Hyperspectral Remote Sensing Imagery Generation From RGB Images Based on Joint Discrimination

Liqin Liu ⓘ, Sen Lei ⓘ, Zhenwei Shi ⓘ, *Member, IEEE*, Ning Zhang, and Xinzhong Zhu

Abstract—Spatial resolution and spectral resolution both play an important role in the recognition of objects in hyperspectral remote sensing. However, the imaging characteristics of hyperspectral images (HSIs) result in a mutually restrictive relationship between the spatial and spectral resolutions. Generative adversarial networks (GANs) have achieved significant success in image generation. The introduce of the discriminators plays a key role in improving the reality. In this article, we propose an RGB to multiband hyperspectral imagery (150 bands) generation method based on GAN (R2HGAN). The method solves the high ill-posed problem and introduces high spectral resolution into RGB images by learning from multiple scenes of HSI. In R2HGAN, we extend the adversarial learning from spatial to spectral dimensions and joint discrimination is designed to generate HSIs closer to the real ones, where two discriminators (the conditional D and the spectral D) are put forward to supervise the spectral similarity and the conditional reality of the HSI jointly. In detail, the conditional discriminator comprehensively judges the quality of each area in the reconstructed HSI. At the same time, to ensure that the generated spectra are close to the real ones, a spectral discriminator based on multilayer perceptron is designed. Through the experiments on GF-5 imagery, the method has significantly improved the quality of the generated images over other state-of-the-art methods.

Index Terms—Generation adversarial network (GAN), hyperspectral image (HSI), remote sensing, spectral superresolution (SSR).

I. INTRODUCTION

THE hyperspectral image (HSI) is a 3-D cube, which contains spatial and spectral information at the same time. Compared with RGB and multispectral remote sensing images (MSIs), HSIs' high spectral resolution is of great significance for identifying the diagnostic spectrum of ground objects. Therefore, HSIs have been widely applied on precision agriculture [1], ecological sciences [2], mineralogy [3], and other fields. In the past 30 years, HSI processing technology has made great progress, including HSI classification [4]–[7], target detection [8], [9], anomaly detection [10], hyperspectral unmixing [11]–[13], etc. [14]. Among the aforementioned techniques, spatial information is widely used and plays an important role. Therefore, we hope to get imagery with both high spatial and spectral resolutions.

In remote sensing imaging, the spectral and spatial resolutions restrict each other under the same imaging time due to the architectural constraints of the imaging system [15]. Many RGB or MSI images with high spatial resolution usually lack spectral information. For example, Chinese Gaofen-1 (GF-1) with spatial resolution of 8 m/pixel has four bands covering wavelength $0.45-0.89$ μm [16] and Gaofen-2 provides four band MSI with spatial resolution of 4 m/pixel. Landsat 8 has seven bands with spatial resolution of 30 m/pixel covering the wavelength range of $0.433-2.3$ μm [17]. Besides, hyperspectral imagery mainly with a low spatial resolution. For satellite HSI, the spatial resolution is even lower than 20 m/pixel, such as the resolution of Chinese GaoFen-5 (GF5) [18] and Earth Observation-1 Hyperion [19] are both 30 m/pixel. At present, it is difficult to obtain remote sensing images with high spatial and spectral resolution at the same time.

Since the spatial resolution and the spectral resolution are very important for HSIs, and the imaging equipment limits the simultaneous acquisition of good spatial and spectral resolutions, some postacquisition methods are widely used [19], [20]. That is, through enhancement techniques to improve the spatial resolution [20] of HSI or to increase the spectral resolution of images with a higher spatial resolution [19] (such as RGB and MSI).

A typical way to improve the spatial resolution of HSI is hyperspectral superresolution (HSI SR) [20], [21]. The SR methods can be classified into two representative categories, one is SR only use the low-resolution HSI (LR-HSI) [22], [23], another contains the methods that enhance the spatial resolution of HSI with the spatial information in high-resolution RGB or multispectral images (MSI) [24], [25]. The HSI SR can improve the spatial resolution on the basis of existing HSI.

Unlike HSI superresolution (SR), spectral SR (SSR) promotes the spectral resolution of the image, which has attracted significant attention in recent years. SSR can provide high spatial resolution HSI only from a multispectral or RGB image without the use of hyperspectral sensors, which saves a lot of imaging hardware. Some SSR methods are designed to increase the channels of the MSI [26], [27]. Besides, a typical SSR application is spectral reconstruction from RGB images [28], [29].

Liqin Liu, Sen Lei, and Zhenwei Shi are with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, also with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China, and also with the Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China (e-mail: liuliqin@buaa.edu.cn; senlei@buaa.edu.cn; shizhenwei@buaa.edu.cn).

Ning Zhang and Xinzhong Zhu Shanghai Aerospace Electronic Technology Institute, Shanghai 201109, China (e-mail: dzs_zhangning@163.com; zhuxz@126.com).

The SSR methods from RGB images can provide HSI only with the RGB input, which makes SSR unique advantages in acquiring HSIs with high spatial and spectral resolution at the same time. However, there are two main shortcomings of the existing remote sensing HSI reconstruction methods, which are as follows.

1) The methods are usually trained and tested on a single scene, lacking a large amount of data support, and cannot reflect the universal mapping between RGB and spectrum of different ground objects. Besides, they usually recover HSI from more than four bands. For example, Mei *et al.* recover HSI from four or six bands [26] use the 3-D CNN on PaviaU or Cuprite dataset only in one scene. In [27], several standard benchmark datasets, which are single scene images or their subsets, are adopted.

2) Most methods use CNN and other structures to constrain the pixelwise distance between the generated image and the real image, MSE or $\mathcal{L}_1$ loss is usually adopted. Mei *et al.* adopt $\mathcal{L}_1$ loss to train networks [26]. Arun *et al.* use multiloss functions, including MSE loss and SAM loss [27]. In [30], Gewali *et al.* use 1-D CNN for single-pixel SSR and propose a new spectral loss function to constrain the Euclidean distances and first and second derivatives between the generated HSI and the real one. It is easy to cause oversmoothing or even some generation that does not meet the actual situation. The framework affects diversity and reality of the spectra.

To overcome these problems, we propose a generative adversarial network (GAN) based framework to generate HSI from RGB. The model can export 150 bands HSI with RGB input without the need for any HSI. It is trained on multiple scenes GF-5 HSI, thus it provides the nonlinear mapping of RGB to spectrum. In addition to the traditional minimize pixelwise error and keep fitting, a GAN-based framework and joint discrimination by two discriminators are devised. One is the conditional discriminator based on PatchGAN [31] to distinguish true or false of multiple regions in the HSI. Another is spectral discriminator based on multilayer perceptron (MLP) to determine whether the spectra sampled from HSI are true. Meanwhile, the random global uniform sampling (RGUS) technique is applied for sampling a limited number of spectra from HSI as the overall expression of it. The experimental results show that R2HGAN outperforms other state-of-the-art methods on various indicators.[1]

In summary, the main contributions can be summarized as follows.

1) We provide a general framework for RGB to HSI (R2HGAN) and experiment it on multiple scenes. The framework reveals the general mapping between RGB values of each pixel and its corresponding spectrum in large amounts of data. Meanwhile, it is based on GAN to avoid overfitting and abnormal generation.

2) We extend the adversarial learning from spatial to spectral dimensions. In addition to the traditional existing conditional discriminator, we design a spectral discriminator based on MLP together with the conditional

discriminator to generate HSIs closer to the real ones by joint discrimination. Two discriminators separately supervise the conditional and spectral reality of the generated HSI, and the generator can export more realistic HSI on spectral and spatial.

The rest of this article is organized as follows. Section II introduces the related works on HSI reconstruction and GAN, and Section III details the proposed method. Section IV provides experimental evaluations of the proposed method on the dataset. Finally, Section V concludes this article.

