HSI-IPNet: Hyperspectral Imagery Inpainting by Deep Learning With Adaptive Spectral Extraction

Rong Wong [ID], Zhijiang Zhang, Yueming Wang [ID], Fansheng Chen [ID], and Dan Zeng [ID]

Abstract—**Feature representation is the key to the hyperspectral images (HSI) inpainting. Existing works mainly focus on using spectral and temporal auxiliary images to inpainting the corrupted region, which were proved to be low robust for all bands missing and high requirements for image acquisition. In this work, we propose an end-to-end inpainting framework for HSI based on convolutional neural networks, which does not require auxiliary images and makes full use of both spectral characteristics and spatial information. For spectral characteristics, a channel attention mechanism is proposed to reduce the redundancy of hyperspectral channels and model the correlation between channels. For spatial information, a local discriminative network is able to cope with the structural continuity of the corrupted regions, and a gradient consistency loss function is proposed to maintain the texture consistency of HSIs. Experimental results in the Airborne Visual Infrared Imaging Spectrometer Indians Pines public dataset and Feicheng Hyperspectral datasets show that our proposed method can provide competitive results compared with state-of-the-art methods.**

Index Terms—**Channel attention mechanism (CAM), global and local discriminators, gradient consistency (GC), hyperspectral remote sensing, image inpainting.**

I. INTRODUCTION

HYPERSPECTRAL remote sensing has been applied to all aspects of earth science and has become an effective technical means in the fields of geological mapping [40], vegetation survey [41], ocean remote sensing [42], agricultural remote sensing [43], atmospheric research [44], and environmental monitoring [45]. However, the problems of instrument instability, slit contamination, and atmospheric interference seriously

affect the quality of the hyperspectral image (HSI), appearing as the lack of various stripes types [12], [13], [14] and the cloud cover [10]. Thus, HSI inpainting is an essential computer vision task in the field of remote sensing [26], [27], [30].

Image inpainting can be described as the process of filling missing data in a specified area of visual input. For the inpainting of grayscale and RGB images, it can be classified into three categories: diffusion-based methods, exemplar-based methods, and learning-based methods.

Diffusion-based methods are developed to fill small or narrow holes by propagating adjacent image pixels into the corrupted region [24], such as partial differential equations (PDE) [15] and total variation (TV) [16]. The core of the PDE method is to fill corrupted regions by spreading information around the area to be repaired along the direction of the illumination line. The TV model used the Euler–Lagrange equation [33] and anisotropic diffusion to maintain the direction of the geometry consistent with the image. These techniques cannot correctly restore texture information, result in less effectiveness in handing large missing areas, and blurring often exists. Chen *et al.* [34] propagated the structure by estimating missing values in the visual tensor to fill the corrupted region by tensor decomposition and completion methods.

Exemplar-based methods try to capture textures with a repetitive pattern and complete the corrupted region with similar neighborhoods [18]. PatchMatch [35] model is a fast-nearest neighbor patch search algorithm and reshuffles the image to make the output more authentic. Huang *et al.* [36] used the mid-level structural cues for automatically guiding patch-based image completion. Owing to the self-similar redundancy in natural images, He and Sun [37] believed that the existed sparsity in the offset between patch blocks and constraints on offsets can produce better results.

Learning-based methods are currently mainly based on Generative Adversarial Network (GAN) [28]. GAN made the samples generated by the generative network obey the real-data distribution as much as possible through adversarial training. Pathak *et al.* [2] developed Context Encoder, proposing the possibility to predict the missing structure with convolutional neural networks (CNNs) and to train the model with the combination of l_2 and adversarial loss function. Based on this work and reducing visible artifacts around the border of the hole, Yu *et al.* [8] proposed a novel contextual attention layer to explicitly attend on related feature patches at distant spatial locations. Iizuka *et al.* [9] proposed a globally and locally consistent adversarial training approach. Liu *et al.* [11] demonstrated that substituting

convolutional layers with partial convolutions and mask updates can achieve state-of-the-art inpainting results. Afterward, some methods [1], [7] employed two-level GAN to achieve a coarse-to-fine inpainting strategy.

For multispectral and HSIs inpainting, partial algorithms for grayscale and RGB images are still applicable, such as diffusion-based methods [21], [22] and exemplar-based methods [52], [53]. These kinds of spatial-based methods take advantage of the relationship between different pixels in the spatial dimension and do not require any other spectral and temporal auxiliary images, so it is impossible to reconstruct a large corrupted region. In addition, there are some spectral-based, temporal-based, and hybrid methods. Spectral-based methods refer to relevant information of other spectral data to reconstruct the corrupted region [18], [48], [54], [55]. The basic idea of this method is to estimate the relationship between the known regions and incomplete bands and then reconstruct the missing regions with this relationship. Due to the redundancy and correlation (RAC) of the remote sensing image, Kang et al. [58] used the tensor decomposition to obtain global RAC in the spectral domain. However, the corrupted region caused by detector damage or cloud cover affects all bands in remotely sensed images. In this case, the spectral-based methods fail in getting a promising result. Temporal-based methods replace the missing area with the same area of different periods [56], [57]. RAC characteristics also exist in the temporal domain. He et al. [59] introduced tensor decomposition into the analysis of time-series remote sensing datasets and proposed a tensor ring completion method for missing information reconstruction. However, such methods increase the complexity of data collection and consumed human and material resources. Some hybrid methods [22], [51] achieve better results to some extent. Ji et al. [51] proposed an improved tensor decomposition method to consider the RAC characteristics of spatial, spectral, and temporal domains while reconstructing the underlying patterns. However, the above problems have not been solved.

To date, by treating the HSI as 3-D cube data, many different algorithms use a 3-D CNN [60], [61] to automatically extract the spatial-spectral feature from HSIs. Li et al. [60] proposed a lightweight 3-D CNN and viewed the HSI cube data altogether without relying on any preprocessing or postprocessing, extracting the deep spectral spatial-combined features effectively. Sellami et al. [61] proposed a novel approach based on adaptive dimensionality reduction and semi-supervised 3-D CNN for the spatial-spectral feature extraction of HSIs. Although the GAN [28] accurate learning methods have been brilliant in the field of RGB image completion, they are not available to obtain large datasets for specific scenes due to the complexity of hyperspectral data collection. At the same time, due to the complex spectral relationship of HSI, using GAN to generate artificial data for learning a better inpainting model is also challenging. Ulyanov et al. [4] proposed that prior information can be obtained from the structure of the neural network itself to guide image inpainting. Oleksii and Hardeberg et al. [3] applied this method to HSI inpainting tasks and achieved good inpainting results subsequently.

A. Contribution

In order to solve the above problems, we propose an end-to-end inpainting framework of HSIs based on CNN to address the corrupted region of all bands. Our proposed method predicts large areas missing without any other spectral and temporal auxiliary images by a deep learning framework [2], [9], [28].

We take full advantage of the RAC characteristics in spectral channels and the texture richness in the spatial dimension. The main contributions can be generalized as follows.

1) The dependence of the HSI channels is fully utilized by a channel attention mechanism (CAM) of assigning different adaptive weights to different channels, to achieve the role of reducing the redundancy.
2) We believe that channel hyperspectral remote sensing images have rich texture features in the spatial dimension. A local discriminative network is added to determine the authenticity of the image in the corrupted regions for structural continuity, and a gradient consistency loss (GC-loss) function is designed to focus on the GC between generated HSI and ground truth.
3) Since there are no large public datasets for precise learning, we develop an HSI dataset named Feicheng Hyperspectral (FCH) containing a total of 3000 HSIs and 3 scenes including farmland, highway, and tenement scenes.

The rest of this article is organized as follows. Section II introduces the materials and the proposed architecture in detail. The experimental results of the proposed FCH dataset and the public dataset are provided in Section III. In Section IV, we discuss the results. Section V briefly summarizes the methodology of our work.

