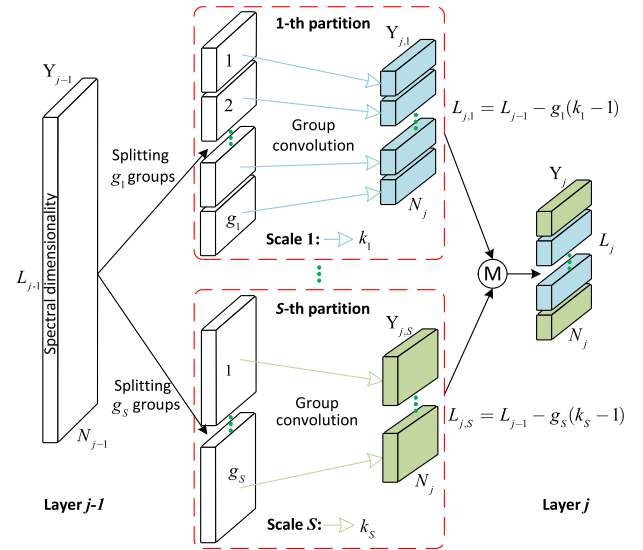

Fig. 3. Illustration of the proposed group spectral feature extraction. The number of spectral dimensionalities reduces progressively from L_{j-1} to L_j, with group convolutions. $\boxed{M}$ denotes the elementwise maximum operation.

scales in parallel. We construct the multiscale representation of the input spectral bands as S partitions in parallel, where each of these partitions corresponds to a scale indexed by $s \in \{1, \ldots, S\}$, as illustrated in Fig. 3. For simplicity, let the kernel size and the number of groups in each partition be the same for all the layers, let the sth partition at the layer j consist of g_s groups of subbands, and let k_s denote the kernel size that convolves these groups separately. We require that

$$A_j = g_1(k_1 - 1) = \cdots = g_S(k_S - 1) \qquad (3)$$

where A_j is the reduced spectral dimensionality from layer $(j-1)$ to j. This requirement ensures the same reduced spectral dimensionality at each scale such that we can merge all the extracted multiscale features. Based on (3), the output spectral dimensionality at layer j is reduced to

$$L_j = L_{j-1} - A_j. \qquad (4)$$

Observe how our approach reduces effectively the spectral dimensionality of the problem. The group convolution with a partition at scale s is g_s times faster than a regular convolution with the same kernel size, i.e., $g_s(k_s-1)$ versus (k_s-1). At the same time, we have a richer multiscale representation. This is an important asset of the proposed method compared to most of the earlier reported ones, including [43] and [44].

To fuse the multiscale features, we employ elementwise maximum operation to identify informative features at multiple scales. This avoids increasing the output dimension of each layer, reducing thereby the number of parameters. Considering the tradeoff between classification performance and the input spectral dimensionality of the data, we exploit three group convolution layers and adopt three scales at each layer in our experiments, as reported in Table I.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we evaluate the proposed IDCN and its two reduced versions, with denoising turned off (I-CN) and with

TABLE I

PROPOSED NETWORK ARCHITECTURE FOR ALL THE DATASETS. IN TYPE COLUMN, 2-D-CON, 1-D-GCON, AND $\mathcal{M}$ DENOTE THE 2-D CONVOLUTION, 1-D GROUP CONVOLUTION, AND GLOBAL MAX POOLING OPERATIONS, RESPECTIVELY. KERNEL DENOTES THE PRODUCT BETWEEN THE KERNEL SIZE AND THE NUMBER OF CHANNELS. B IS THE NUMBER OF BANDS. GROUP REFERS TO THE NUMBER OF GROUPS

Network	Layer	Type	Kernel	Group	Activation
Denoising	1		Input shape: $7 \times 7 \times B$		
	2	2D-Conv	$1 \times 1 \times 288$	1	ReLU
	3-5	2D-Conv	$3 \times 3 \times 288$	1	ReLU
	6	2D-Conv	$1 \times 1 \times B$	1	Sigmoid
	7	$\mathcal{M}$	–	–	–
Classification	1		Input shape: $B \times 1$		
	2-4	1D-GConv (three scales)	3×128	12	ReLU
			7×128	4	ReLU
			9×128	3	ReLU
	5	$\mathcal{M}$	–	–	–

classification turned off (ID-N), and compare these to some state-of-the-art denoising and classification methods.

A. Datasets and Settings

1) Datasets: We perform experiments both on data with simulated noise and real noise. The gray values in each dataset are normalized to the range [0, 1]. To analyze various noise scenarios in a simulated setting, we take am HSI image, which is nearly noise-free: Washington DC Mall. This HSI image, collected by the HYDICE sensor, is commonly used to simulate degraded data with different kinds of noises [24], [25], [28], [45]. It contains 280×307 pixels with six classes and 210 spectral bands; 19 opaque bands were removed. The performance of HSI images with real noise will be evaluated on two other benchmark datasets: Indian Pines and the University of Pavia (PaviaU). The Indian Pines image, captured by the AVIRIS sensor, contains 145×145 pixels with 16 ground-truth classes and 224 spectral bands, and 24 water absorption bands were removed. The PaviaU image, obtained from the ROSIS-03 sensor, consists of 610×340 pixels with nine classes and 103 spectral bands.

2) Reference Methods: We compare the proposed IDCN and its reduced versions ID-N (without classification) and I-CN (without denoising) with several recent methods from two aspects. First, to evaluate the denoising performance, four mainstream denoising methods are used.

1) *Traditional Denoising Methods:* KBR [32], FastHyDe [46], and NGmeet [33].
2) *Deep Denoising Method:* HDCNN [25] (we perform the noise reduction on the testing data blindly by applying the well-trained HDCNN).

Second, to evaluate the classification performance, four deep spectral classifiers are included: LCNN [47], SCNN [44], SRNN [43], and CSRNN [48]. The parameters of the reference methods are set to the values indicated in the original works and conducted under the same experimental settings for a fair comparison.

3) Evaluation Criterion: Two criteria, peak signal-to-noise ratio (PSNR) [49] and structure similarity (SSIM) [50], are utilized for denoising evaluation on the simulated data, while a blind image quality assessment criterion (QSFL) [51] is utilized for denoising evaluation on the real data. Note that PSNR and SSIM values are the mean results averaged over all bands. In comparison, larger PSNR and SSIM, and smaller QSFL denote better performance. Two objective metrics, overall accuracy (OA) and kappa coefficient (κ), are used for classification evaluation. For each experiment, we report the mean results averaged over five runs with randomly selected training samples.

4) Experimental Settings: For training, randomly select 12% of the labeled samples in Washington DC and Indian Pines, and 2% in PaviaU. The remaining labeled samples are used as the test set to evaluate the performance of the proposed method. 10% randomly selected training samples are used as the validation set to determine the hyperparameters. The parameters and, where applies, training data of the reference denoisers are set to the default configurations indicated in the original works. All the classifiers are trained with the same training data and the same training configuration as the proposed method. The network hyperparameters of the proposed method are the same for both simulated and real datasets, and are given in Table I. The numbers of maximum epochs and batch size are empirically set to 70 and 64, respectively. The initial learning rate is empirically set to 3, 2, and 0.3 for Washington DC, PaviaU, and Indian Pines, respectively, and it reduces with a clever strategy (i.e., the ReduceLROnPlateau function with a factor of 0.5 and with the patience of 3). The optimizer is Adadelta [37], with weight decay parameter (i.e., rho) as 0.95 and fuzz factor (i.e., epsilon) as 1e-6. Reasoning that more training data can effectively fit our framework, the common data augmentation operations (horizontal and vertical flipping, random angle flipping, and adding a small amount of Gaussian noise) are utilized in our experiments.

B. Simulated Data Experiments

1) Noise Settings: To thoroughly evaluate the performance of the proposed method, we add six types of noise to the supposedly noise-free Washington DC image following [28]. These will be referred to as the following cases.

Case 1: All bands are polluted by zero-mean independent identically distributed (i.i.d.) Gaussian noise $N(0, \sigma^2)$ with $\sigma = 0.05$.

Case 2: All bands are polluted by non-i.i.d. Gaussian noise with different noise levels, yielding PSNR values in the range of 10–20 dB.

Case 3: All bands are polluted by Gaussian noise as in Case 2. Besides, 40 bands were randomly selected to add stripe noise, with a different number of stripes ranging from 20 to 40.

Case 4: All bands are polluted by Gaussian noise as in Case 2, and 40 bands were randomly selected to add dead lines, with a different amount of dead lines ranging from 5 to 15.