II. RELATED WORK

In this section, we introduce the two HSI reconstruction methods, including HSI SR and SSR from RGB images. Meanwhile, a brief introduction of GAN and its use on image generation, remote sensing image processing is exhibited. Various indicators used for evaluating the quality of reconstruction are demonstrated at last.

A. HSI Superresolution

The HSI SR mainly includes two types of methods: SR from LR-HSI only and fusion of high spatial resolution RGB/MSI and low spatial resolution HSI.

Akgun *et al.* [20] represent different wavelengths as weighted linear combinations of a small number of aliased and blurred basis image planes and solve the SR problem by a set-theoretic method. Similarly, many research works use the interband information for SR [22], [32]. In [32], Hu *et al.* use very deep CNN for keybands, SDCNN for other bands, and spatial-error-correction model to correct the spatial error. Intrafusion network [22] utilizes spatial–spectral information and reconstructs high-resolution HSI (HR-HSI) directly. Hu *et al.* [23] propose a 3-D-CNN-based method that integrates multiscale features for SR. Many GAN-based methods introduce the discriminator to judge the reality of HR-HSI. Li *et al.* [23] adopt band attention to promote the consistency of generated spectra. In addition, Bayes-based single-image SR method proposed by Irmak *et al.* [33] transforms the problem of ill-posed SR reconstruction in the frequency domain into a quadratic optimization problem of abundance mapping and solves it by energy minimization method based on MRF. These methods require LR-HSI without need of corresponding MSI/RGB images.

In addition to the previous means, more HSI SR methods use other images with high spatial resolution and make full use of the spatial information [34], [35]. On this basis, some research base on the sparsity of HSI [21], [24], [35]–[38] to reconstruct the HR-HSI, some authors study the self-similarity between local and nonlocal patches [25], [34]–[36] and the low rank of them [38]–[40]; some literatures have use or model the imaging principles and degradation process for SR [24], [41], [42]. From the perspective of the solution process: dictionary-based methods are common, such as the works in [25], [38], and [39]. Since HSI is a 3-D tensor, Xu *et al.* [35] provide a *t-product* to establish the relationship between HR-HSI and LR-HSI and constraint nonlocal similarity, Li *et al.* [37] propose a fusion method based on coupled sparse tensor decomposition, Dian *et al.* [40] compose 4-D tensor with similar HSI patches, and

transform HSI SR problem into LTTR regularization constraint by using low-rank property. In [41] and [42], CNN is used to fit the reconstruction process, and in [41], GAN [43] is adopted for SR. Kwon and Tai [21] guide the up-sampling process of LR-HSI by RGB and spectrum substitution to refine the upsampled spectra. Borsoi *et al.* [44] adopt spectral unmixing and model the spectral variability to study the SR problem of seasonal variation. The fusion methods of HSI and MSI (or RGB) make use of spatial information to achieve more accurate SR. Compared with HSI only methods, they need to correspond other images in the scene, which is difficult to obtain in many cases.

The two kinds of HSI HR methods need HSI with low resolution, which still need hyperspectral sensors. Usually, the cost of hyperspectral sensors is expensive due to complicated optics and electronics [45]. Therefore, many times the acquisition of LR-HSI is hard to satisfy.

B. Spectral SR

Different from HSI SR needs HSI input, the SSR reduces requirements on the spectral resolution of input images. SSR restores HSI with the high spatial resolution images, such as MSI or RGB images.

To our knowledge, SSR methods for remote sensing mainly recover HSI from MSI and most methods work on one scene, which lack applicability. For example, Mei *et al.* [26] use 3-D CNN for simultaneous spatial and SSR and experiment on datasets for classification to select a patch for testing, whereas others for training. Besides, many CNN-based methods are adopted [27], [30]. In [30], a 1-D CNN is designed with tunable spectral subsampling layer, and in loss function, not only the Euclidean distances are constrained but also the first and second derivatives. The works in [15], [19], and [46] enhance spectral resolution by dictionary learning and spectral sparse. In [15], Fotiadou *et al.* adopt coupled dictionary learning along with sparse representations based on the assumption that the sparse codes in MSI and HSI are the same. Yi *et al.* [19] use both the spectral improvement and the spatial preservation strategies to enhance the spectral information while retaining the spatial one for HSI reconstruction. Meanwhile, Arun *et al.* [27] and Sun *et al.* [47] extract endmembers from MSI and reconstruct HSI from the abundance map. Arun *et al.* [27], combines the collaborative unmixing process and correct the pixel-level reconstruction results, where 3-D CNN is used to combine spatial information.

Compared to MSI, RGB imaging is simpler and requires less hardware. Many methods generate HSI from RGB images in natural scene. Some methods use image priors, such as Yan *et al.* [48] introduce category and location information into the network. In [49], Arad and Ben-Shahar adopt dictionary learning and in [50], manifold learning is used for the HSI generating. The most common RGB to HSI reconstruction methods is based on CNN [51], [52]. Pixel and channel attention are adopted in methods [53], [54] to expand the receptive field or focus on key channels. Li *et al.* provide 3-D RAN to extract contextual information between bands [53] and attention is adopted to weight different channels. Zhao *et al.* [54] propose a four-layer hierarchical regression network and increase the receptive field by establishing an attention mechanism for the residual global

module. In [28] and [55], the authors take the imaging mechanism and physically plausible into account to constrain the SSR process. Nie *et al.* [55] learn spectral response functions (SRFs) using modern film filter techniques. Shi *et al.* [56] upsample the bands and learn residuals to achieve accurate spectral recovery. Several methods [57], [58] use a GAN discriminator to ensure the reality of generated HSI. These methods are all based on natural HSI data, such as ICVL [49], [50], Bgu HS [51], [53], and Arad HS [54], [59]. They mainly generate 31 bands from RGB images, which different from the remote sensing HSI generating to 150 bands.

We can find that there lacks SSR technology for reconstructing multiband remote sensing HSI from RGB images.

C. GAN on Generation and Remote Sensing Image Processing

GAN was proposed in 2014 [43] and has been widely used in various fields, such as image generation [31], [60], image style transform [61], [62], and image SR [63], [64]. GAN is composed of a generator G and a discriminator D, the G and D zero-sum game jointly promote each other.

Image generation is to generate specific images from other images or text. The HSI reconstruction can be seen as a generation from RGB to HSI. GAN has been widely used on image generation. A generator aims to generate a target image from an image or other input and the discriminator aims to distinguish whether the image generated is real or not. Early methods used GAN [43] and its improvements to generate [65], [66]. After that, a number of hierarchical generation methods are used to generate an image closer and closer to the target one, such as Stack-GAN [67], LAPGAN [68], and PGGAN [69]. In [31], a generic image-to-image translation framework based on CGAN [70] is proposed and PatchGAN is proposed to improve the discriminator on output. Different from the pair image generation or style transform, unsupervised generation methods have also been gradually developed. CycleGAN [62] use cycle-consistent to realize unpair image translation. Meanwhile, DualGAN [71] and DiscoGAN [72] adopt a similar approach. In addition to one-to-one generation, many one-to-many generation methods have been developed, most of which are implemented by introducing latent code [73], [74]. A typical method is StyleGAN [61], [75], it aims to generate image with a specific style or obtain corresponding style image from latent code.

GAN has also achieved ideal results on remote sensing image processing [64], [76]. Ma *et al.* [77] is a GAN model for unsupervised pan-sharpening. Zhao *et al.* study remote sensing change detection via GAN [78]. In [64] and [76], the authors use GAN for remote sensing SR. GAN is also used for cloud detection [79] and scene classification [80]. Many studies use GAN for hyperspectral anomaly detection [10] and classification [81], whereas Mehta *et al.* [82] adopt GAN for HSI dehazing.

In this article, we will adopt GAN for the reconstruction of HSI from RGB images.