II. PROPOSED METHOD

Our method is based on the GAN [28] for HSI inpainting process, in which we simultaneously train two networks: a generative network G to capture the data distribution and a discriminative network D to estimate the probability that a sample comes from the training data or G. Once the training is stable, that is, D cannot judge whether the image is true or fake and G also cannot generate a more realistic image to deceive the D, we get the optimal generator G^*. However, due to the particularity of spectral and spatial information of HSI, adaptive improvement of ordinary GAN is needed. In order to make the network pay more attention to the bands with large structural differences, the CAM SE-Net is added by the method of [17] and assigns adaptive weights to different channels during the generative network training. In the discriminator section, similar to the method of [9], we use the global discriminative network to determine the overall characteristics of the image and the local discriminative network to focus more attention on the missing region. In order to ensure that the texture of the generated HSI is more authentic and abundant, we design a GC-loss function to constrain the GC between the generated HSI and the ground truth image. Finally, we combine the GC-loss with reconstruction loss, adversarial loss, and TV loss as the final
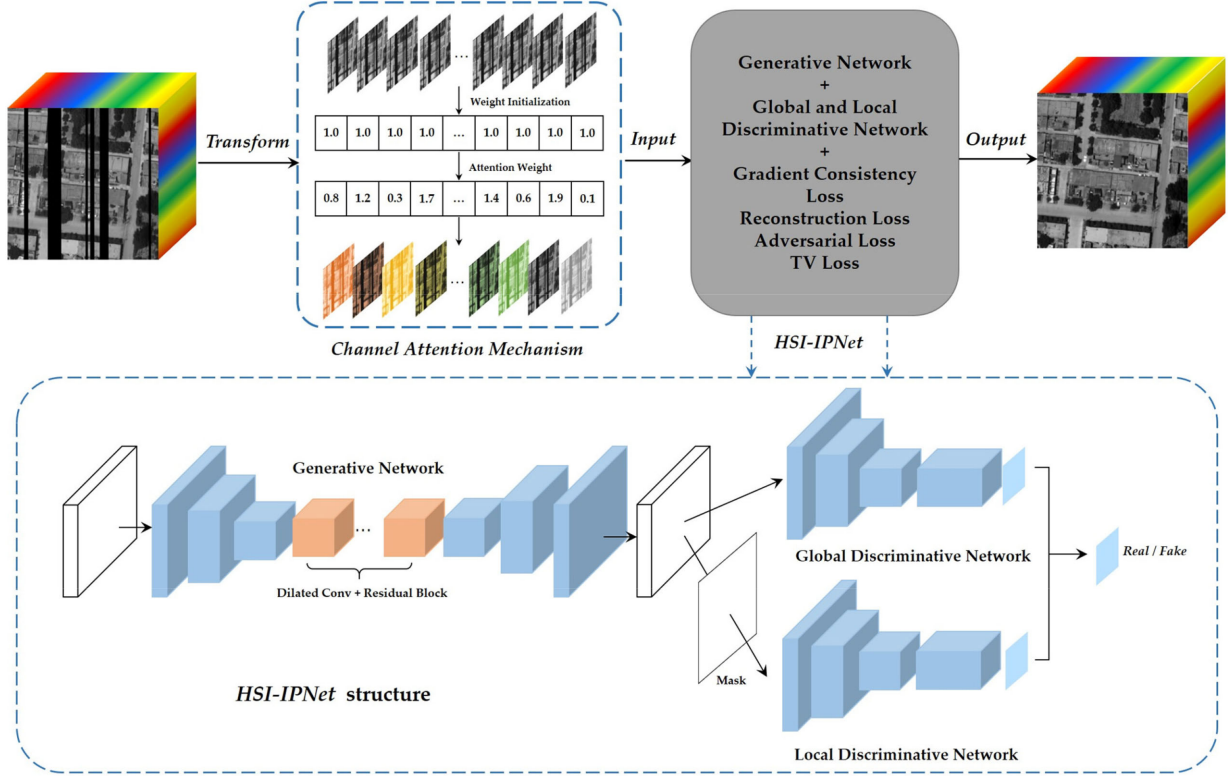
Fig. 1. Overall framework of the proposed HSI inpainting method.

optimization function to joint training. The overall framework of our inpainting system is shown in Fig. 1.

A. Generative Network

The input of generative network I_{in} is a $63 \times 256 \times 256$ HSI with some black missing parts. The structure of our network is similar to the generative network in [5], which is composed of the encoder part, middle part, and decoder part. In the encoder part, a 7×7 convolution kernel is used to obtain an initial feature map, on that basis, SE-Net is added to gain the weight values of each layer. Then, in order to generate nonblurred texture in the corrupted regions, we down-sample twice with 4×4 convolutions. Symmetrical on the encoder part, the decoder part up-samples images back to the original size. In the middle part, we have eight residual blocks with dilated convolutions [31] instead of regular convolutions to gain a larger receptive field, while still using the same amounts of parameters and computational power. An overview of the generative network model architecture can be seen in Table I.

CAM: The number of HSI bands can reach tens or even hundreds, and there is a strong correlation among adjacent channels. However, these very similar channels have no extra effect on HSI inpainting tasks, and repairing only one can achieve the goal of repairing multiple adjacent channels. By assigning different weight coefficients to the respective channels of the HSI, the purpose of eliminating channel redundancy can be achieved. We use SE-Net [17] to implement the CAM. In

TABLE I
ARCHITECTURE OF THE HSI GENERATIVE NETWORK

Type	Kernel	Dilation	Stride	Outputs
conv.	7×7	1	1×1	64
conv.	4×4	1	2×2	128
conv.	4×4	1	2×2	256
dilated conv.	3×3	2	1×1	256
dilated conv.	3×3	2	1×1	256
dilated conv.	3×3	2	1×1	256
dilated conv.	3×3	2	1×1	256
dilated conv.	3×3	2	1×1	256
dilated conv.	3×3	2	1×1	256
dilated conv.	3×3	2	1×1	256
dilated conv.	3×3	2	1×1	256
deconv.	4×4	1	2×2	128
deconv.	4×4	1	2×2	64
conv.	7×7	1	1×1	63

order to take advantage of global contextual information, the squeeze function maps global spatial information to 1-D channel descriptors.

Assuming that the squeeze function is represented by $\mathbf{F}_{\text{sq}}$, the input feature maps $U \in \mathbb{R}^{H \times W \times C}$, where

$$z_c = \mathbf{F}_{\text{sq}}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j) \qquad (1)$$

where $\mathrm{U} = [u_1, \ u_2, \ \ldots, \ u_C]$, $c = 1, \ 2, \ \ldots, \ C$, z_c is the cth element of $\mathbf{z}$. A rectified linear unit (ReLU) [30] function and

Type	Kernel	Stride	Outputs
conv.	4×4	2×2	64
conv.	4×4	2×2	128
conv.	4×4	2×2	256
conv.	4×4	2×2	512
conv.	4×4	2×2	1

a sigmoid activation with two fully connected (FC) layers are opted to fully capture channel-wise dependencies. Then we can write the activate output as $\mathbf{F}_{ex}$

$$s_c = \mathbf{F}_{ex}(z) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 z)) \tag{2}$$

where s_c is the cth element of $\mathbf{s}$. δ and σ refer to the ReLU function and sigmoid activation, respectively. $\mathbf{W}_1 \in \mathbb{R}^{C' \times C}$, $\mathbf{W}_2 \in \mathbb{R}^{C \times C'}$, and $C' = \frac{C}{r}$. The FC layer including dimensionality reduction process reflected in the setting of reduction ratio r, so as to reduce the parameters in the training course. The final output of the CAM can be described as

$$x_c = s_c u_c \tag{3}$$

where $\mathbf{X} = [x_1, \ x_2, \ \ldots, \ x_C]$ is the final weight vector.