Case 5: All bands are polluted by Gaussian noise as in Case 2. Besides, 40 bands were randomly selected to add impulse noise with a different intensity, in the range of 50%–70%.

Case 6: Each band is polluted randomly by at least one kind of noise in Cases 2–5.

TABLE II
DENOISING AND CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS OVER WASHINGTON DC MALL DATASET IN THE SIX NOISE CASES

| Noise type	Pseudo reference			LCNN		SCNN		SRNN		CSRNN		I-CN		IDCN			
	Denoiser	PSNR	SSIM	OA	κ	OA	κ	OA	κ	OA	κ	OA	κ	PSNR	SSIM	OA	κ
Case 1	Noisy image	26.02	0.434	79.41	74.63	83.31	79.47	82.32	78.29	86.18	82.98	86.32	83.16	40.81	0.957	98.68	98.37
	HDCNN	38.74	0.940	91.57	89.64	90.91	88.81	90.50	88.30	91.52	89.56	94.88	93.70	42.23	0.969	98.73	98.44
	KBR	44.37	0.975	94.85	93.68	95.61	94.59	98.21	97.80	97.78	97.27	98.78	98.49	44.21	0.965	98.84	98.57
	NGmeet	42.13	0.967	92.53	90.77	91.96	90.09	90.59	88.42	92.75	91.07	93.49	91.97	45.79	0.972	98.84	98.57
	FastHyDe	45.63	0.986	87.42	84.48	89.99	87.64	92.72	91.04	92.09	90.25	94.53	93.26	44.62	0.980	98.90	98.65
	Clean image	–	–	95.02	93.95	93.39	91.86	96.68	95.91	94.73	93.50	97.38	96.77	47.43	0.979	98.68	98.37
Case 2	Noisy image	26.03	0.469	79.27	74.41	81.79	77.63	82.51	78.48	84.14	80.45	86.89	83.87	41.24	0.949	98.86	98.59
	HDCNN	36.99	0.912	89.37	86.98	90.23	87.96	91.92	90.05	91.28	89.25	95.98	95.04	42.17	0.964	98.91	98.66
	KBR	37.92	0.888	89.01	86.33	91.71	89.79	95.19	94.08	94.20	92.86	95.57	94.53	42.80	0.951	98.72	98.42
	NGmeet	40.03	0.941	80.84	76.32	85.04	81.56	85.54	82.22	85.06	81.60	88.03	85.26	41.87	0.961	98.76	98.48
	FastHyDe	42.83	0.948	89.20	86.73	88.67	85.99	91.88	90.00	91.40	89.39	94.19	92.84	43.28	0.957	98.89	98.63
	Clean image	–	–	95.02	93.95	93.39	91.86	96.68	95.91	94.73	93.50	97.38	96.77	47.87	0.977	98.26	97.86
Case 3	Noisy image	27.26	0.517	77.96	72.63	82.74	78.78	83.41	79.60	84.75	81.23	87.39	84.49	42.53	0.958	98.72	98.42
	HDCNN	37.51	0.919	93.09	91.50	88.96	86.39	91.72	89.80	91.33	89.32	95.60	94.58	43.50	0.968	98.96	98.72
	KBR	41.13	0.939	94.92	93.75	93.76	92.30	96.80	96.05	96.42	95.58	97.41	96.81	43.04	0.957	98.99	98.76
	NGmeet	42.92	0.969	89.99	87.67	89.20	86.73	90.98	88.89	91.33	89.31	92.55	90.82	44.95	0.974	98.92	98.67
	FastHyDe	44.68	0.962	91.53	89.59	90.55	88.35	91.73	89.81	91.34	89.32	94.02	92.62	44.92	0.973	98.91	98.65
	Clean image	–	–	95.02	93.95	93.39	91.86	96.68	95.91	94.73	93.50	97.38	96.77	47.73	0.979	98.87	98.61
Case 4	Noisy image	27.33	0.525	75.82	70.08	83.06	79.18	83.33	79.51	84.80	81.28	87.36	84.45	41.43	0.956	98.81	98.53
	HDCNN	37.55	0.924	90.02	87.75	90.30	88.05	92.10	90.27	91.77	89.86	95.87	94.92	43.24	0.969	98.98	98.74
	KBR	41.48	0.951	87.02	84.22	94.51	93.23	96.07	95.15	96.30	95.44	97.86	97.36	42.26	0.957	98.92	98.67
	NGmeet	42.43	0.969	84.71	81.23	90.55	88.37	91.35	89.35	91.27	89.23	92.87	91.22	43.90	0.974	98.95	98.70
	FastHyDe	44.59	0.967	86.81	85.55	90.66	88.49	91.80	89.91	91.53	89.58	94.00	92.60	45.49	0.978	98.92	98.67
	Clean image	–	–	95.02	93.95	93.39	91.86	96.68	95.91	94.73	93.50	97.38	96.77	48.24	0.979	98.89	98.63
Case 5	Noisy image	26.85	0.486	40.03	25.12	81.59	77.35	81.51	77.26	83.90	80.15	85.22	81.82	42.51	0.958	98.44	98.08
	HDCNN	38.33	0.932	90.90	88.81	90.04	87.74	90.41	88.20	90.89	88.77	94.32	93.00	43.98	0.967	98.56	98.23
	KBR	44.27	0.969	90.38	88.19	95.56	94.52	98.02	97.56	95.87	94.91	98.50	98.15	44.16	0.959	98.70	98.39
	NGmeet	41.98	0.961	84.70	81.08	89.32	86.85	87.78	84.93	89.02	86.47	90.16	87.87	43.94	0.972	98.68	98.38
	FastHyDe	47.22	0.988	92.45	90.72	90.11	87.83	92.46	90.72	91.93	90.06	93.88	92.46	46.95	0.980	98.67	98.36
	Clean image	–	–	95.02	93.95	93.39	91.86	96.68	95.91	94.73	93.50	97.38	96.77	47.34	0.979	98.54	98.21
Case 6	Noisy image	25.84	0.456	73.65	67.29	80.51	76.05	82.52	78.49	83.87	80.14	86.40	83.27	41.50	0.950	98.74	98.44
	HDCNN	36.90	0.912	93.14	91.56	89.60	87.20	92.34	90.57	92.23	90.42	95.80	94.83	42.25	0.960	98.86	98.60
	KBR	37.88	0.886	90.59	90.05	91.36	89.35	94.47	93.20	94.68	93.45	95.45	94.39	42.92	0.949	98.69	98.39
	NGmeet	39.64	0.938	85.31	81.84	86.90	83.89	85.92	82.69	85.63	82.30	88.20	85.47	41.21	0.955	98.51	98.16
	FastHyDe	42.63	0.946	88.52	85.85	90.07	87.78	91.18	89.13	90.92	88.81	93.69	92.22	43.29	0.957	98.84	98.57
	Clean image	–	–	95.02	93.95	93.39	91.86	96.68	95.91	94.73	93.50	97.38	96.77	46.63	0.976	98.83	98.55
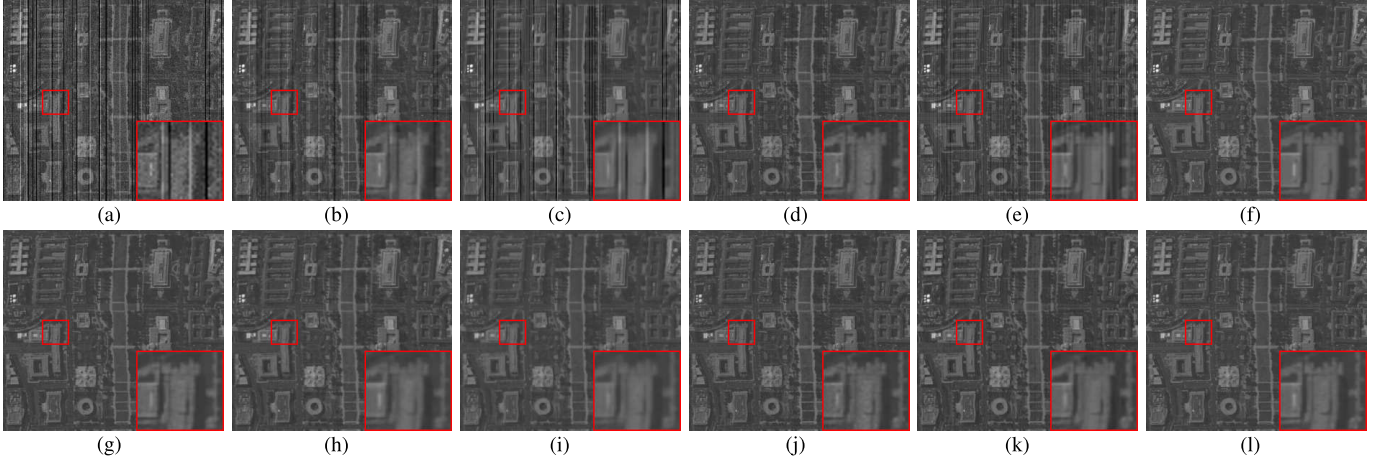
Fig. 4. Denoising results for band 30 in the Washington DC image under mixture noise of Case 6. Top: pseudoreferences. Bottom: proposed IDCN using the corresponding pseudoreferences. (a) Noisy image. (b) HDCNN. (c) KBR. (d) NGmeet. (e) FastHyDe. (f) Clean image. (g) Noisy image+IDCN. (h) HDCNN+IDCN. (i) KBR+IDCN. (j) NGmeet+IDCN. (k) FastHyDe+IDCN. (l) Clean image+IDCN.