D. Quantitative Metrics on HSI Reconstruction

At present, the evaluation indicators used in the literature that recover HSI from RGB are different, and we will introduce the commonly used ones in the following. Let $I_R^{(i)}$ represents the ith

spectrum in the real HSI and $I_G^{(i)}$ represents the ith spectrum in the generated HSI. n is the total number of spectra. d represents the number of bands in HSI.

1) *Root-mean-squared error (RMSE)* measures the pixelwise square root error between spectra generated and the real. It is the most commonly used metric in RGB-to-HSI research [19], [28], [30], [48], [49], [51]–[58]

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (I_R^{(i)} - I_G^{(i)})^2}. \tag{1}$$

2) *Mean relative absolute error (MRAE)* is the magnitude of the error relative to the true value. Same as RMSE, they both measure the distance between the spectrum and its true value, except that MRAE is the error related to the L_1 norm [28], [48], [52]–[54], [56], [58]

$$\text{MRAE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|I_R^{(i)} - I_G^{(i)}|}{I_G^{(i)}}. \tag{2}$$

3) *Mean peak signal-to-noise ratio (MPSNR)*. PSNR is often used as a measurement of signal reconstruction quality in image compression and other fields, and it can be derived from mean squared error (MSE). For HSI, we need to calculate PSNR for different bands separately, and then take the average value, which is called MPSNR [19], [26], [27], [53]. Similar to MRAE, MPSNR reflects the relative magnitude between the error and real HSI

$$\text{MSE}_j = \frac{1}{n} \sum_{i=1}^{n} (I_R^{(i,j)} - I_G^{(i,j)})^2$$

$$\text{MPSNR} = \frac{1}{d} \sum_{j=1}^{d} 10 \times \lg \left(\frac{\text{MAX}_j}{\text{MSE}_j} \right). \tag{3}$$

In (3), MSE_j is the MSE of jth band and MAX_j is the max possible value in jth band of real HSI. $I_{R/G}^{(i,j)}$ is the jth intensity in ith spectrum.

4) *Spectral angle mapper (SAM)* reflects the spectral angular distance between the generated spectrum and the real one. Different from MPSNR or MRAE, SAM has more constraints on the shape of the generated spectrum [19], [26], [27], [30], [47], [48], [51]–[53]

$$\text{SAM} = \frac{1}{n} \sum_{i=1}^{n} \cos^{-1} \frac{I_R^{(i)T} I_G^{(i)}}{(I_R^{(i)T} I_R^{(i)})^{1/2} (I_R^{(i)T} I_R^{(i)})^{1/2}}. \tag{4}$$

5) *Mean structural SIMmilarity (MSSIM)*. SSIM is a measurement proposed in [83]. Unlike RMSE and MRAE, which measure error visibility, SSIM measures the structural similarity of two bands. MSSIM is the mean of SSIM of every generated HSI band and the real [19], [26], [27], [53]

$$\text{MSSIM} = \frac{1}{d} \sum_{j=1}^{d} \frac{(2\mu_R^j \mu_G^j + C_1)(2\sigma_{RG}^{jj} + C_2)}{(\mu_R^{j^2} + \mu_G^{j^2} + C_1)(\sigma_R^{j^2} + \sigma_G^{j^2} + C_2)} \tag{5}$$

where $\mu_{G/R}^j$ and $\sigma_{G/R}^j$ denote the mean and standard of patch in jth band of generated or real HSI, respectively.

σ_{RG}^{jj} is the covariance between the patches in jth band in R and jth band in G. $C_1 = (K_1 \times L)^2$ and $C_2 = (K_2 \times L)^2$ are constants where L denotes the dynamic range of pixel values and we select $K_1 = 0.01$, $K_2 = 0.03$, and $L = 65535$.

In addition, there are also other indicators adopted, such as NRMSD [50], ERGAS [47], and SSM [27], which evaluate pixelwise error; GFC [58], CC [19], and R [30] from the perspective of relevance; ΔE [58] and BPMRAE [54], that is colorimetric based.

In this article, we will compare R2HGAN with other state-of-the-art methods on RMSE, MRAE, SAM, MSSIM, and MPRNR to comprehensively evaluate the recovery result.

III. PROPOSED METHOD

In this article, we propose a 3 to 150 bands mapping framework for HSI reconstruction and experiment it on multiple scene images. To promote the diversity and the reality of the generated HSI, we design a joint discrimination method under the GAN framework. One discriminator is designed as a conditional PatchGAN, which supervises the consistency between HSI and input RGB. It outputs a patch in which each element represents the performance of a region in HSI. Another discriminator is proposed to ensure the physical reality of the generated spectra. Meanwhile, RGUS technique samples the spectra in HSI, and then they are input to the MLP to get reality scores for everyone. The flowchart of R2HGAN is illustrated in Fig. 1.

A. U-Net Generator

To make use of multiscale features of RGB image and generate an HSI containing semantic characteristics and detailed information in both, we design a U-Net for the generator (G). In G encoder, to realize rapid reduction of the image size, there are eight conv-layers with kernel_size $= 4 \times 4$, stride $= 2$ each followed by leaky ReLU. To save computing resources, we adopt convolutional kernel with Kernel_size $= 4 \times 4$, stride $= 2$ instead of kernel_size $= 44$, stride $= 1$ and 2×2 pooling. In the encoder, the size of feature maps is down by half layer by layer. Since the input image size is cropped to 256×256, to leverage information at the deepest semantic level, eight conv-layers are used one by one. After eight downsamplings, the size becomes $1/2^8 = 1/256$ of the input image. We utilize deconv kernel_size $= 4 \times 4$, stride $= 2$ for upsample and before it, ReLU is used for activation. In the first three layers in the decoder, dropout with keep_prob $= 0.5$ is adopted to avoid overfitting. After the concatenate with seven encoder layers, a deconv layer followed by Tanh activation restores the 150-channel HSI. It is worth noting that we removed all the BN layers in G as we select batchsize $= 1$ for the generator because it is well known that BN will cause poor effect when batch size is small [51].

B. Joint Discrimination: Conditional Discriminator

We design a joint discrimination for prompting the generator to produce more realistic spectra and ensure the correspondence between generated samples and input RGB. The joint discrimination is conducted by two discriminators, one is the conditional
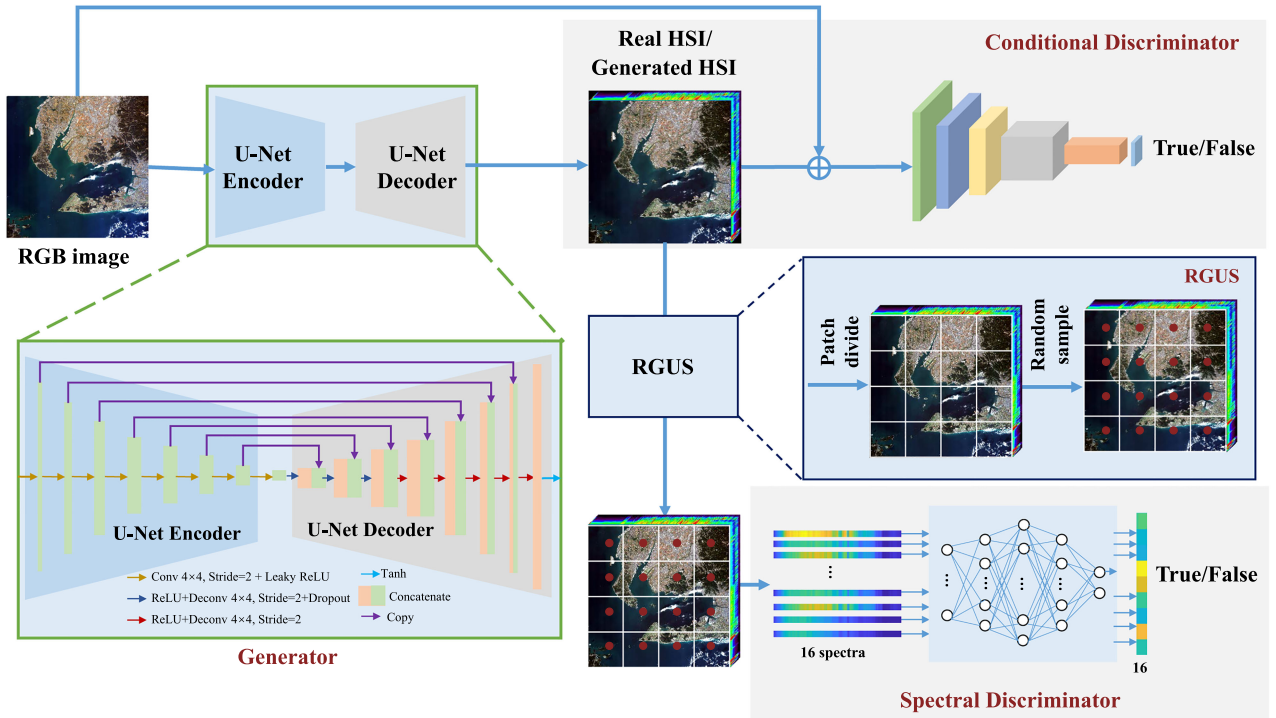
Fig. 1. Flowchart of R2HGAN. We design a U-Net generator to fully integrate features of different semantic levels, thus get a more precise HSI. The generator consists of an eight-layer convolutional encoder and a corresponding decoder. Joint discrimination is conducted by conditional discriminator and spectral discriminator. The conditional discriminator takes the RGB image and generated or real HSI for input, and output an $N \times N$ metric for correspondence of the two inputs. Before the spectral discriminate, RGUS is used for sample spectra from HSI, including steps patch divide and random sample. Then, the selected spectra are input to the MLP-based spectral discriminator.