B. Global and Local Discriminative Network

A global discriminative network and a local discriminative network have the objective of discerning whether an HSI is real or fake (HSI from generative network output). The networks are based on CNNs that squeeze the HSI into small feature maps. An overview of the networks can be seen in Table II.

The global discriminator takes as an input of the entire image rescaled to 256×256 pixels. Then, we use a 70×70 PatchGAN [5] [49] [50] architecture which consists of five convolutional layers with the size of the final outputs feature map is 70×70. All the convolutional layers, except the last one, employ a stride of 2×2 pixels to decrease the HSI resolution while increasing the number of output channels. The last layer keeps the space size constant and compresses the channel dimension to one. In contrast with the generative network, all convolutions use 4×4 kernels. The global discriminator looks at the entire image to assess if it is coherent as a whole, but cannot pay good attention to the consistency of the missing part. Thence, we design the local discriminator looks only at a small corrupted area at the completed region to ensure the local consistency of the generated patches. The local discriminator has the same pattern except the input is the missing part that needs to be fixed. The missing area of HSIs caused by instrument instability and slit contamination is usually striped, so we consider various striped masks during the training process. We also consider the issue of cloud cover and add a cloud-like irregular mask.

C. Loss Function

The HSI generative network uses the incomplete HSI $\tilde{x} = x \odot (1 - M)$ as input, where x refers to the complete HSI, and M refers to the binary mask with the value 1 inside regions to be filled-in and 0 elsewhere. $G(\cdot)$ denotes the output of the generative network and $D(\cdot)$ denotes the combined discriminative network in a functional form. In order to make the network generate more realistic HSI, four loss functions are jointly used: a proposed GC-loss to constraint gradient differences between generated images and ground-truth, a reconstruction (L_1) loss to capture the overall structure of the generative HSI, an adversarial loss to make the final network output look more realistic, and a TV loss refers to the gradient diffusion concept in the TV model [16] to make the adjacent pixel values of the output as similar as possible. We now introduce each loss function in detail.

GC-Loss: Gradient reflects high-frequency information of the image. For image inpainting tasks, especially remote sensing images, texture information reflected by gradient is very important. There are also some image inpainting methods or networks [5], [6] using edge information to assist image inpainting, but they use the GC of neighboring pixels to smooth the output image. We use the gradient loss function to focus on the required gradient information and maintain the consistency of the gradient between the generated images and the ground truth images. Then we can describe the gradient loss as L_{grad}

$$\mathcal{L}_{\text{grad}} = \|T(x) - T(G(x \odot (1 - M)))\|_1 \tag{4}$$

where $T(\cdot)$ represents the function used to extract and select gradients. In this article, we use the canny operator to achieve this part.

Reconstruction Loss: We use the L1 distance as the reconstruction loss function to measure the difference between the generator's output and the ground truth HSI, L_{rec}

$$\mathcal{L}_{\text{rec}}(x) = \|x - G(x \odot (1 - M))\|_1 \tag{5}$$

where $\odot$ is the element-wise product operation. Reconstruction loss is the most commonly used loss function for general image reconstruction. However, this loss function can only obtain the approximate structural information of the corrupted region, some high-frequency parts cannot be well reconstructed. Nevertheless, it is a good choice to use it to get the initial structural information.

Adversarial Loss: The adversarial loss reflects the adversarial process of GAN [28]. In order to learn the generative distribution $P_G(x)$, GAN proposed a discriminator D to continue adversarial learning with the aim to measure the difference between $P_G(x)$ and true distribution $P_{\text{data}}(x)$. We can find the optimal generator G^* as follows:

$$\min_G \max_D \mathbb{E}_{x \sim P_{\text{data}}(x)}[\log D(X)] + \mathbb{E}_{x \sim P_G(x)}[\log(1 - D(Z))] \tag{6}$$

where Z refers to the output of generator G. We first train the discriminator and learn the probabilities that belong to the real data or the false data, respectively. When this probability reaches the maximum, we can get the optimal discriminator D^*. Then we start training the generator, hoping that the fake data generated can fool the D^*. Fix the parameters of the discriminator, optimize the generator through back-propagation, and make the data obtained by the generator after the discriminator closer to one. Then we can describe the adversarial loss as L_{adv}

$$\mathcal{L}_{\text{adv}} = \max_D \mathbb{E}_{x \sim P_{\text{data}}(x)}[\log(D(x))$$

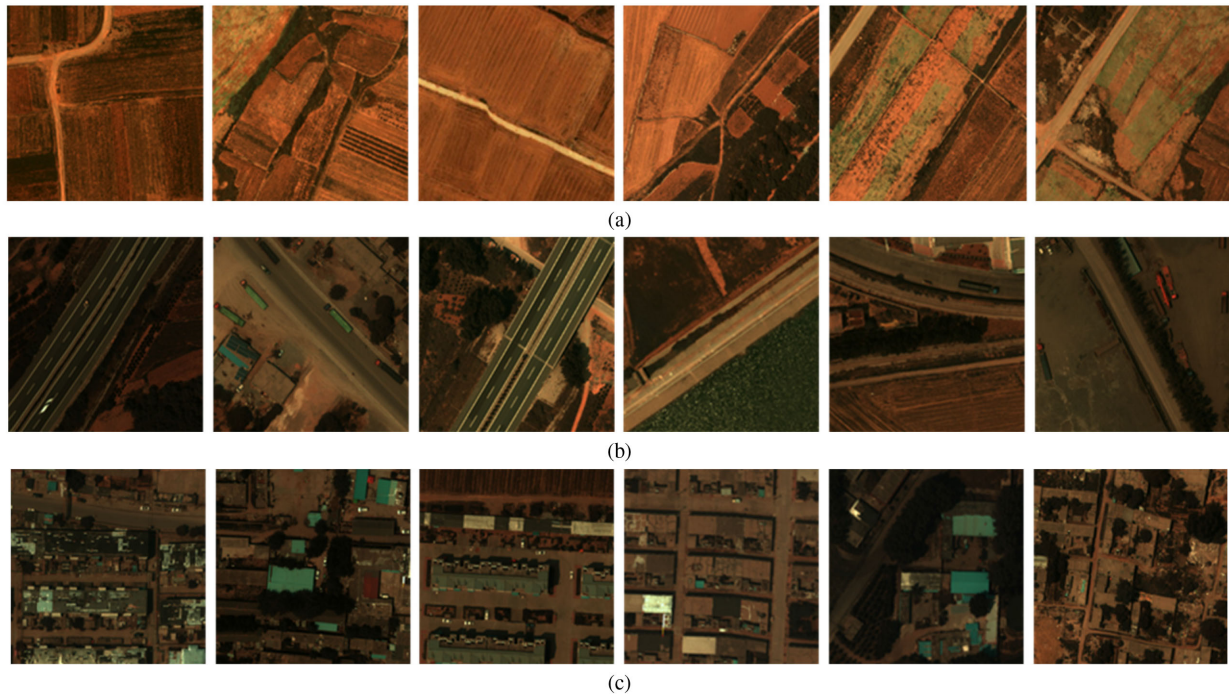$$+ \log(1 - D(G(x \odot (1 - M))))] . \tag{7}$$

Fig. 2. Class representatives of the FCH dataset: (a) farmland; (b) highway; (c) tenement.

We use optimizer Adam [38] to optimize G and D jointly. It is worth noting that the adversarial loss makes the overall generative images more realistic, not just in the corrupted regions.