2) Denoising and Classification Performance: Table II shows the denoising and classification performance of different denoising and classification methods over the Washington DC image in the six noise cases. In each case, we test the noisy image, the denoised results by using the four comparative denoisers (HDCNN [25], KBR [32], NGmeet [33], and FastHyDe [46]), and the clean (noise-free) image as the pseudoreferences, respectively. The classification results of the four comparative classifiers (LCNN [47], SCNN [44], SRNN [43], and CSRNN [48]) and our I-CN take as input these pseudoreferences for classification. Our IDCN takes as input the noisy image and is supervised by the class label and these pseudoreferences for training, producing the final denoised and classification results simultaneously. For the

convenience of comparison, the best performance in each case is marked in bold, and the best performance for each row is marked as underlined numbers.

Regarding the denoising performance, the proposed IDCN yields better mean PSNR and SSIM than the pseudoreference methods in most cases. This verifies the effectiveness of the proposed method over different pseudoreferences. It is worth highlighting that the improvement in PSNR and SSIM is significant in all six cases when using the noisy image as the pseudoreference. The essential reason for this improvement is the benefit from the joint denoising and classification (see Section IV-D2) and the certain denoising ability of ID-N (see Section IV-D4). It is also evident that the gains in PSNR and SSIM are considerable when employing the denoised results of HDCNN and NGmeet as the pseudoreferences. The results indicate that our IDCN mainly improves the classification performance when the initial denoising result (pseudoreference) is already excellent. For instance, the denoising methods, KBR and FastHyDe, already yield impressive denoising performance, far superior to all other reference methods in Case 1 and Case 5 (with mean PSNR above 44 dB). Starting from this excellent pseudoreference, our IDCN further suppresses noise in some bands (aiming to improve the classification performance) at the expense of deteriorating somewhat the mean PSNR. The drop in mean PSNR is relatively small compared to the differences in performance of the different reference methods.

Fig. 4 shows the corresponding denoising results of band 30 for the Washington DC image in Case 6. We display visual results for Case 6 (mixture noise) because it is the most realistic noise type. The first row shows the pseudoreferences and the second row the denoising results obtained by the proposed IDCN using the corresponding pseudoreferences. Fig. 4(a) shows the input noisy image, which is degraded by a combination of Gaussian and impulse noise, and by stripes and dead lines. It can be observed that the comparative methods HDCNN, KBR, and FastHyDe can hardly remove this complex mixture noise from the images, and NGmeet still maintains some dead line noise compared with the clean image in Fig. 4(f). Comparatively, our IDCN consistently yields appealing visual quality for all the pseudoreferences, smoothing out flat regions, and preserving well sharp edges, as well as subtle texture details.

To illustrate the preserving of spectral details, we plot the spectral reflectance difference curves between the clean image (ground truth) and the denoising results by all the methods in Fig. 5. The smaller the absolute values are, the better spectra preservation. We show spectral reflectance differences at three spatial positions (marked by red, green, and blue boxes). Compared to the noisy image, all the denoisers remove some of the noise, and our IDCN yields the smallest difference across almost all bands at the three positions. Thus, in these experiments, our IDCN yields favorable denoising results both visually and quantitatively compared to the reference methods.

Without the denoising step (i.e., the noisy image case), as shown in Table II, our I-CN yields better classification results than LCNN, SCNN, SRNN, and CSRNN in all the six noise cases, which verifies the effectiveness of our approach.

It is clear that all the classification results have been indeed improved when involving the denoising step, which confirms the benefit of denoising for classification. Our IDCN consistently yields the best OA and κ with a significant improvement over the comparative methods in all the noise cases, especially in the noisy image case where the gain of OA is more than 11%. Compared to I-CN (our network where the denoising part is switched OFF), IDCN consistently yields better classification results due to the benefit of the proposed joint learning. It is worth emphasizing that our IDCN yields stable classification performance regardless of the noise type and the pseudoreference. This confirms important earlier findings, which also served as motivation for our work as stated in the Introduction section, which separates denoising and classification and leads to suboptimal classification accuracy.

Using the clean image as the pseudoreference with any method, including the proposed IDCN, yields the best or nearly the best denoising and classification performance in all noise cases. One possible reason for this nearly and not absolutely best classification performance is that the clean image also contains some small amount of noise and a certain spectral variability within each class.

Fig. 6 shows the full classification maps using the same classifier (our I-CN) on Washington DC in Case 6 and on its denoised results obtained by different denoisers. Visually, they are consistent with the results reported in Table II and Fig. 4. It is obvious that the visual results of comparative denoisers are better than that of the noisy image [see Fig. 6(a)]. Our IDCN [see Fig. 6(f)] exhibits a smoother appearance and preserves the local details better than all the other methods.

We further analyze the influence of the percentage of training samples Tr on denoising and classification performance. We adopt the noisy images as the pseudoreferences and conduct experiments on the Washington DC image in the six noise cases, respectively. The results in Fig. 7 show that PSNR first increases with increasing Tr and then tends to stabilize or fluctuate in some cases. The improved performance when more training data are given is indeed expected for a deep network. The reason for fluctuations in PSNR is that training samples are here noisy data, and thus, increasing Tr means also that a larger amount of possibly erroneous training data will be employed. Contrary to this, the class labels of all the training samples are correct, and thus, OA monotonically increases with Tr as expected, until it reaches stable values. It is worth highlighting that our IDCN yields stable OA for all six cases, exhibiting robustness to different noise types.