discriminator (Con-D), which with a input concating HSI and RGB images. For the Con-D, we adopt PatchGAN to avoid one value between $0-1$ cannot express the whole generative result. With the input of RGB and HSI combined, the output of the Con-D is an $N \times N$ patch where each one correspond to the conditional reality of a local area in the generated image. Different from the traditional GAN, which the variable output evaluates the whole picture, our conditional D can consider the reality of each patch and integrate each one. See Table I for the detailed structure of conditional D.

C. Joint Discrimination: Spectral Discriminator

To improve the physical reality of the generated spectra, a spectral discriminator is provided. The main structure of the spectral discriminator (Spe-D) is an MLP with an input of one dim with 150 elements (a spectrum) and an output between 0–1 with details in Table II.

To comprehensively percept the spectral curve shape information, MLP is adopt instead of 1-D CNN. The output is to select the first of the softmax layer, which represents the probability of whether the spectrum is true.

We extract a certain number of spectra from the HSI for spectral discriminator to improve the calculation efficiency and prevent overfitting. To make the selected spectra more representative of the global spectral characteristics, we propose an RGUS technique. The RGUS can sample spectra evenly distributed in

TABLE I
STRUCTURE OF THE CONDITIONAL DISCRIMINATOR

Operation	Name	Size
–	Input	$153 \times 256 \times 256$
Conv 3×3		
Conv 3×3	Layer1	$256 \times 256 \times 256$
Leaky ReLU		
Conv 4×4, Stride=2		
Conv 3×3		
BatchNorm	Layer2	$512 \times 128 \times 128$
Leaky ReLU		
Conv 4×4, Stride=2		
Conv 3×3		
BatchNorm	Layer3	$1024 \times 64 \times 64$
Leaky ReLU		
Conv 4×4, Stride=2		
BatchNorm	Layer4	$1024 \times 32 \times 32$
Leaky ReLU		
Conv 4×4	Output	$1 \times 31 \times 31$
Sigmoid		

each position of the HSI, with greater randomness. Details of RGUS is shown in Algorithm 1.

TABLE II
STRUCTURE OF THE SPECTRAL DISCRIMINATOR

Operation	Name	Output Size
Linear(150,128)	Input	128
Linear(128,256)	Hidden-1	256
Linear(256,128)	Hidden-2	128
Linear(128,2)	Hidden-3	2
Softmax and select the first	Output	1

Algorithm 1 Random Global Uniform Sampling (RGUS)

Input: HSI data with size $N \times N$; Patch size for division $S \times S$;

Output: $M \times M$ spectra selected from the HSI;

1: Divide the input HSI into $M \times M$ patches with size $S \times S$, where $S \times M = N$;
2: Randomly generate two numbers a, b between 0 and S as the horizontal and vertical coordinates respectively;
3: Extract the spectra with the location $[a, b]$ in each patch and get $M \times M$ spectra;
4: **return** $M \times M$ spectra;

Force the output to 1 when training G to generate more real spectra and optimize them to 0 of the generated HSI (and the output of the real HSI to 1) to improve the discrimination performance of the spectral discriminator.

D. Loss Function

We use the L_1 loss and binary cross-entropy (BCE) loss to construct our loss functions. The goal of the generator is to generate pseudo-HSI to obey the real data distribution $p\_data(x)$ as much as possible, which can fool the discriminator. For this purpose, we first constrain the proximity of each generated spectra with the real ones. L_1 loss is a pixelwise loss and it can avoid the sensitivity to outliers and prone to oversmoothing of L_2 loss. Therefore, we choose L_1 loss in generator. Let y be the real HSI and x be the RGB image, z be the parameters in G, the output of G is $G(x, z)$, the L_1 loss can be calculated according to the following:

$$\mathcal{L}_{L1} = \mathbb{E}_{x,y,z}\left[\|y - G(x, z)\|_1\right]. \quad (6)$$

In addition, in order to fool the conditional discriminator, BCE is used for adversarial loss

$$\mathcal{L}_{\text{conadv}}(G, D_{\text{con}}) = \mathbb{E}_y\left[\log D_{\text{con}}(y)\right]$$
$$+ \mathbb{E}_{x,z}\left[\log(1 - D_{\text{con}}(G(x, z)))\right]. \quad (7)$$

In the same way, the adverse loss of the spectral discriminator is

$$\mathcal{L}_{\text{speadv}}(G, D_{\text{spe}}) = \mathbb{E}_y\left[\log D_{\text{spe}}(y)\right]$$
$$+ \mathbb{E}_{x,z}\left[\log(1 - D_{\text{spe}}(G(x, z)))\right]. \quad (8)$$

In summary, the loss of generator is in (9), where λ_1 and λ_2 represent the weight coefficient between the three losses

$$\mathcal{L}_G = \mathcal{L}_{L1} + \lambda_1 \mathcal{L}_{\text{conadv}}(G, D_{\text{con}}) + \lambda_2 \mathcal{L}_{\text{speadv}}(G, D_{\text{spe}}). \quad (9)$$

For discriminators, the optimization goal is to distinguish between generated HSI and the real, as (10) and (11)

$$\max_{D_{\text{con}}} \mathcal{L}_{\text{conadv}}(G, D_{\text{con}}) = \mathbb{E}_y\left[\log D_{\text{con}}(y)\right]$$
$$+ \mathbb{E}_{x,z}\left[\log(1 - D_{\text{con}}(G(x, z)))\right] \quad (10)$$

$$\max_{D_{\text{spe}}} \mathcal{L}_{\text{speadv}}(G, D_{\text{con}}) = \mathbb{E}_y\left[\log D_{\text{spe}}(y)\right]$$
$$+ \mathbb{E}_{x,z}\left[\log(1 - D_{\text{spe}}(G(x, z)))\right]. \quad (11)$$

In total, the optimization process can be expressed as follow:

$$\mathcal{G}^* = \arg\min_G \max_{D_{\text{con}}, D_{\text{spe}}} \mathcal{L}_{L1}(G) + \lambda_1 \mathcal{L}_{\text{conadv}}(G, D_{\text{con}})$$
$$+ \lambda_2 \mathcal{L}_{\text{speadv}}(G, D_{\text{spe}}). \quad (12)$$

E. Implementation

The training samples are a set of 256×256 image patches randomly cropped from RGB and corresponding HSI. During the spectral discrimination, we first select 4×4 spectra from the generated HSI and the spectra in the corresponding real HSI are also extracted. As for loss function, we set $\lambda_1 = \lambda_2 = 0.01$. For optimization, we adopt Adam Optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and during the process, the two discriminators are trained together and take turn with G. Learning rate decreasing strategy is adopted, the initial learning rate of G, conditional D, and spectral D are $1 \times e^{-4}$ and become 0.1 times the previous after every 300 epochs. In order to generate reliable results early, we first train 100 epochs of G to obtain preliminary results and then add the two discriminators into the training process to adjust the quality of generation. After two Ds join, every training G 3 iters, training the two Ds 1 iter.