TV loss: TV loss is evolved from the traditional image inpainting method TV model. It was originally used for denoising and deblurring tasks. However, due to the diffusion performance of the TV model, it can also be used in the field of image inpainting. We use the TV loss as follows:

$$\mathcal{L}_{\text{TV}} = \sum_{i,j} \left(x_{i,j+1} - x_{i,j}\right)^2 + \left(x_{i+1,j} - x_{i,j}\right)^2 \quad (8)$$

where i and j are the coordinate position of a pixel in the images. We use the TV loss to make the output image smoother, as well as increase the authenticity of the image.

Joint Loss: We define the overall loss function as

$$\mathcal{L} = \lambda_g \mathcal{L}_{\text{grad}} + \lambda_{\text{rec}} \mathcal{L}_{\text{rec}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}} + \lambda_{\text{TV}} \mathcal{L}_{\text{TV}}. \quad (9)$$

For our experiments, we choose $\lambda_{\text{rec}} = 1, \lambda_{\text{adv}} = 0.1, \lambda_{\text{TV}} = 1, \lambda_g = 5$.

III. RESULTS

A. Proposed Datasets

Since there is no large hyperspectral dataset for accurate training of HSI inpainting tasks, we collected a hyperspectral dataset on June 23, 2018, from Feicheng, Shandong Province, China, and named FCH dataset. The spectroscopic equipment used for acquisition is the new-generation airborne hyperspectral imager instrument of China, named Wide Swath and High-Resolution Airborne Pushbroom Hyperspectral Imager, which contains 63 spectral channels with 0.4–1.0 μm spectral resolution and the spatial resolution is about 12.5 cm per pixel. The size of the HSI we captured is 262748×10983. However, due to the camera shooting angle, there are some black borders above the image. After cropping these black areas, we retain valid areas. Subsequently, the big HSI is randomly cropped to a size of 256×256, and 3000 HSIs are obtained. These HSIs mainly include three scenes: farmland, highway, and tenement. As shown in Fig. 2, the pseudocolor image is formed with bands (5, 10, and 30).

B. Experimental Setup

The overall network architecture is designed using the Pytorch framework. On the FCH dataset, 80% of the samples are randomly selected as the training set and 20% of the samples are used as the test set. In order to expand the training samples, the method of random cropping is used to make the samples of each epoch different, so as to achieve the purpose of data augmentation. In our experiments, we use three types of image masks: single band, multiple bands, and irregular patterns. For the single band and multiple bands masks, the size is fixed and the width is equal to the width of the image. The single mask takes up 1/8 of the total image pixels with the random location.

Multiple masks have multiple stripes of different sizes. For the irregular mask, we use a cloud mask for the case of cloud cover.

Our network is trained 256×256 HSI as input with the joint loss function defined in (10). The hyperparameter settings during training are as follows. We use the stochastic gradient descent solver Adam [38] to optimize the model and set $\beta_1 = 0$ and $\beta_2 = 0.9$. The generative network G is trained with learning rate
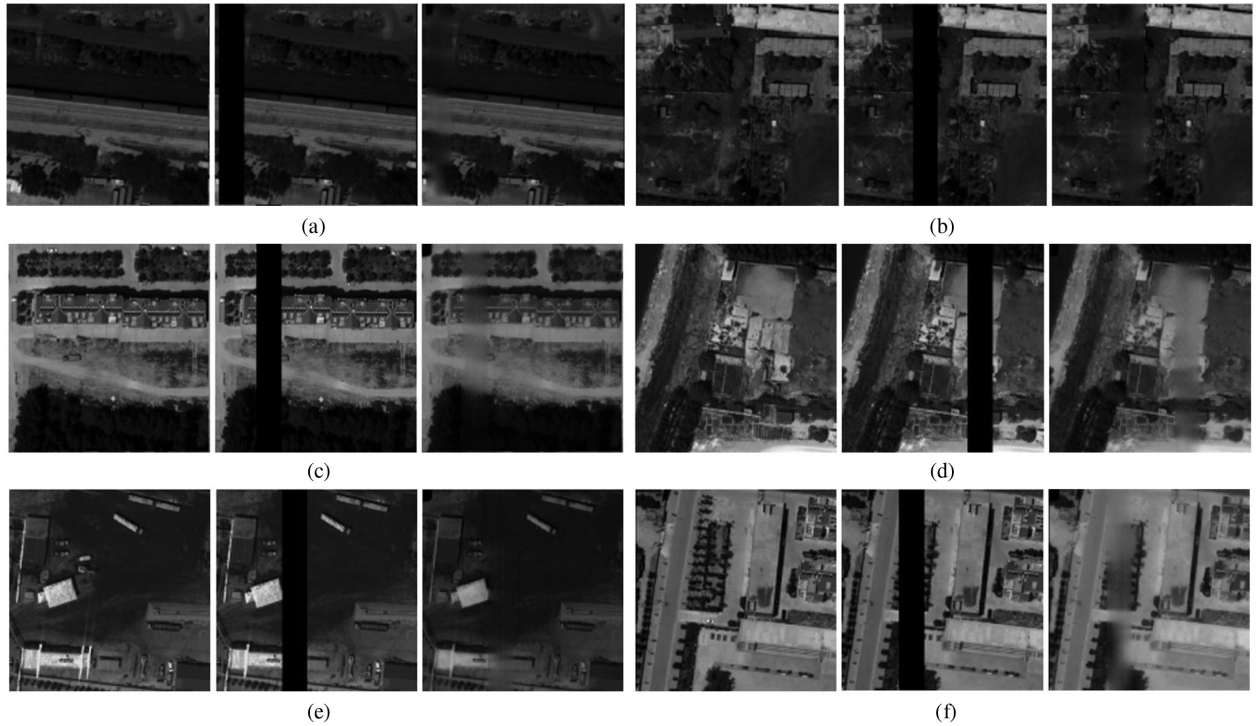
Fig. 3. Inpainting results for FCH dataset with random single band mask. For each subpicture like (a), from left to right are ground truth HSIs, generator input HSIs, and generator output HSIs.

10^{-3} which is ten times higher than the discriminative network D. The batch size and the maximum number of iterations are set as 8 and 30 000, respectively. It is worth noting that we use the pretrained inpainting model in the normal image from the Places2 dataset. When we add the SE-Net to assign channel weight, the pretrained inpainting model with no SE-Net model in the FCH dataset will be used. In all experiments, the filter weights of the generator and the discriminator are initialized by Gaussian distribution with 0 mean and 0.2 variances.

C. Comparative Experiment

First, we observe the experimental results on the developed FCH dataset, selecting 80% of the samples to train and 20% to test. Fig. 3 shows some experimental results under the random single mask. Fig. 4 shows some experimental results under the fixed multiple masks. Fig. 5 shows some experimental results under the fixed cloud mask. The experimental images are all from the proposed FCH dataset, and the 30th channel is taken for visualization. We observed that most of the corrupted regions can be well reconstructed, but there are still obvious borders and blurs on some images, as shown in Figs. 3(c) and 4(f). After analysis, we found that these blurred images usually have higher pixel values. For example, the average pixel of the ground truth image of the missing areas in Fig. 4(f) is 60, while the average pixel of the ground truth image of the missing part in Fig. 4(a) is only 29. We can easily observe that Fig. 4(f) has more information than Fig. 4(a) needed to be restored because of the initial value of the missing areas is zero. Overall, the inpainting results of multiple masks are better than the single

TABLE III
OVERALL QUANTITATIVE INDICATORS OF DIFFERENT METHODS ON INDIANS PINES DATASET

Methods	Input	Mumford-Shah	TV-H^{-1}	FastHyIn	**Ours**
MPSNR	17.75	24.74	27.68	28.08	**28.39**
MSSIM	0.722	0.890	0.911	0.920	**0.930**

mask. This is because the single band mask has a larger corrupted area compare with one single band in the multiple bands mask having smaller corrupted regions.