C. Real Data Experiments

Table III reports the denoising and classification performance of different denoising and classification methods over the Indian Pines and PaviaU images. Regarding the denoising performance, the proposed IDCN yields better or comparable QSFL than the pseudoreference methods on the two datasets, which demonstrates the effectiveness of the proposed IDCN over different pseudoreferences. Note that our IDCN effectively improves QSFL when using the noisy image as the pseudoreference. The main reason is the effectiveness of the
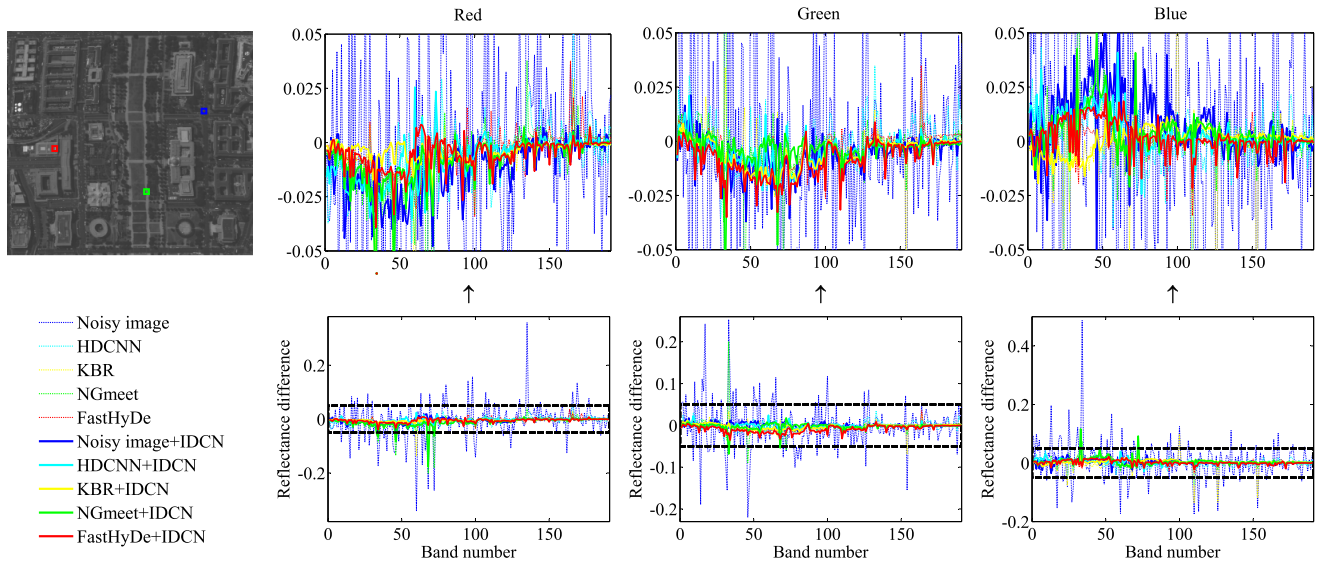
Fig. 5. Spectral reflectance difference curves between the clean image (ground-truth) and the recovered results obtained by different denoising methods at three marked positions (red, green, and blue boxes) for Washington DC image in Case 6. Note that dash lines show the curves of different pseudoreferences, and solid lines give the curves of the proposed IDCN using the corresponding pseudoreferences.
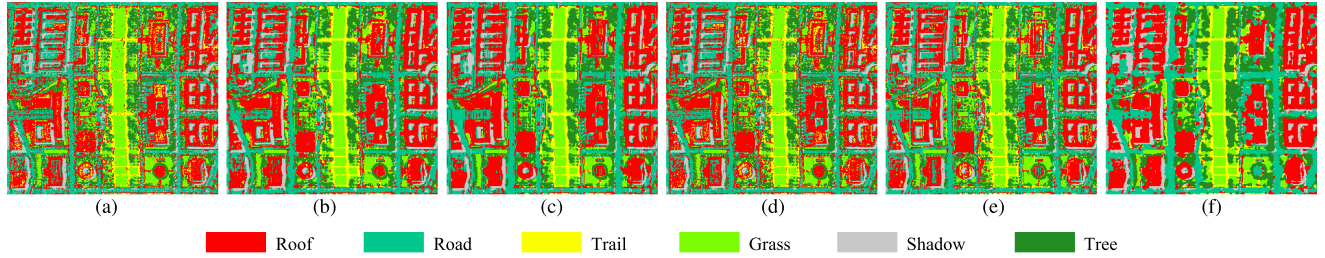

Fig. 6. Full classification maps for the Washington DC image in Case 6 using the same classification method (our I-CN) before and after denoising. (a) Noisy image. (b) HDCNN. (c) KBR. (d) NGmeet. (e) FastHyDe. (f) Our IDCN using the noisy image as the pseudoreference (Noisy image+IDCN).

TABLE III
DENOISING AND CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS OVER THE INDIAN PINES AND PAVIAU IMAGES

| Real data	Pseudo reference		LCNN		SCNN		SRNN		CSRNN		I-CN		**IDCN**		
	Denoiser	QSFL	OA	κ	OA	κ	OA	κ	OA	κ	OA	κ	QSFL	OA	κ
Indian Pines	Noisy image	33.54	79.50	76.40	75.83	72.31	83.41	81.06	81.96	79.39	89.11	87.55	17.50	97.85	97.54
	HDCNN	23.30	88.20	86.50	85.59	83.52	89.07	87.52	84.51	82.33	91.32	90.07	17.93	97.87	97.57
	KBR	18.39	86.04	84.04	87.01	85.16	90.07	88.67	83.88	81.62	90.92	89.63	18.27	98.00	97.72
	NGmeet	22.63	79.19	76.20	75.48	71.91	85.02	82.86	79.29	76.38	85.53	83.45	23.61	**98.10**	97.83
	FastHyDe	19.82	86.48	84.53	84.79	82.62	85.93	83.95	79.70	76.78	89.57	88.08	17.69	97.86	97.56
PaviaU	Noisy image	17.01	85.52	80.23	89.70	86.23	89.51	86.07	84.00	78.69	93.11	90.82	15.28	98.01	97.36
	HDCNN	**14.83**	87.74	83.33	91.79	89.04	91.11	86.70	82.83	81.84	94.81	93.09	15.49	98.01	97.36
	KBR	16.73	81.00	73.74	90.51	87.17	89.46	86.16	88.74	84.95	94.15	92.22	15.55	97.99	97.33
	NGmeet	18.38	84.44	78.58	93.76	91.57	96.14	94.86	95.64	94.21	97.29	96.39	16.71	**98.18**	97.58
	FastHyDe	16.46	84.96	79.45	86.51	81.80	90.40	87.21	84.72	79.59	93.32	91.10	14.87	97.93	97.25
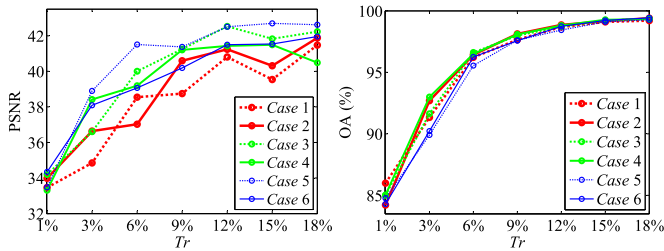
Fig. 7. Performance of the proposed method in terms of PSNR (left) and OA (right) depending on the percentage of the training samples per class Tr for the Washington DC image with the six noise cases.

proposed denoiser by making use of global spectral correlations. A detailed analysis will be given in Section IV-D4). Figs. 8 and 9 show the restoration results for the Indian

Pines and PaviaU images, respectively. It is clear that all the denoising methods improve the visual effects compared to the noisy images, and our IDCN still yields appealing visual quality compared to the reference denoising methods.