IV. EXPERIMENT

In this section, we experiment with R2HGAN on GF5 dataset. We conduct a detailed introduction of the dataset and present different quantitative metrics on HSI construction. Then, a comparison with other state-of-the-art generation methods is carried out. Finally, ablation studies of the architecture are stated.

A. Datasets and Experiment Setup

The dataset includes six scenes of hyperspectral, including inland, island, and sea. It covers cities, mountains, arable land, forests, ports, etc., and the imaging time includes four seasons of 2018/2019. The HSIs are imaged by GF5 visible and near-infrared (VN) sensor with 16 b (the range of reflection value is 0–65535). Each HSI has a total of 150 bands covering the wavelength range of $390 - 1035\tilde{\ }$nm with a nominal spectral resolution of 5 nm. The RGB images are extracted from the HSIs' bands 72, 35, and 23. Each scene of the dataset has a spatial size of 2083×2008 and is cropped to 512×512 batches. Thus, the dataset has 120 512×512 HSIs and corresponding RGB images. 115 image pairs are selected for training and 5 image pairs for testing.
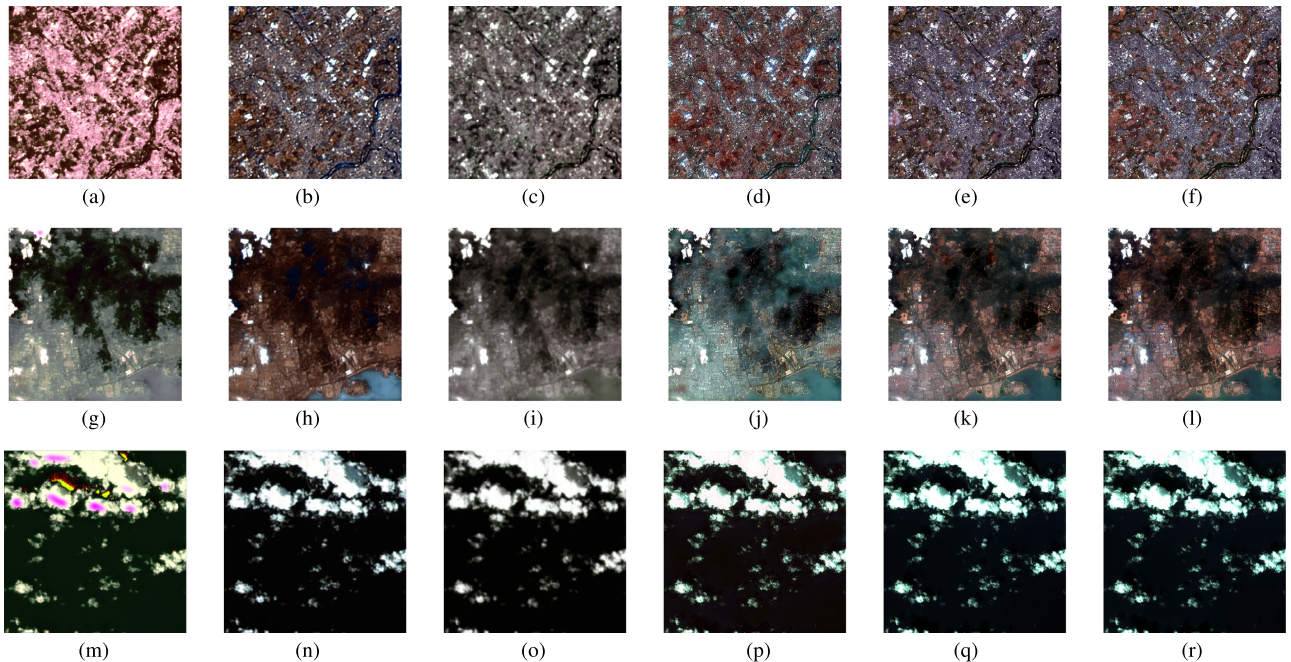
Fig. 2. False-color image of the generated HSI for three test images. (a)–(e) Result of MsCNN, HSCNN+, HSRNet, FMNet, and R2HGAN, respectively, for image 1. (f) False color of real HSI. (g)–(l) Result and real HSI for image 2. (m)–(r) Result and real HSI for image 3.

The experiment is carried out under win10 system with Intel (R) Core (TM) i7-7700 K CPU @ 4.20 GHz and NVIDIA GeForce GTX 1080. The training process of R2HGAN takes approximately 19 h. Under the same conditions, we compare R2HGAN to four SSR methods, including MsCNN [51], HSCNN+ [56], FMNet [84], and HSRNet [85]. MsCNN and HSCNN+ are based on U-Net and DenseNet, respectively. FM-Net adopts pixel-aware to keep pixel-learning ability. HSRNet uses physical mechanism, reconstructing according to SRF.

To train all the deep-learning-based methods adequately, we train them on the whole training set (115 image pairs) and chose different parameters by evaluating them on the testing set (five image pairs). Therefore, we choose the parameters by selecting the best result of each method. All experiment settings are set same for R2HGAN and other four methods.

B. Comparison With Other Methods

We compare our R2HGAN with some state-of-the-art SSR methods, including MsCNN [51], HSCNN+ [56], FMNet [84], and HSRNet [85]. All these methods are fully optimized on our dataset under same condition to obtain the best performance for fair competition. The exception is that HSRNet is cropped during the test due to GPU memory limitations. Fig. 2 shows a comparison of the false-color images of the HSI generated by these methods. We can find that R2HGAN generates HSI whose false-color image is closer to the real HSI. MsCNN has a large loss of both spatial and color information. HSCNN+ and HSRNet can maintain the spatial structure well, but lost information of the color. FMNet can better preserve color information while maintaining spatial structure relatively completely. R2HGAN has a good recovery of spatial and color information at the same time. The HSI generated by HSRNet has seams because

TABLE III
COMPARISON OF DIFFERENT METHODS ON OUR DATASET

Method	RMSE ↓	MRAE ↓	SAM ↓	MSSIM ↑	MPSNR ↑
MsCNN [51]	23682	136.33	1.3539	0.7108	28.056
HSCNN+ [56]	2242	6.829	0.192	0.9432	46.273
HSRnet [84]	2625.2	5.997	0.4832	0.9426	48.166
FMNet [85]	7793.4	40.787	0.4842	0.8741	42.792
R2HGAN	**178.3**	**0.126**	**0.0435**	**0.9972**	**61.479**

Note: For RMSE, MRAE, SAM, and RSDS, a lower score indicates better, whereas for MSSIM and MPSNR, a higher score indicates better.

of the limitation of GPU memory, we cut the RGB image input 512×512 into 256×256 and resplicing.

Table III shows the performance of different comparison methods on various indicators, including MPSNR, MRAE, and RMSE to measure the result of image reconstruction, spectral angle mapping (SAM) to quantity related to spectral shape and category information, and MSSIM that measure the spatial structure. The bold entities indicate the best performance on corresponding indicators. MPSNR represents RMSE and MRAE to a certain extent and $MAX_j = 65535$. From these indicators, we can find that FMNet is underperforming in terms of spatial structure, which is related to its use of pixel-aware. HSCNN+ and HSRNet perform relatively well on both spectral and spatial information except R2HGAN. R2HGAN performs better than any other comparing methods in five indicators. R2HGAN gets the highest MPSNR of 61.479 followed by HSRNet with 48.166. Meanwhile, R2HGAN preserves the most structural information with MSSIM = 0.9972 followed by HSCNN+. The SAM between R2HGAN generated HSIs and the real ones reach as low as 0.0435, whereas HSCNN+ gets a second at 0.192 more than four times of R2HGAN.