Then, we perform comparative experiments on the Airborne Visual Infrared Imaging Spectrometer (AVIRIS) Indians Pines dataset. The size of the dataset is $145 \times 145 \times 200$, which is collected by the AVIRIS with the imaging wavelength range, and the spatial resolution is about 0.4–2.5 μm and 20 m per pixel, respectively. Fig. 6 compares images generated by the proposed inpainting method with three traditional methods: Mumford-Shah [25], fourth-order TV (TV-H^{-1}) [39] 2-D methods and the state-of-the-art HSI inpainting method FastHyIn [18], and uses the result from [3]. To measure the quality of our results, we use two indicators: signal-to-noise ratio (PSNR) and structural similarity (SSIM). The specific quantitative indicators are described in Table III.

Although the proposed method has improved in quantitative indicators, the phenomenon of ambiguity still exists in Fig. 6(f). Through analysis, we believe that it is caused by the difference between training and testing data. The training set uses the proposed FCH dataset and the spatial resolution is 0.25 cm per
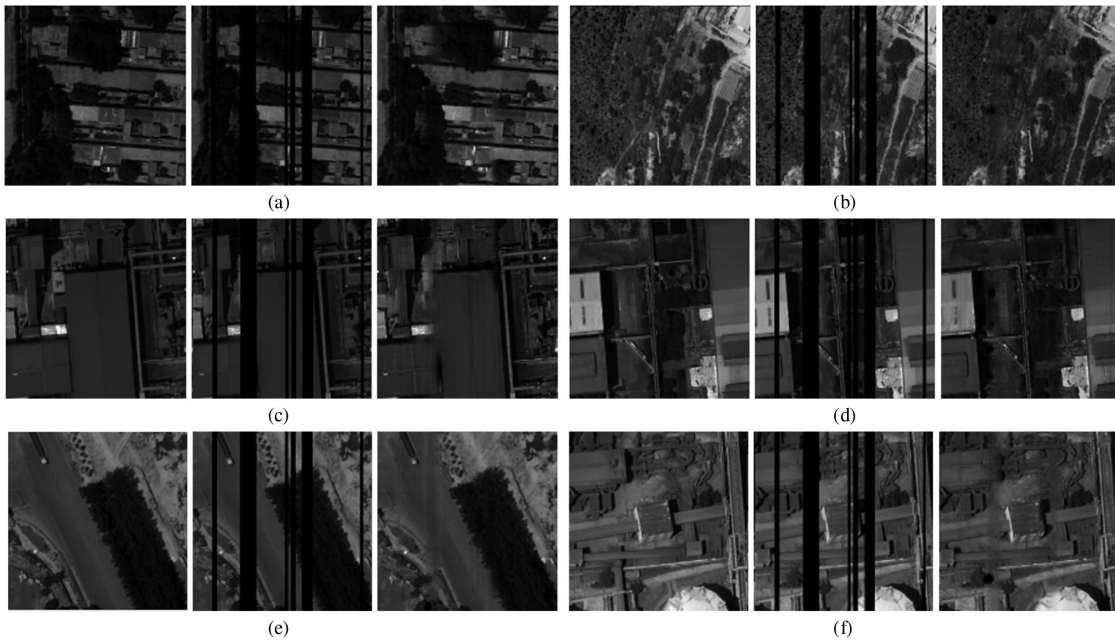
Fig. 4. Inpainting results for FCH dataset with fixed multiple bands mask. For each subpicture like (a), from left to right are ground truth HSIs, generator input HSIs, and generator output HSIs.
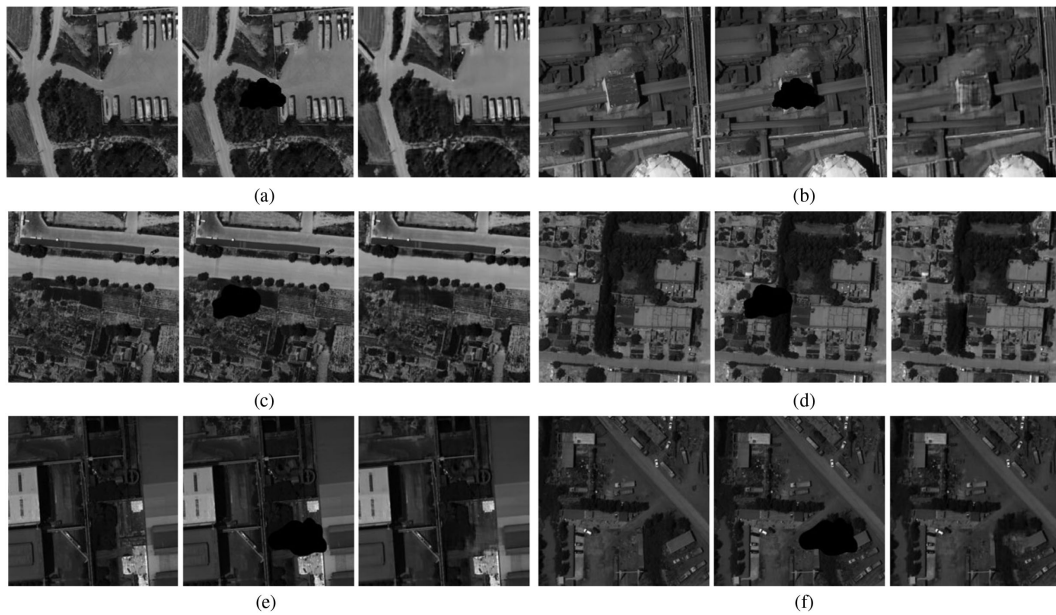

Fig. 5. Inpainting results for FCH dataset with cloud mask. For each subpicture like (a), from left to right are ground truth HSIs, generator input HSIs, and generator output HSIs.
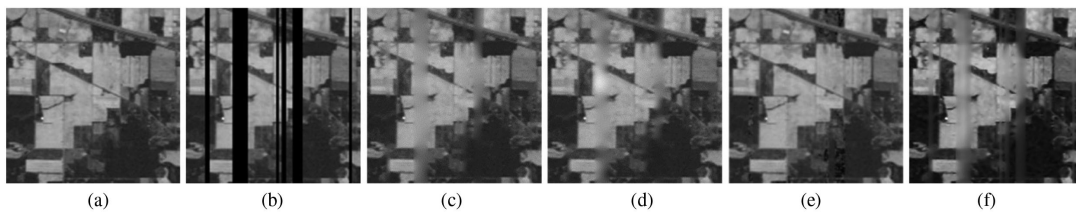