Without the denoising step, our I-CN consistently yields the best OA with a significant improvement over the comparative classification methods for the two datasets. For example, on Indian Pines, the improvement in OA compared to LCNN, SCNN, SRNN, and CSRNN is about 9.6%, 13.3%, 5.7%, and 7.1%, respectively. Different from the simulated data, not all the classification results in real data have been improved when the denoising and classification are processed separately, e.g., NGmeet+CSRNN and NGmeet+I-CN on Indian Pines, and NGmeet+LCNN and FastHyDe+SCNN on PaviaU. A similar case also happened in [30]. One possible reason is that the
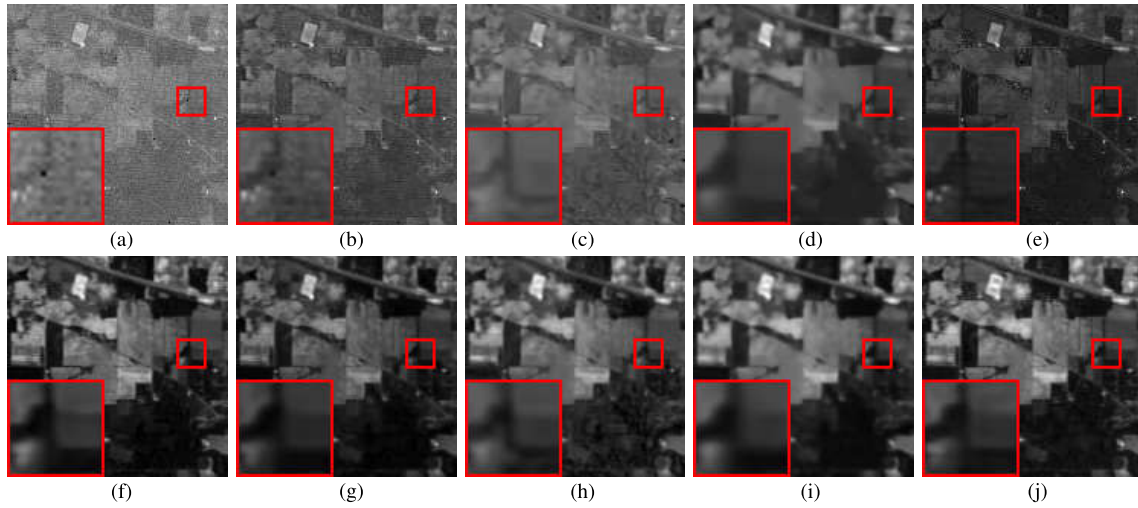
Fig. 8. Restoration results for band 2 in Indian Pines. Top: pseudoreferences. Bottom: proposed IDCN using the corresponding pseudoreferences. (a) Noisy image. (b) HDCNN. (c) KBR. (d) NGmeet. (e) FastHyDe. (f) Noisy image+IDCN. (g) HDCNN+IDCN. (h) KBR+IDCN. (i) NGmeet+IDCN. (j) FastHyDe+IDCN.
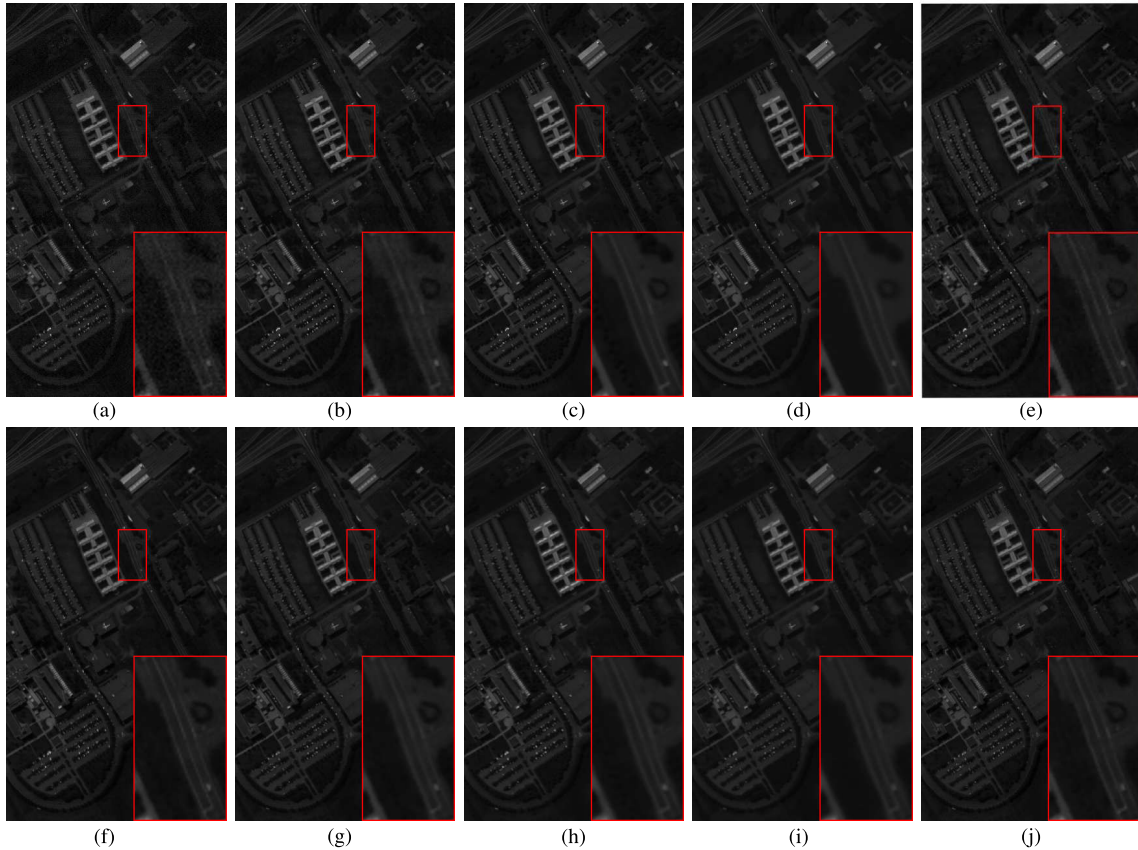

Fig. 9. Denoising results for band 3 in PaviaU. Top: pseudoreferences. Bottom: proposed IDCN using the corresponding pseudoreferences. (a) Noisy image. (b) HDCNN. (c) KBR. (d) NGmeet. (e) FastHyDe. (f) Noisy image+IDCN. (g) HDCNN+IDCN. (h) KBR+IDCN. (i) NGmeet+IDCN. (j) FastHyDe+IDCN.

denoising methods have degraded the noise-free bands since the real noisy data have more complex noise distribution and structure than the simulated noisy data. Our IDCN consistently yields stable classification performance for the two testing datasets, regardless of the pseudoreferences, proving the benefit of its end-to-end joint learning.

Figs. 10 and 11 show the full classification maps using our I-CN classifier on Indian Pines and PaviaU, and their denoising results obtained by different denoisers. Visually, they are consistent with the denoising results, as shown in Figs. 8 and 9. The comparative denoisers deliver different degrees of improvement [except Fig. 10(d)] than the noisy image. It is obvious that our IDCN shows the best visual performance, preserving the local details and structural information.

We further analyze the influence of the proposed IDCN using different pseudoreferences with the percentage of training samples Tr on the two test datasets. The results in Fig. 12 show that the OA values improve first abruptly and then continue to rise gradually with increasing Tr in all the pseudoreferences. It is clear that our IDCN yields stable
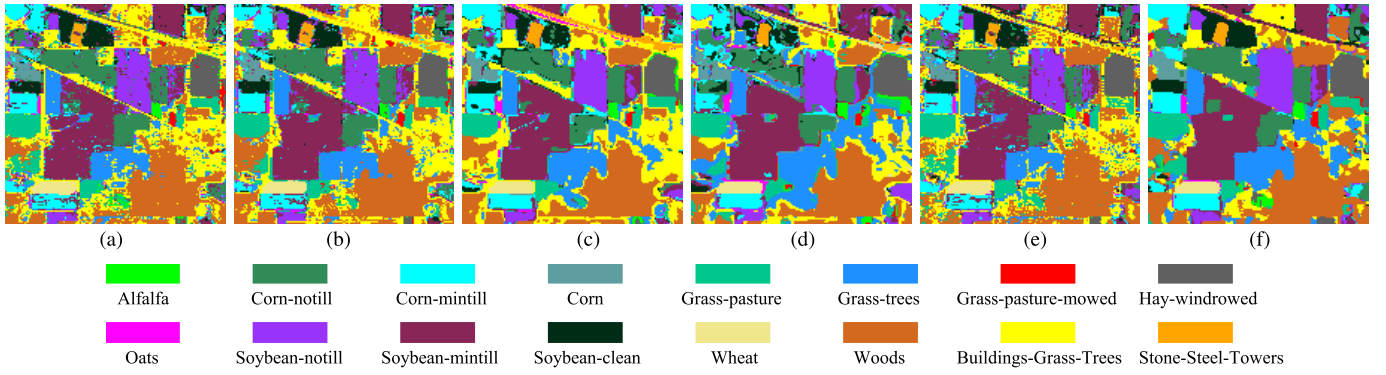
Fig. 10.   Full classification maps for the Indian Pines image using the same classification method (our I-CN) before and after denoising. (a) Noisy image. (b) HDCNN. (c) KBR. (d) NGmeet. (e) FastHyDe. (f) Our IDCN using the noisy image as the pseudoreference (Noisy image+IDCN).
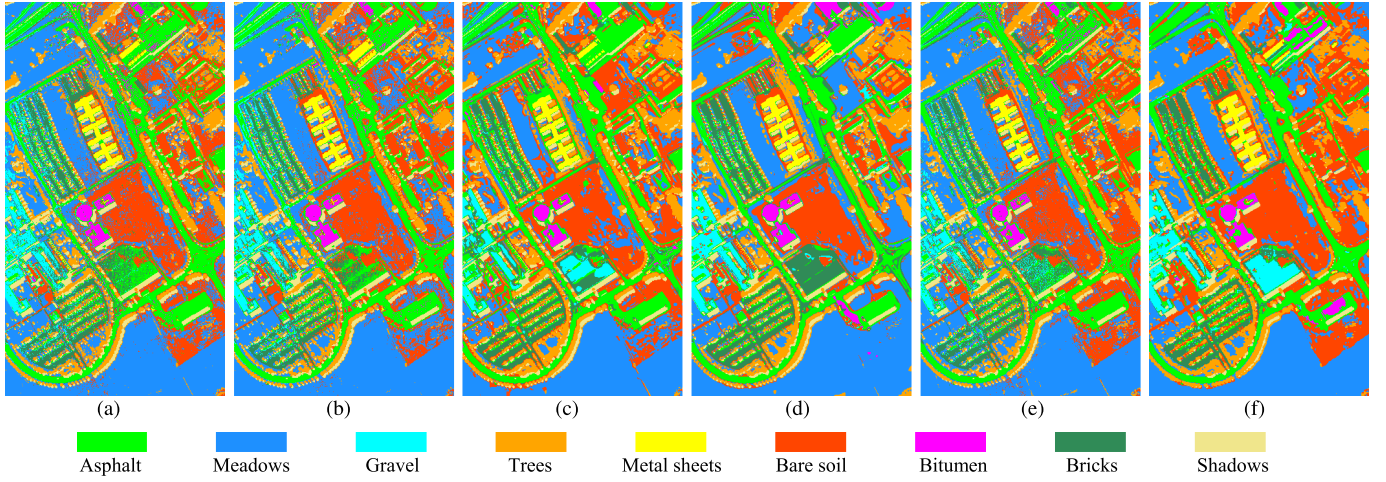