Besides, our method can avoid generating abnormal or nonphysical spectral curves. As shown in Fig. 3, the relatively
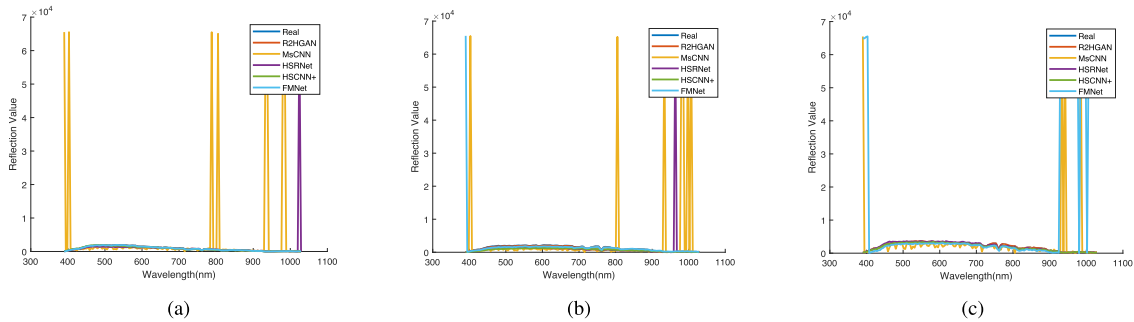
Fig. 3. Abnormal spectra generated by comparison methods. The title of the figure represents the methods generating abnormal spectra. In (a), MsCNN and HSRNet have abnormal generation. In (b), except for R2HGAN and HSCNN+, other methods all generated abnormal spectra. In (c), MsCNN and FMNet have abnormal generation. (a) MsCNN and HSRNet. (b) MsCNN, HSRNet, and FMNet. (c) MsCNN and FMNet.
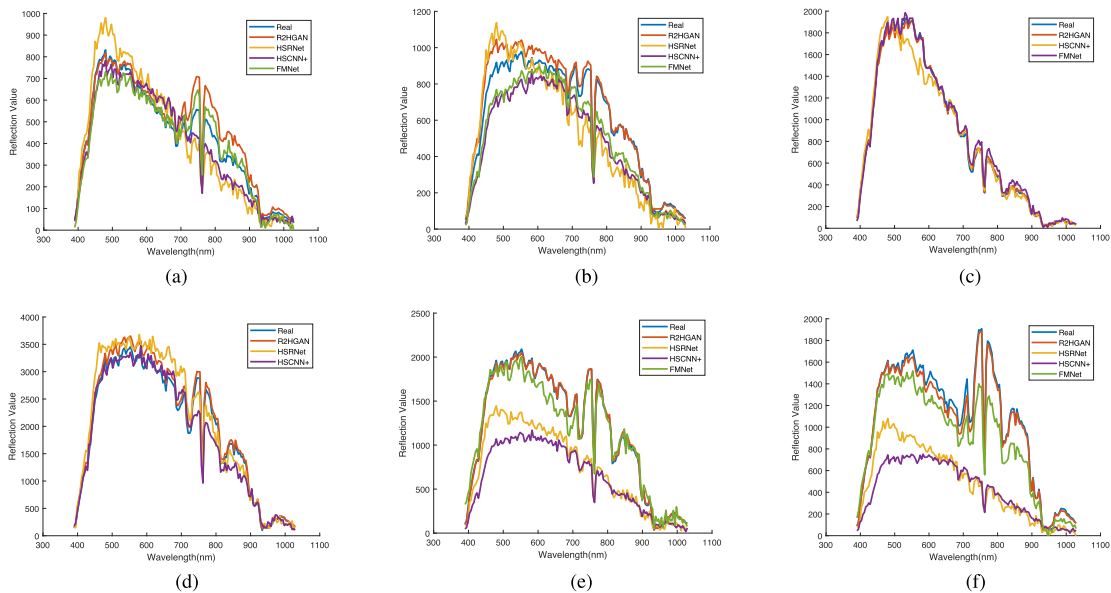

Fig. 4. Compare real spectra with those generated by various methods. (a)–(d) Spectra of different ground objects produced by five methods. The titles represent objects selected to plot spectra. R2HGAN generates spectra that are more like real ones on multiobjects. (a) Building. (b) Soil. (c) Water. (d) Cloud. (e) Road. (f) Tree.

competitive HSRNet and FMNet in the comparison method are both prone to outliers. Furthermore, MsCNN is almost impossible to produce spectra without anomalies. R2HGAN has not been found to produce abnormal spectra.

We randomly select different objects and plot their spectra generated as well as the real ones. Fig. 4 shows the difference between spectra generated by five methods and the real ones. Note that the lack of spectral curves from a certain method in some figures is due to the abnormal spectrum generated by the method. We only show the generated spectra, which are similar to the real ones. It is obvious that R2HGAN generated spectra more like the real ones for various objects. Besides, FMNet performed relatively well, HSRNet and HSCNN+ performed poorly, and MsCNN was the worst, barely able to generate a normal spectrum.

In Fig. 5, some typical bands of the HSI generated by the five methods are exhibited. From left to right are, respectively, bands of MsCNN, HSCNN+, HSRNet, FMNet, R2HGAN, and the GroundTruth. The performance of MsCNN in all bands is far from the true value. Besides, all methods except MsCNN get a

relatively close generation to the real HSI for band 60 in image 1. For band 10 and 80 in image 1, the five methods distorted slightly in spatial details of the band and FMNet also caused brightness distortion. But for band 135 in image 1 and band 100 in image 2, all methods failed to generate reliable bands, only R2HGAN can reserve more spatial information. Test image 3 is relatively simple with cloud and sea water, but FMNet and MsCNN all cause abnormal points. Test image 3 is relatively simple with cloud and sea water, but FMNet and MsCNN cause abnormal points.

C. Ablation Studies

The ablation studies are conducted to improve the effectiveness of the two discriminators. We compare the joint discrimination with conditional discriminator only or spectral discriminator only, respectively.