Fig. 6. Comparison of HSI inpainting results with existing models in AVIRIS Indian Pines dataset; band 150. (a) Ground truth. (b) Ground truth with mask. (c) Mumford-Shah [25]. (d)TV-H^{-1} [39]. (e) FastHyIn [18]. (f) Ours.
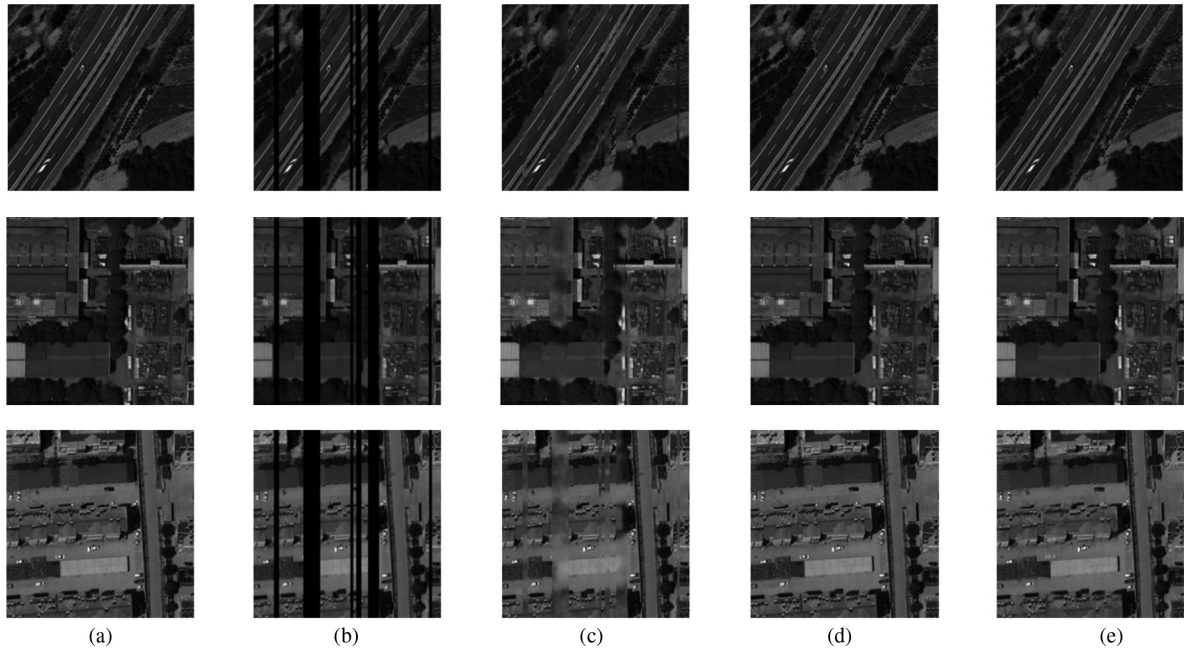
Fig. 7. Comparison of HSI inpainting results with existing models in FCH datasets; band 30. (a) Ground truth. (b) Ground truth with mask. (c) Deep-HS-prior 3D [3]. (d) (d) Deep-HS-prior 2D [3]. (e) Ours.

TABLE IV
OVERALL QUANTITATIVE INDICATORS OF DIFFERENT METHODS ON THE FCH DATASETS

Methods	Input	Deep-HSI-prior 3D	Deep-HSI-prior 2D	Ours
MPSNR	13.32	29.54	33.21	33.49
MSSIM	0.342	0.936	0.959	0.963

TABLE V
OVERALL QUANTITATIVE INDICATORS OF ABLATION EXPERIMENTS ON THE FCH DATASET

Methods	No-CAM	No-Local-D	No-GC-loss	All
MPSNR	20.94	32.65	29.18	33.49
MSSIM	0.748	0.943	0.922	0.963

pixel. Even though the problem of insufficient scene information caused by the high spatial resolution is considered and use 3 × 3 convolutional fuzzy processing, comparing to the 20 m per pixel spatial resolution of the AVIRIS Indians Pines dataset, the difference is still huge. There are also discrepancies between the training scene and the testing scene. We test on the AVIRIS Indian Pines public dataset and the image scene is urban data, but there are only about one hundred HSI of urban scenes in our developed dataset.

Finally, we compare the deep-HSI-prior 3-D and deep-HSI-prior 2-D methods proposed in [2] on the FCH dataset. The visual results are shown in Fig. 7, and the quantitative indicators are given in Table IV. It can be seen that our proposed method is better than the deep-HSI-prior methods on the proposed FCH dataset.

D. Ablation Experiments

To evaluate the effectiveness of our proposed method, ablation experiments are conducted by using no CAM, no local discriminator, and no GC-loss function on the proposed FCH dataset. When we add the CAM, a network with no CAM should be pretrained. Table V shows all experimental results and we can draw the following conclusions.

1) Results from no no-CAM experiments are the worst because of the RAC of HSI. HSI have many spectral channels, for example, 63 channels in the proposed FCH dataset and 200 channels in the AVIRIS Indians Pines public dataset. Such multichannel characteristics not only increase the distinguishability of materials but also generate a certain degree of redundancy. Therefore, it is very important to establish the dependence relationship between channels for the HSI.
2) If the local discriminative network is not added, the experimental results are slightly reduced. Simplify experiments without the local discriminative network as No-Local-D. The reason is that the global discriminator looks at the entire image to assess if it is coherent as a whole, but cannot pay good attention to the consistency of the missing part.
3) The results of no-GC-loss function experiments prove that the validity of the texture consistency idea. Since HSI are a type of remote sensing images with rich scene information, the texture information represented by the gradient is very important for the inpainting task of HSI.
4) Our proposed method achieves the best performance compared to the three situations mentioned above, which is a result of combining CAM, local discriminator, and gradient consistency loss function (All). The CAM adaptively

TABLE VI
INVESTIGATE THE EFFECT OF DIFFERENT POOLING LAYERS IN THE SQUEEZE
OPERATOR ON THE RESULTS OF HSI INPAINTING

Squeeze	Average-pooling	Max-pooling
MPSNR	33.28	**33.49**
MSSIM	0.960	**0.963**

TABLE VII
INVESTIGATE THE EFFECT OF DIFFERENT POOLING LAYERS IN THE EXCITATION
OPERATOR ON THE RESULTS OF HSI INPAINTING

Excitation	ReLU	Tanh	Sigmoid
MPSNR	32.98	33.28	**33.49**
MSSIM	0.950	0.959	**0.963**

TABLE VIII
OVERALL QUANTITATIVE INDICATOR OF DIFFERENT NUMBER OF SE-NET
WITH THE TEST RESULTS

Numbers of the SE-Net	0	1	2	3
MPSNR	20.94	33.34	33.41	**33.49**
MSSIM	0.748	0.960	0.961	**0.963**
Params	47.2M	**50.1M**	52.5M	54.2M

The bold values in represents the optimal result.

selects the required channels and establishes the dependency relationship between channels. The local discriminative network strengthens the judgment on the authenticity of the image of the corrupted region. The gradient consistency loss function guarantees the consistency of the corrupted region texture between the generated image and the ground truth.

IV. DISCUSSION

In this section, we mainly explore how to maximize the benefits of SE-Net in the generative network and draw corresponding conclusions. In all experiments, the training set and the testing set use the developed FCH dataset to analyze the above factors. Details are as follows.

A. Evaluation of Squeeze Operator

The squeeze operator squeezes a 3-D image cube into a 1-D channel descriptor and the pooling layer plays an important role. The result of pooling is to reduce features and parameters, but the purpose of pooling is not only that. The purpose of pooling is to maintain some invariance, such as rotation, translation, scaling, etc. Average-pooling and max-pooling are commonly used. Average-pooling can reduce the error caused by the increase in the variance of the estimated value and the limited size of the neighborhood, which retain more background information of the image. Max-pooling can reduce the bias of the estimated mean caused by the error of the convolution layer parameters and retain more texture information. We study the effect of these two pooling layers on the image inpainting results, and the quantitative analysis results are shown in Table VI. While both the max-pooling and the average-pooling are effective, the performance of the max pool is slightly better, which shows that it is reasonable to choose it as the basis of squeeze operation. However, we also notice that the performance of the SE block is quite robust to the choice of a particular aggregation operator.

B. Evaluation of Excitation Operator

Next, we evaluate the nonlinear selection of the excitation operator. The role of the excitation operator is to assign attention weights and establish a dependency relationship between channels. The operator is mainly implemented by a nonlinear activation function. In this section, we explore the effects of Sigmoid function, Tanh function, and ReLU function on the results of HSI inpainting. The quantitative analysis results are reported in Table VII. We can see that replacing sigmoid with the Tanh intersection will slightly reduce the performance while using ReLU will significantly worsen and actually cause the

inpainting performance to drop below the baseline (Deep-HS-prior 2D [3]). This suggests that the excitation operator must be carefully selected to make the SE block valid.