Fig. 11.   Full classification maps for the PaviaU image using the same classification method (our I-CN) before and after denoising. (a) Noisy image. (b) HDCNN. (c) KBR. (d) NGmeet. (e) FastHyDe. (f) Our IDCN using the noisy image as the pseudoreference (Noisy image+IDCN).
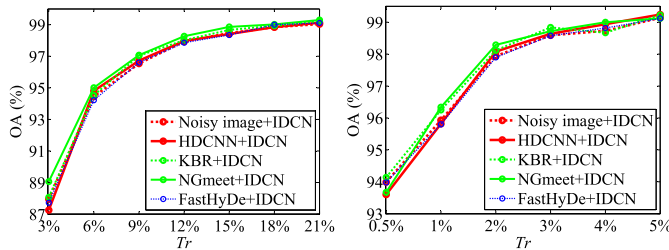

Fig. 12.   OA of the proposed IDCN using different pseudoreferences with different percentages of training samples per class Tr on Indian Pines (left) and PaviaU (right).

TABLE IV

PERFORMANCE OF OUR METHOD TRAINED WITH THE MSE LOSS AND THE PROPOSED LAE LOSS ON SIMULATED AND REAL NOISY DATASETS

Data sets	MSE los				Our LAE loss			
Simulated	PSNR	SSIM	OA	κ	PSNR	SSIM	OA	κ
Case 1	38.52	0.935	98.50	98.15	40.81	0.957	98.68	98.37
Case 2	36.86	0.822	98.30	97.91	41.24	0.949	98.86	98.59
Case 3	39.37	0.897	98.54	98.20	42.53	0.958	98.72	98.42
Case 4	39.67	0.911	98.44	98.07	41.43	0.956	98.81	98.53
Case 5	39.53	0.883	98.00	97.54	42.51	0.958	98.44	98.08
Case 6	36.39	0.804	98.01	97.55	41.50	0.950	98.74	98.44
Real	QSFL		OA	κ	QSFL		OA	κ
Indian Pines	20.63		97.58	97.24	17.50		97.85	97.54
PaviaU	15.50		98.01	97.36	15.28		98.03	97.38

classification accuracy for different qualities of pseudoreferences. This verifies the robustness of the proposed method.

D. Ablation Study

We analyze several key components and settings of the proposed method in view of denoising and classification performance in both simulated and real datasets. Note that we choose the noisy image as the pseudoreference in all the experiments to better verify the proposed method. We experimentally find that the six simulated noise cases hold similar trends. We, thus, report the results for two representative cases: Case 1 and Case 6 in some situations.

1) Effect of the Denoising Loss: The proposed compound loss function consists of the proposed LAE denoising loss and the cross-entropy classification loss. We test the effectiveness

of LAE by comparing it with the commonly used mse loss under the same classification loss. The results in Table IV show that our LAE consistently yields superior denoising performance under comparable classification performance compared to mse, which verifies its effectiveness.

We further analyze the restoration results of the proposed LAE across all bands in terms of PSNR and SSIM. Fig. 13 provides a comparative analysis among the noisy image, the restoration results obtained by mse, and the proposed LAE for the Washington DC image in the six types of noise. It is clear that both mse and LAE yield better PSNR and SSIM than the noisy image in all noise cases, with a significant improvement. It is also evident that our LAE
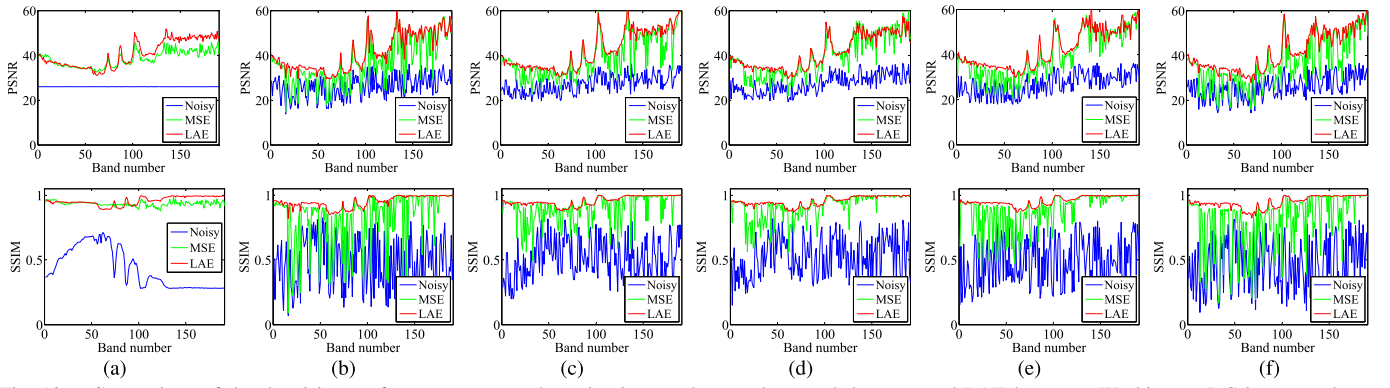
Fig. 13. Comparison of the denoising performance among the noisy image, the mse loss, and the proposed LAE loss over Washington DC image under a certain type of noise. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4. (e) Case 5. (f) Case 6.
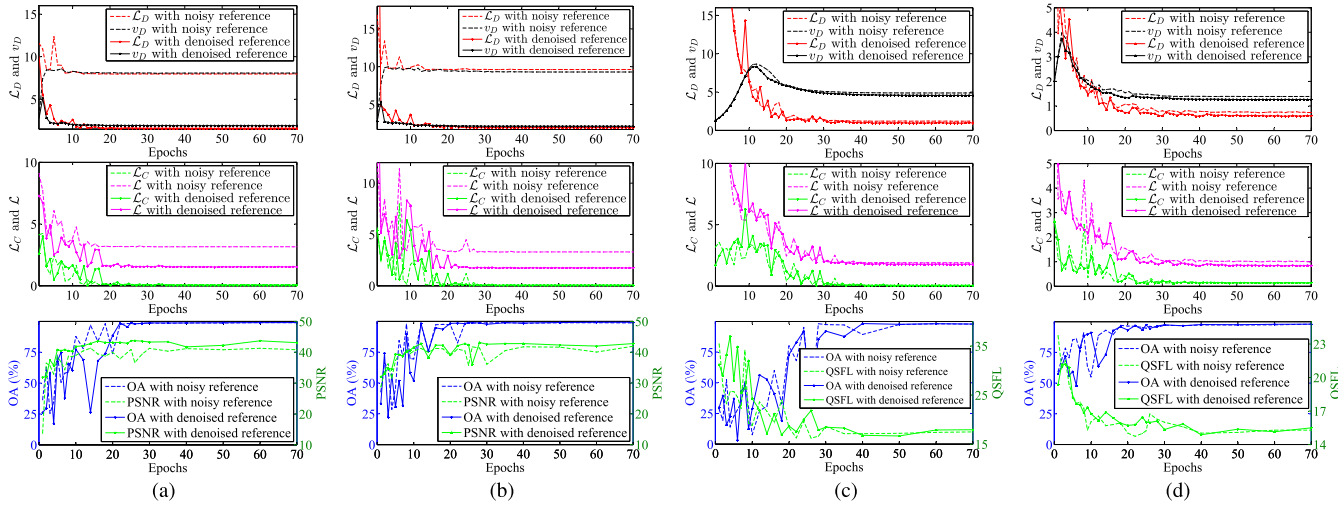

Fig. 14. Comparison of the evolution of each component in the compound loss with the noisy pseudoreference and a denoised reference HDCNN [25] for simulated and real noisy data. Top: noise variance v_D and the denoising loss $\mathcal{L}_D$. Middle: classification loss $\mathcal{L}_C$ and the compound loss $\mathcal{L}$. Bottom: performance of the proposed method in terms of OA (left Y-axis) and PSNR or QSFL (right Y-axis). (a) Case 1. (b) Case 6. (c) Indian Pines. (d) PaviaU.