1) Study of Joint Discrimination: Here, we mainly analyze the excellence of joint discrimination. In Table IV, we demonstrate RMSE, MRAE, SAM, MSSIM, and MPSNR of
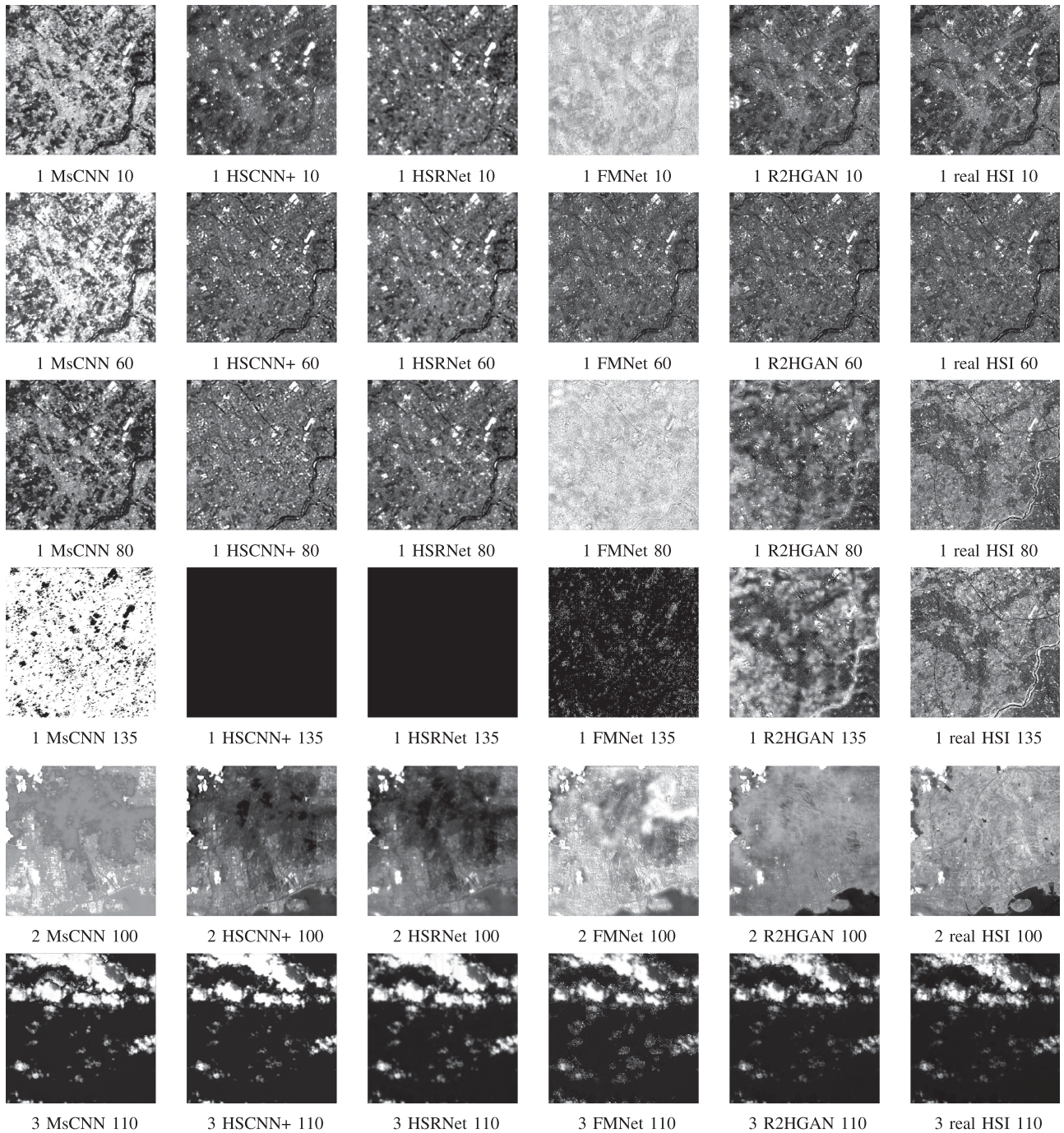
Fig. 5. Different bands of the generated HSI on three test images. Each row represents the effect of different methods on the same band, and each column represents the different bands of the same method to generate HSI. Numbers before the methods represent different test images and the following ones represent the index of the band. For example, 1 MsCNN 30 means it shows the 30th band in the HSI generated for image 1 by MsCNN.

six ablation experiments and the bold entities indicate the best performance on corresponding indicators.

Comparing experiments 1 and 3 in Table IV, we can find that the introduce of the conditional discriminator greatly improved the generation effect that the MPSNR gains an increase of around 8.2. Meanwhile, the MSSIM also has a 1.4% improvement, which means that the conditional discriminator enables the generator to recover more spatial structure information.

In experiment 4, only the spectral discriminator is adopted together with the RGUS. It can be seen that the addition of the

spectral discriminator has greatly improved the quality of the generated spectra, and the PSNR has been increased from 51.29 to 60.08. At the same time, the spectral discriminator can obtain a lower SAM than the conditional discriminator.

The experiment 6 in Table IV shows the results of adopting joint discrimination and using RGUS sampling for spectral discriminator. And it is shown that with the discriminators, all indicators have improved. When we use conditional discriminator and spectral discriminator simultaneously, the spectra generated are more similar to the real ones that use G alone or G and

TABLE IV
ABLATION STUDIES, THE CROSS IN BN MEANS MOVING BN FROM G, THE CHECK MARK MEANS WITH THE D OR RGUS

Name	BN	Con-D	Spe-D	RGUS	RMSE ↓	MRAE ↓	SAM ↓	MSSIM ↑	MPSNR ↑
1	✗				1820.5	4.4045	0.1404	0.981	51.2877
2		✓	✓	✓	311.02	0.382	0.0742	0.991	58.1387
3	✗	✓			357.62	0.37	0.0553	0.9952	59.4947
4	✗		✓	✓	266.56	0.227	0.0477	0.9962	60.0801
5	✗	✓	✓		204.91	0.162	0.0462	0.9965	60.7535
6	✗	✓	✓	✓	**143.23**	**0.1066**	**0.0418**	**0.9975**	**62.058**

Note: Con-D represents the conditional discriminator, whereas Spe-D means the spectral one.
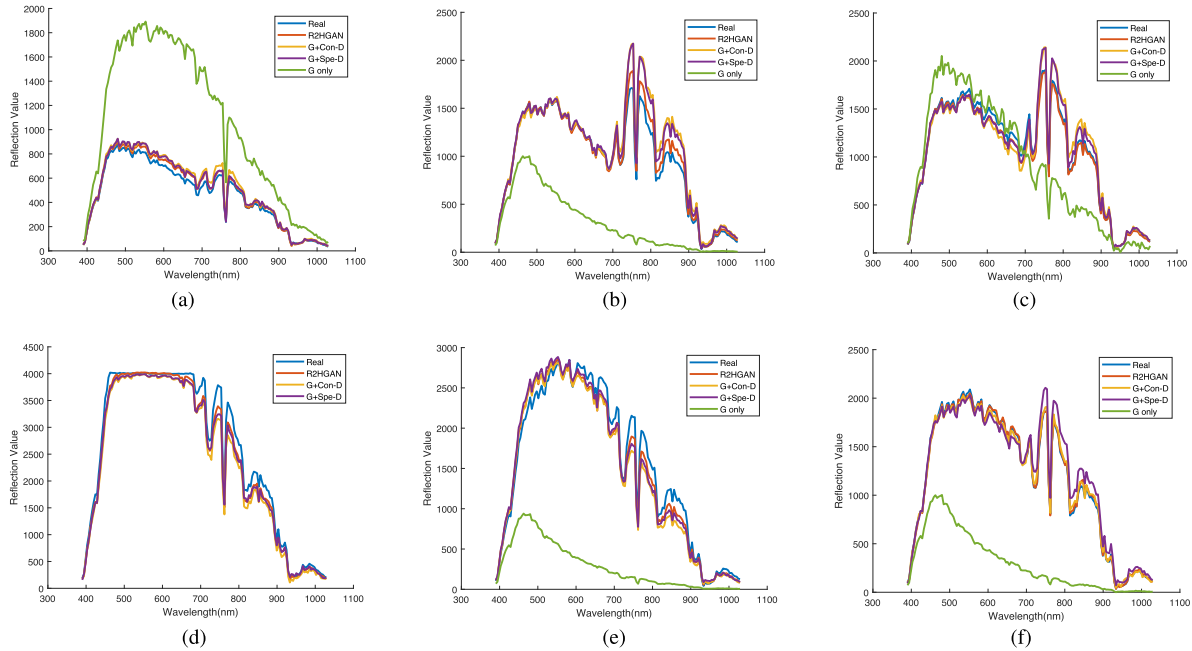

Fig. 6. Comparison of generated spectra after adding discriminators on different objects. R2HGAN represents the joint discrimination, G+Con-D means only use conditional D and G only is the spectral without any discriminator. (a) Building. (b) Soil. (c) Tree. (d) Cloud. (e) Port. (f) Road.

conditional D. As is shown in Fig. 6, G alone method causes a huge deviation in the generated spectra. R2HGAN represents the joint with the conditional and spectral discriminator, it generates spectra very close to the method with conditional D only. The spectral discriminator fine-tunes the spectra to make it closer to real. For example, Fig. 6(d) shows the spectra of a cloud pixel. G only method generates spectrum far from the real. The conditional D and R2HGAN obtain spectra exactly alike, and the one from R2HGAN is more similar to the real.