C. Evaluation of the Number of SE-Net

We explore the influence of the different numbers of SE-Net blocks by integrating SE-Net blocks into the generator G. Specifically, we add SE-Net blocks to the encoder stage: stage_1, stage_2, and stage_3, and the quantitative analysis results in Table VII. We observe that the quality of HSI is steadily improved as the number of SE-Net blocks is gradually increased. In the stage_1, the main role of SE-Net block is to eliminate the redundancy on the HSI spectrum and establish the dependency between channels. By comparing the second column with the third column in Table VIII, we can observe the obvious improvement of the picture quality evaluation indicators. At the same time, it is also proved the necessity to establish a dependency relationship between channels for multichannel images such as HSI. In the stage_2 and stage_3, the SE-Net blocks work on the feature maps generated by the forward propagation of the CNN and establish the dependency relationship on the channel dimensions of the feature maps. It can be seen from the fourth and fifth columns of Table VIII that with the increasing of the number of SE-Net layers introduced, the quality of HSI is also continuously improved. We believe that the benefits of SE blocks at different stages are complementary and can be effectively combined to further improve network performance. Both MPSNR and MSSIM indicators have obvious upward trends. The third row in Table VIII shows that with the increase of SE-Net blocks, the parameters of the network are also increasing. Hence, it is needed to make a tradeoff between performance and efficiency. It may not be optimal to use the same ratio in the entire network (due to the different roles played by different layers) in practice. In order to reduce the complexity of network parameters, we reduce the layer by layer four times the ratio to meet the needs of the infrastructure.
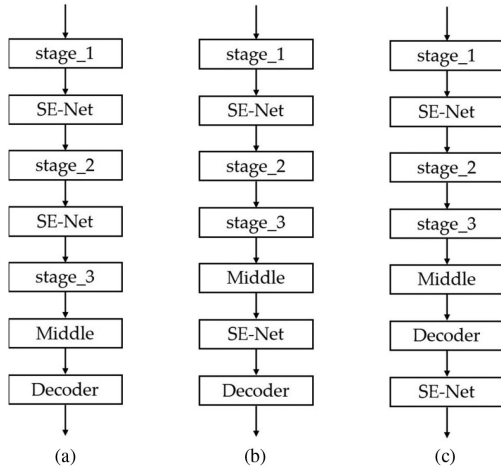
Fig. 8. SE-Net block location designs explored in the discussion.

TABLE IX
OVERALL QUANTITATIVE INDICATOR OF DIFFERENT POSITION OF SE-NET
WITH THE TEST RESULTS ON THE FCH DATASETS

Position of the SE-Net	Initial	Encoder	Middle	Decoder
MPSNR	33.40	**33.49**	33.39	33.25
MSSIM	0.962	**0.963**	0.958	0.956

The bold values in represents the optimal result.

D. Evaluation of the Position of SE-Net

We remark that the performance of the network is improved with the increase in the number of SE-Net, which cannot only establish the dependence relationship between the spectral channels of HSIs but also establish the channel connection of the 3-D feature map after convolution. Since it can establish the dependency relationship between the feature maps, learning which part (our generative network consists of Encoder, Middle, and Decoder) of the generative network to add is also important. Next, we study the impact of SE-Net on network performance in various parts of the generation network. In the experiment, SE-Net is added to the initial layer in each subexperiment, and then a comparison test is performed at the end of the encoder part, the end of the middle part, and the end of the decoder part. The specific experimental method is shown in Fig. 8 and the results are reported in Table IX. While all kinds of experiments are effective, the performance of the Encoder is slightly better and the performance of the decoder is slightly worse. However, we also notice that the performance of the SE block is quite robust to the choice of specific placement.

V. CONCLUSION

Our proposed method uses a deep learning framework to predict large areas missing without any other spectral and temporal auxiliary images. We design an end-to-end inpainting framework HSI-IPNet and make full use of both spectral characteristics and spatial information. For spectral characteristics, we employ SE-Net to the adaptive extraction of the spectral dimension, reducing the redundancy of hyperspectral channels

and modeling the correlation between channels. For spatial information, a local discriminative network is introduced in the discriminator to determine the structural continuity of the corrupted region images, and a gradient consistency loss function to focus the texture consistency between the generated images and the ground truth images.

We test with three kinds of masks and prove that HSI-IPNet can solve the lack of various stripes types caused by instrument instability and slit contamination, as well as the lack of irregular graphics caused by cloud occlusion. To test the performance of our method, we experiment on the AVIRIS Indians Pines public dataset and the proposed FCH dataset. Extensive experimental results consistently show that our architecture is superior to current state-of-the-art methods. In future work, we will pay more attention to the fusion of HSI with different spectral resolutions and the methods to expand HSI datasets, such as using style transfer to expand the limited HSI datasets.

REFERENCES

[1] H. Zhang, Z. Hu, C. Luo, W. Zuo, and M. Wang Zhang, "Semantic image inpainting with progressive generative networks," in *Proc. 26th ACM Int. Conf. Multimedia*, 2018, pp. 1939–1947.

[2] D. Pathak, K. Philipp, D. Jeff, D. Trevor, and A. E. Alexei, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern. Recognit.*, 2016, pp. 2536–2544.

[3] S. Oleksii and J. Y. Hardeberg, "Deep hyperspectral prior: Denoising, inpainting, super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019.

[4] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vis. Pattern. Recognit.*, 2018, pp. 9446–9454.

[5] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, "Edge connect: Generative image inpainting with adversarial edge learning," 2019, *arXiv:1901.00212*.

[6] W. Xiong *et al.*, "Foreground-aware image inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern. Recognit.*, 2019, pp. 5840–5848.

[7] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 4471–4480.

[8] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE Conf. Comput. Vis. Pattern. Recognit.*, 2018, pp. 5505–5514.

[9] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," in *Proc. ACM Trans. Graph.*, 2017, pp. 1–14.

[10] C.-H. Lin, K.-H. Lai, Z.-B. Chen, and J.-Y. Chen, "Patch-based information reconstruction of cloud-contaminated multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 163–174, Jan. 2014.

[11] G. Liu, F.A. Reda, K.J. Shih, and T.C. Wang, "Image inpainting for irregular holes using partial convolutions," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 85–100.

[12] J. Chen, Y. Shao, H. Guo, W. Wang, and B. Zhu, "Destriping CMODIS data by power filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 2119–2124, Sep. 2003.

[13] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.

[14] L. Wang, J. J. Qu, X. Xiong, X. Hao, Y. Xie, and N. Che, "A new method for retrieving band 6 of Aqua MODIS," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 2, pp. 267–270, Apr. 2006.

[15] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Berlin, Germany: Springer, Nov. 2006, vol. 147.

[16] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1, pp. 89–97, Jan. 2004.

[17] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern. Recognit.*, 2018, pp. 7132–7141.

[18] 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.

[19] H. Shen, X. Li, L. Zhang, and D. Tao, "Compressed sensing-based inpainting of Aqua moderate resolution imaging spectroradiometer band 6 using adaptive spectrum-weighted sparse Bayesian dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 894–906, Feb. 2014.

[20] 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.

[21] P. Addesso *et al.*, "Hyperspectral image inpainting based on collaborative total variation," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2017, pp. 17–20.

[22] Q. Cheng, H. Shen, L. Zhang, and P. Li, "Inpainting for remotely sensed images with a multichannel nonlocal total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 175–187, Jan. 2014.

[23] K. Degraux, V. Cambareri, L. Jacques, B. Geelen, C. Blanch, and G. Lafruit, "Generalized inpainting method for hyperspectral image acquisition," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 315–319.

[24] F. Bousefsaf, M. Tamaazousti, S.H. Said, and R. Michel, "Image completion using multispectral imaging," *IET Image Process.*, vol. 12, no. 7, pp. 1164–1174, Jul. 2018.