TABLE V
COMPARISON OF THE DENOISING PERFORMANCE OF THE PROPOSED ID-N AND IDCN FOR SIMULATED AND REAL NOISY DATA. FOR REFERENCE, HDCNN [25] IS INCLUDED AS WELL

Data sets	Noisy HSI		ID-N		IDCN		HDCNN	
Simulated	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Case 1	26.02	0.434	32.93	0.788	40.81	0.957	38.74	0.940
Case 2	26.03	0.469	33.80	0.768	41.24	0.949	36.99	0.912
Case 3	27.26	0.517	34.97	0.824	42.53	0.958	37.51	0.919
Case 4	27.33	0.525	35.37	0.839	41.43	0.956	37.55	0.924
Case 5	26.85	0.486	36.62	0.855	42.51	0.958	38.33	0.932
Case 6	25.84	0.456	32.61	0.727	41.50	0.950	36.90	0.912
Real	QSFL		QSFL		QSFL		QSFL	
Indian Pines	33.54		16.86		17.50		23.30	
PaviaU	17.01		14.83		15.28		14.83	

achieves the best PSNR and SSIM values across almost all bands. This verifies the robustness of the proposed LAE loss over entire spectral bands. Compared to mse, our LAE yields stable denoising performance with a small fluctuation over entire spectral bands, especially in terms of SSIM. This is because the proposed LAE considers the nonlinear characteristic of spectra and nonuniform noise statistics across the bands.

2) Effect of the Compound Loss: It is also of interest to analyze the evolution of each component of the proposed compound loss during the training. Fig. 14 provides a

comparative analysis between the noisy pseudoreference and a denoised pseudoreference of HDCNN [25] for two simulated and real noisy data. Similar trends hold for other simulated noise cases and other denoised pseudoreferences. The results in the first row of Fig. 14 show that noise variance v_D first increases quickly with increasing epoch and then declines (with some fluctuations in some cases) until it reaches stable values for both noisy and denoised pseudoreferences in all the testing data. It is also evident that the proposed method with a denoised reference achieves smaller v_D than the one with the noisy pseudoreference across almost all epochs, especially for the simulated cases. The reason is that the denoised pseudoreference has a higher image quality than the noisy pseudoreference, corresponding to a higher weight $(1/v_D)$ as well.

With the joint learning, the denoising loss $\mathcal{L}_D$, the classification loss $\mathcal{L}_C$, and the compound loss $\mathcal{L}$ converge in similar epochs (in around 30 or 40 epochs) as the evolution of v_D for all the testing data. This implies that all the components of the compound loss are working together to achieve the same objective of minimizing $\mathcal{L}$, so as to obtain favorable denoising and classification performance, which is verified by the results in the bottom row of Fig. 14. Considering an environment where the pseudoreference is extremely noisy and v_D becomes very large (i.e., the impact of $\mathcal{L}_D$ is eliminated nearly), there

TABLE VI
COMPARISON OF COMPUTATIONAL COMPLEXITY OF DIFFERENT DENOISING AND CLASSIFICATION METHODS FOR THE WASHINGTON DC IMAGE

Methods	KBR	NGmeet	FastHyDe	HDCNN	LCNN	SCNN	SRNN	CSRNN	IDCN
Running time (s)	1603.9	110.7	3.9	27011.7	74.6	41.5	33.5	123.9	910.3
# Params (×M)	-	-	-	0.40	0.08	0.17	0.02	0.35	5.46
FLOPs (×10^6)	-	-	-	34311.88	0.30	0.89	0.04	8.11	221.92
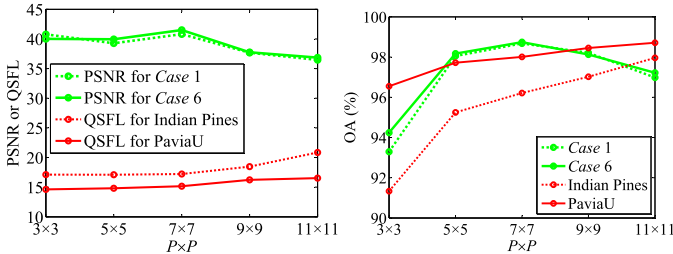
Fig. 15. Effect of the input patch size $P \times P$. Left: PSNR for simulated noisy images and QSFL for real noisy images. Right: OA.

Fig. 16. Effect of the scale on classification accuracy.

is still joint learning because the gradients of the classifier inevitably update the denoiser in the back-propagation (see Section III-A and Fig. 1).

3) Effect of the Input Patch Size: The effect of the input patch size $P \times P$ on the denoising and classification performance of the proposed method is illustrated in Fig. 15. For the analyzed real and simulated datasets, the denoising results first tend to be stable and then decline slightly when $P \geq 7$. A possible reason is that a large P contains more heterogeneous regions, which distorts the spectral reflectance of the center pixel. In contrast, the OA values first significantly increase ($P \leq 5$) because the use of spatial context in the denoising network benefits the classification and then tend to be stable or slightly decline when P becomes so large that the patches are heterogeneous, which complies with the findings in [43]. We choose $P = 7$ as a tradeoff between denoising and classification performance over all the test images.

4) Effect of the All-Band Denoiser: The results in Table V provide a comparative denoising analysis between our IDCN and its reduced version without classifier ID-N, both using the noisy images as pseudoreferences and a reference deep denoiser HDCNN [25]. Observe that ID-N indeed improves all the metrics compared to the noisy input. This is because ID-N makes use of all spectral information and employs the data augmentation and min-batch training strategies, which randomly introduces the error between the output of ID-N and its pseudoreference, improving thereby denoising ability. However, ID-N alone shows unsatisfactory denoising results compared to any reference denoisers, including HDCNN, and compared to the full IDCN in all the six simulated cases (see Table II). An interesting result is that ID-N yields better QSFL than other denoisers on the two datasets with real noise. This may be due to the fact that the level of noise in these datasets is very small, and the state-of-the-art denoisers and IDCN may smooth out some features that are important for QSFL.

5) Effect of the Spectral Classification: We also analyze the effect of our group spectral classification approach (I-CN) without the denoising step, which extracts multiscale spectral features. The parameter settings are given in Table I. The results in Tables II and III demonstrate that our I-CN yields
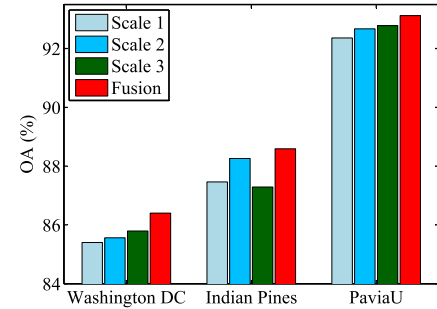
better accuracy than the comparative methods [43], [44], [47], [48] for both simulated and real datasets. We also test the effect of the scale in our I-CN on both simulated and real datasets. The results in Fig. 16 show that our I-CN with multiscale extraction yields better OA for all the test datasets compared to using an arbitrary single scale. In practice, this is even more important because it may not be evident upfront which scale is optimal for processing the data at hand.

6) Analysis of the Computational Efficiency: The results in Table VI provide a comparative analysis of the computational complexity of the proposed IDCN, four comparative denoisers: HDCNN [25], KBR [32], NGmeet [33], and FastHyDe [46], and four comparative classifiers: LCNN [47], SCNN [44], SRNN [43], and CSRNN [48]. The following attributes are reported: the running time for all the methods (i.e., the total processing time for the traditional denoising methods: KBR, NGmeet, and FastHyDe, the training and testing time in total for the deep learning methods), the number of parameters, and the number of floating point operations (FLOPs) for the deep learning models. The reported values correspond to one of the datasets (Washington DC) and are similar to the other two test datasets. All experiments are conducted on an Intel Core i7-7820X CPU with an Nvidia TITAN Xp GPU.

The proposed IDCN is more time-consuming than most of the combinations of the comparative denoisers and classifiers (except KBR and HDCNN) due to much more parameters and FLOPs (except HDCNN). This is attributed to the training of the unified network, which employs the group convolutions over the spectral dimension, inputting both local spatial and full spectral information, and using the data augmentation strategy. This is the price for the improved denoising and classification performance that IDCN offers and its robust behavior across different datasets and different noise scenarios.

V. CONCLUSION

In this article, we proposed an end-to-end deep framework to explore the interplay between the denoising and classification of HSIs. In this framework, we developed an all-band denoising network to connect with the classifier in

an end-to-end fashion. We formulated an appropriate compound loss to optimize our framework from scratch, making the denoising and classification benefit each other iteratively. An important component of this compound loss is a nonuniform denoising loss, defined to take into account nonuniform noise statistics across the bands. Experimental results on simulated and real data demonstrated favorable denoising and robust classification performance compared to the reference methods. In our future research, we will try to extend the proposed framework to more image applications, such as super-resolution [52], [53], segmentation [54], and object detection [55].