[25] S. Esedoglu and J. Shen, "Digital inpainting based on the Mumford-Shah-Euler image model," *Eur. J. Appl. Math.*, vol. 13, no. 4, pp. 353–370, Sep. 2002.

[26] 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.

[27] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

[28] I. Goodfellow *et al.*, "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 41, pp. 2672–2680, 2014.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern. Recognit.*, 2016, pp. 770–778.

[30] 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.

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

[32] F. Yu and V. Koltun, "Multi-Scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Representations*, 2016, pp. 1–13.

[33] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, pp. 259–268, 1992.

[34] Y.L. Chen, C. T. Hsu, and H. Y. M. Liao, "Simultaneous tensor decomposition and completion using factor prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 577–591, Mar. 2014.

[35] C. Barnes, E. Shechtman, and A. Finkelstein, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, Jul. 2009, Art.no. 24.

[36] J. B. Huang, S. B. Kang, N. Ahuja, and J. Kopf, "Image completion using planar structure guidance," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–10, Jul. 2014.

[37] K. He and J. Sun, "Image completion approaches using the statistics of similar patches," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2423–2435, Dec. 2014.

[38] D.P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–13.

[39] A. Bertozzi and C.B. Schönlieb, "Unconditionally stable schemes for higher order inpainting," *Commun. Math. Sci.*, vol. 9, no. 2, pp. 413–457, 2011.

[40] S. Chabrillat, P.C. Pinet, and G. Ceuleneer, "Ronda peridotite massif: Methodology for its geological mapping and lithological discrimination from airborne hyperspectral data," *Int. J. Remote Sens.*, vol. 21, no. 12, pp. 2363–2388, Nov. 2010.

[41] P. Kempeneers, B. Deronde, and L. Bertels, "Classifying hyperspectral airborne imagery for vegetation survey along coastlines," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2004, vol. 2, pp. 1475–1478.

[42] B. C. Gao, M. J. Montes, Z. Ahmad, and C. O. Davis, "An atmospheric correction algorithm based on vector radiative transfer modeling for hyperspectral remote sensing of ocean color," *Proc. SPIE*, vol. 3753, pp. 70–78, Oct. 1999.

[43] B. Siegmann, T. Jarmer, T. Selige, and H. Lilienthal, "Using hyperspectral remote sensing data for the assessment of topsoil organic carbon from agricultural soils," *Proc. SPIE*, vol. 8531, Oct. 2012, Art. no. 85312C.

[44] M.C. Wang, Y. N. Wang, and S. B. Chen, "Atmospheric correction research on vegetation patterns of multi-angle hyperspectral CHRIS/Proba data," *J. Jilin Univ.*, vol. 41, no. 2, pp. 609–614, 2011.

[45] R. J. Ellis and P.W. Scott, "Evaluation of hyperspectral remote sensing as a means of environmental monitoring in the St. Austell China clay (kaolin) region, Cornwall, UK," *Remote Sens. Environ.*, vol. 93, no. 1–2, pp. 118–130, Oct. 2014.

[46] F. Melgani, "Contextual reconstruction of cloud-contaminated multitemporal multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 442–455, Feb. 2006.

[47] A.B. Salberg, "Land cover classification of cloud-contaminated multitemporal high-resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 377–387, Jan. 2011.

[48] H. Shen, X. Li, L. Zhang, D. Tao, and C. Zeng, "Compressed sensing-based inpainting of Aqua moderate resolution imaging spectroradiometer band using adaptive spectrum-weighted sparse Bayesian dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 894–906, Feb. 2014.

[49] P. Isola, J.Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern. Recognit.*, 2017, pp. 1125–1134.

[50] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.

[51] Y. Ji, N. Yokoya, X. Zhu, and T. Huang. "Nonlocal tensor completion for multitemporal remotely sensed images' inpainting," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3047–3061, Jun. 2018.

[52] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.

[53] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1033–1038.

[54] P. Rakwatin, W. Takeuchi, and Y. Yasuoka, "Restoration of Aqua MODIS band 6 using histogram matching and local least squares fitting," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 2, pp. 613–627, Feb. 2009.

[55] H. Shen, C. Zeng, and L. Zhang, "Recovering reflectance of Aqua MODIS band 6 based on within-class local fitting," *IEEE J. Sel. Top ics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 185–192, Mar. 2011.

[56] C. Zeng, H. Shen, M. Zhong, L. Zhang, and P. Wu, "Reconstructing MODIS LST based on multitemporal classification and robust regression," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 512–516, Mar. 2015.

[57] J. Wang, P. A. Olsen, A. R. Conn, and A. C. Lozano, "Removing clouds and recovering ground observations in satellite image sequences via temporally contiguous robust matrix completion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun.c 2016, pp. 2754–2763.

[58] J. Kang, Y. Wang, M. Schmitt, and X. X. Zhu, "Object-based multipass InSAR via robust low-rank tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3062–3077, Jun. 2018.

[59] W. He, N. Yokoya, L. Yuan, and Q. Zhao, "Remote sensing image reconstruction using tensor ring completion and total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8998–9009, Jul. 2019.

[60] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, Jan. 2017, Art. no. 67.

[61] A. Sellami, M. Farah, I. R. Farah, and B. Solaiman, "Hyperspectral imagery classification based on semi-supervised 3-D deep neural network and adaptive band selection," *Expert Syst. Appl.*, vol. 129, pp. 246–259, Apr. 2019.
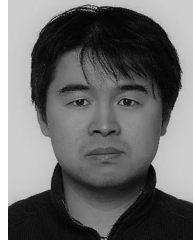
Rong Wong received the B.E. degree in electronic information engineering from Shanghai University, Shanghai, China, in 2018. She is currently working toward the M.E. degree in signals and information processing at Shanghai University, Shanghai, China.

Her research interests include hyperspectral image processing.

Zhijiang Zhang received the B.E degree from Harbin University of Science and Technology, Heilongjiang, China, in 1991, the M.E. and Ph.D. degrees from Harbin University of Technology, Heilongjiang, China, in 1996 and 1999, respectively.

Since 1999, he has been teaching with the Communication Engineering Department of Shanghai University and is currently a Professor. His research interests include image processing, visual inspection, digital holography and 3D display.

Dr. Zhang is the Director of Measurement Instrument Professional Committee of China Measurement and Test society.

Fansheng Chen received the B.E. degrees in optoelectronic information engineering from the Shandong University, Shandong, China, in 2002, and the Ph.D. degree in physical electronics from Shanghai institute of the University of Chinese Academy of Sciences, Beijing, China, in 2007.

He is currently a Research Professor with the Shanghai Institute of Technical Physics, Chinese Academy of Sciences. He has a responsibility for research and development of the infrared staring camera. His research interests include intelligent, infrared photoelectric, autonomous system, high time sensitivity, polarization remote sensing, and space-based infrared limit detection overall technology.

Yueming Wang received the B.E. degree in mechanical engineering from Tsinghua University, Beijing, China, in 2000, and the Ph.D. degree in electronic science and technology from the University of Chinese Academy of Sciences, Beijing, China, in 2005.

He is currently a Research Professor with the Shanghai Institute of Technical Physics, Chinese Academy of Sciences. He has a responsibility for research and development of high-performance imaging spectrometers. His research interests include the system design and simulation of infrared and hyperspectral imaging system.

Dan Zeng received the B.S. degree in electronic science and technology and Ph.D. degree in circuits and systems from the University of Science and Technology of China, Hefei, China, in 2003 and 2008, respectively.

Since 2008, she has been teaching with the Communication Engineering Department of Shanghai University and is currently a Professor. From 2001 to 2004, she was a Visiting Scholar with the University of Texas, San Antonio, School of Computer Science, and the National Scholarship Council. Her research interests include computer vision, machine learning, and multimedia technology.