REFERENCES

[1] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2018.

[2] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[3] C. Ramirez-Reyes et al., "Reimagining the potential of Earth observations for ecosystem service assessments," *Sci. Total Environ.*, vol. 665, pp. 1053–1063, May 2019.

[4] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1037–1050, Mar. 2022.

[5] J. Marcello, E. Ibarrola-Ulzurrun, C. Gonzalo-Martin, J. Chanussot, and G. Vivone, "Assessment of hyperspectral sharpening methods for the monitoring of natural areas using multiplatform remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8208–8222, Oct. 2019.

[6] H. Shi, Z. Li, H. Ye, H. Luo, W. Xiong, and X. Wang, "First level 1 product results of the greenhouse gas monitoring instrument on the GaoFen-5 satellite," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 899–914, Feb. 2021.

[7] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1529–1543, May 2020.

[8] L. Zhang, J. Nie, W. Wei, Y. Li, and Y. Zhang, "Deep blind hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2388–2400, Jun. 2020.

[9] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, Jul. 2013.

[10] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[11] J. Jiang, J. Ma, and X. Liu, "Multilayer spectral–spatial graphs for label noisy robust hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 839–852, Feb. 2022.

[12] S. Jia, Z. Lin, B. Deng, J. Zhu, and Q. Li, "Cascade superpixel regularized Gabor feature fusion for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1638–1652, May 2019.

[13] Z. Gong, P. Zhong, and W. Hu, "Statistical loss and analysis for deep learning in hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 322–333, Jan. 2021.

[14] N. He, L. Fang, S. Li, J. Plaza, and A. Plaza, "Skip-connected covariance network for remote sensing scene classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1461–1474, May 2020.

[15] C. Liu, J. Li, L. He, A. J. Plaza, S. Li, and B. Li, "Naive Gabor networks for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 376–390, Jan. 2021.

[16] X. Li, M. Ding, and A. Pižurica, "Deep feature fusion via two-stream convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2615–2629, Apr. 2020.

[17] H.-C. Li, W.-S. Hu, W. Li, J. Li, Q. Du, and A. Plaza, "A³CLNN: Spatial, spectral and multiscale attention ConvLSTM neural network for multisource remote sensing data classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 747–761, Feb. 2022.

[18] N. Audebert, B. L. Saux, and S. Lefevre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, Jun. 2019.

[19] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 427–436.

[20] P. Zhong and R. Wang, "Jointly learning the hybrid CRF and MLR model for simultaneous denoising and classification of hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1319–1334, Jul. 2014.

[21] D. Liu, B. Wen, J. Jiao, X. Liu, Z. Wang, and T. S. Huang, "Connecting image denoising and high-level vision tasks via deep learning," *IEEE Trans. Image Process.*, vol. 29, pp. 3695–3706, 2020.

[22] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[23] 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.

[24] Q. Zhang, Q. Yuan, J. Li, X. Liu, H. Shen, and L. Zhang, "Hybrid noise removal in hyperspectral imagery with a spatial–spectral gradient network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7317–7329, Oct. 2019.

[25] Q. Yuan et al., "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.

[26] B. Lin, X. Tao, and J. Lu, "Hyperspectral image denoising via matrix factorization and deep prior regularization," *IEEE Trans. Image Process.*, vol. 29, pp. 565–578, 2020.

[27] T. Plotz and S. Roth, "Benchmarking denoising algorithms with real photographs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1586–1595.

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

[29] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1712–1722.

[30] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5174–5189, Jul. 2019.

[31] H. Fan, C. Li, Y. Guo, G. Kuang, and J. Ma, "Spatial–spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6196–6213, Oct. 2018.

[32] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1888–1902, Aug. 2018.

[33] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6861–6870.

[34] V. Lukin, S. Abramov, S. Krivenko, A. Kurekin, and O. Pogrebnyak, "Analysis of classification accuracy for pre-filtered multichannel remote sensing data," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6400–6411, Nov. 2013.

[35] D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2650–2658.

[36] R. Cipolla, Y. Gal, and A. Kendall, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7482–7491.

[37] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," 2012, *arXiv:1212.5701*.

[38] W. Liu and J. Lee, "A 3-D atrous convolution neural network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5701–5715, Aug. 2019.

[39] V. Nair and G. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[40] T. Zhang, G.-J. Qi, B. Xiao, and J. Wang, "Interleaved group convolutions," in *Proc. IEEE CVPR*, Oct. 2017, pp. 4373–4382.

[41] G. Huang, S. Liu, L. V. D. Maaten, and K. Q. Weinberger, "CondenseNet: An efficient DenseNet using learned group convolutions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2752–2761.

[42] N. Ma et al., "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. ECCV*, Sep. 2018, pp. 116–131.

[43] Y. Xu, L. Zhang, B. Du, and F. Zhang, "Spectral–spatial unified networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5893–5909, Oct. 2018.

[44] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[45] 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.

[46] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.

[47] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, pp. 1–12, Jan. 2015.

[48] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019.

[49] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.

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

[51] 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, Mar. 2017.

[52] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral images super-resolution via learning high-order coupled tensor ring representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4747–4760, Nov. 2020.

[53] J. Xue, Y.-Q. Zhao, Y. Bu, W. Liao, J. C.-W. Chan, and W. Philips, "Spatial–spectral structured sparse low-rank representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 3084–3097, 2021.

[54] X. Chang, H. Pan, W. Sun, and H. Gao, "YolTrack: Multitask learning based real-time multiobject tracking and segmentation for autonomous vehicles," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5323–5333, Dec. 2021.

[55] L. Jiao et al., "New generation deep learning for video object detection: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3195–3215, Aug. 2022, doi: 10.1109/TNNLS.2021.3053249.

Mingli Ding received the B.S. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 2000 and 2005, respectively.

From 2009 to 2010, he was a Visiting Scholar with the French National Center for Scientific Research, Toulouse, France. He is currently a Full Professor with the School of Instrumentation Science and Engineering, Harbin Institute of Technology. His research interests include deep learning, image classification, object detection, automation test technology, and information processing.

Yanfeng Gu (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 2005.

He joined the School of Electronics and Information Engineering, HIT, as a Lecturer. He was appointed as an Associate Professor at HIT in 2006; meanwhile, he was enrolled in the first Outstanding Young Teacher Training Program of HIT. From 2011 to 2012, he was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, CA, USA. He is currently a Professor with the Department of Information Engineering, HIT. He has published more than 100 peer-reviewed papers and four book chapters. He is the inventor or a coinventor of 20 patents. His research interests include space-intelligent remote sensing and information processing, multimodal hyperspectral remote sensing, and spaceborne time-series image processing.

Aleksandra Pižurica (Senior Member, IEEE) received the Diploma degree in electrical engineering from the University of Novi Sad, Novi Sad, Serbia, in 1994, the M.Sc. degree in telecommunications from the University of Belgrade, Belgrade, Serbia, in 1997, and the Ph.D. degree in engineering from Ghent University, Ghent, Belgium, in 2002.

She is currently a Professor of statistical image modeling with Ghent University. Her research interests include the area of signal and image processing and machine learning, including multiresolution statistical image models, Markov random field models, sparse coding, representation learning, and image and video reconstruction, restoration, and analysis.

Prof. Pižurica received the scientific prize "de Boelpaepe" for 2013–2014, awarded by the Royal Academy of Science, Letters and Fine Arts of Belgium for her contributions to statistical image modeling and applications to digital painting analysis. The work of her team has been awarded twice the Best Paper Award of the IEEE Geoscience and Remote Sensing Society Data Fusion Contest in 2013 and 2014. She has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2012 to 2016 and a Senior Area Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2016 to 2019. She was a Lead Guest Editor of the *EURASIP Journal on Advances in Signal Processing* for the Special Issue "Advanced Statistical Tools for Enhanced Quality Digital Imaging With Realistic Capture Models" in 2013. She is also an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.

Xian Li (Member, IEEE) received the Ph.D. degree in instrument science and technology from the Harbin Institute of Technology (HIT), Harbin, China, in 2021.

From 2018 to 2020, he was a Doctoral Researcher with the Department of Telecommunications and Information Processing, Ghent University, Ghent, Belgium, supported by the China Scholarship Council. He is currently an Assistant Professor with the School of Electronics and Information Engineering, HIT. His research interests include deep learning, hyperspectral remote sensing, and multisource remote sensing data processing.