2) Ablation on RGUS: For the spectral discriminator, we randomly select several spectra from the HSI. We design an RGUS technique to extract spectra, which represent global spectral information of the HSI. For experiment 5 in Table IV, we randomly select the same number of spectra as RGUS for spectral discrimination. As compared between 5 and 6, the RGUS promotes the MPSNR from 60.75 to 61.30. Therefore, the global information by the RGUS is helpful for the generation.

Meanwhile, we compare the patch size of RGUS in Table V, the bold entities indicate the best performance on corresponding indicators. A larger patch size results in fewer selected spectra, which reduces training time. Smaller patch size selects more spectra and causes consumption of the training time. As is shown

TABLE V
EXPERIMENT ON THE PATCH SIZE OF RGUS

Name	M in RGUS	RMSE ↓	MRAE ↓	SAM ↓	MSSIM ↑	MPSNR ↑
1	16	178.97	0.1189	0.0425	0.9973	60.939
2	32	178.28	0.1257	0.0435	0.9972	61.479
3	64	**143.23**	**0.1066**	0.0418	0.9975	**62.058**
4	128	162.32	0.1178	**0.0417**	**0.9976**	61.842

in Table V, with the increase of patch size in RGUS (M), the spectra input to Spe-D reduce, the various indicators gradually improve till $M = 64$. Thus, we select the patch size as 64.

3) Effect of Removing BN in G: From Table IV, experiments 2 and 6 show the different effects of whether to remove BN. Comparing with BN in G, the removal of it won a 3.16 improvement of MPSNR as well as promoting of other indicators.

V. CONCLUSION

We propose a general framework for generating HSI from RGB images. The R2HGAN can generate reliable spectra while retaining more spatial information. We collect six scenes of HSI from GF5 for the experiment, which is different from

the previous methods of conducting testing and training on one-scene HSI. And to overcome the ill-posedness during 3–150 bands generation and overfitting (over smooth), we design a joint discrimination method to distinguish the reality of the generation by two discriminators. The conditional discriminator based on PatchGAN measures the correspondence between the generated HSI and the input RGB. The spectral discriminator adopts MLP architecture to avoid the problem of CNN being affected by the receptive field. Furthermore, we newly design an RGUS method to extract spectra from the generated HSI for the spectral discriminator. The RGUS is closer to obtaining a subset of the global spectral expression. Ablation studies improve the effectiveness of our framework. Our method achieves better generation results than other state-of-the-art methods under the evaluation of multiple indicators.[2]

REFERENCES

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

[2] M. Borhani and H. Ghassemian, "Kernel multivariate spectral-spatial analysis of hyperspectral data," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 6, pp. 2418–2426, Jun. 2015.

[3] N.-B. Chang, B. Vannah, and Y. Jeffrey Yang, "Comparative sensor fusion between hyperspectral and multispectral satellite sensors for monitoring microcystin distribution in Lake Erie," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 6, pp. 2426–2442, Jun. 2014.

[4] B. Pan, Z. Shi, and X. Xu, "MugNet: Deep learning for hyperspectral image classification using limited samples," ISPRS J. Photogrammetry Remote Sens., vol. 145, pp. 108–119, 2018.

[5] L. Liu, Z. Shi, B. Pan, N. Zhang, H. Luo, and X. Lan, "Multiscale deep spatial feature extraction using virtual RGB image for hyperspectral imagery classification," Remote Sens., vol. 12, no. 2, 2020, Art. no. 280.

[6] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[7] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 5, pp. 4340–4354, May 2021.

[8] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 6, pp. 5131–5153, Jun. 2021.

[9] Z. Zou and Z. Shi, "Hierarchical suppression method for hyperspectral target detection," IEEE Trans. Geosci. Remote Sens., vol. 54, no. 1, pp. 330–342, Jan. 2016.

[10] T. Jiang, Y. Li, W. Xie, and Q. Du, "Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection," IEEE Trans. Geosci. Remote Sens., vol. 58, no. 7, pp. 4666–4679, Jul. 2020.

[11] Z. Shi, T. Shi, M. Zhou, and X. Xu, "Collaborative sparse hyperspectral unmixing using L_0 norm," IEEE Trans. Geosci. Remote Sens., vol. 56, no. 9, pp. 5495–5508, Sep. 2018.

[12] X. Xu, Z. Shi, and B. Pan, "L_0-based sparse hyperspectral unmixing using spectral information and a multi-objectives formulation," ISPRS J. Photogrammetry Remote Sens., vol. 141, pp. 46–58, 2018.

[13] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," IEEE Trans. Image Process., vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[14] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction," ISPRS J. Photogrammetry Remote Sens., vol. 158, pp. 35–49, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271619302199

[15] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Spectral super resolution of hyperspectral images via coupled dictionary learning," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 5, pp. 2777–2797, May 2019.

[16] Z. Li, H. Shen, H. Li, G. Xia, P. Gamba, and L. Zhang, "Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery," Remote Sens. Environ., vol. 191, pp. 342–358, 2017.

[17] D. P. Roy et al., "Landsat-8: Science and product vision for terrestrial global change research," Remote Sens. Environ., vol. 145, pp. 154–172, 2014.

[18] Y.-N. Liu et al., "The advanced hyperspectral imager: Aboard China's GaoFen-5 satellite," IEEE Geosci. Remote Sens. Mag., vol. 7, no. 4, pp. 23–32, Dec. 2019.

[19] C. Yi, Y.-Q. Zhao, and J. C.-W. Chan, "Spectral super-resolution for multispectral image based on spectral improvement strategy and spatial preservation strategy," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 11, pp. 9010–9024, Nov. 2019.

[20] T. Akgun, Y. Altunbasak, and R. Mersereau, "Super-resolution reconstruction of hyperspectral images," IEEE Trans. Image Process., vol. 14, no. 11, pp. 1860–1875, Nov. 2005.

[21] H. Kwon and Y.-W. Tai, "RGB-guided hyperspectral image upsampling," in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 307–315.

[22] J. Hu, X. Jia, Y. Li, G. He, and M. Zhao, "Hyperspectral image super-resolution via intrafusion network," IEEE Trans. Geosci. Remote Sens., vol. 58, no. 10, pp. 7459–7471, Oct. 2020.

[23] J. Li et al., "Hyperspectral image super-resolution by band attention through adversarial learning," IEEE Trans. Geosci. Remote Sens., vol. 58, no. 6, pp. 4304–4318, Jun. 2020.

[24] K. Zheng et al., "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 3, pp. 2487–2502, Mar. 2021.

[25] R. Dian, S. Li, L. Fang, T. Lu, and J. M. Bioucas-Dias, "Nonlocal sparse tensor factorization for semiblind hyperspectral and multispectral image fusion," IEEE Trans. Cybern., vol. 50, no. 10, pp. 4469–4480, Oct. 2020.

[26] S. Mei, R. Jiang, X. Li, and Q. Du, "Spatial and spectral joint super-resolution using convolutional neural network," IEEE Trans. Geosci. Remote Sens., vol. 58, no. 7, pp. 4590–4603, Jul. 2020.

[27] P. Arun, K. M. Buddhiraju, A. Porwal, and J. Chanussot, "CNN based spectral super-resolution of remote sensing images," Signal Process., vol. 169, 2020, Art. no. 107394

[28] B. J. Fubara, M. Sedky, and D. Dyke, "RGB to spectral reconstruction via learned basis functions and weights," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2020, pp. 1984–1993.

[29] S. Koundinya et al., "2D-3D CNN based architectures for spectral reconstruction from RGB images," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2018, pp. 957–9577.

[30] U. B. Gewali, S. T. Monteiro, and E. Saber, "Spectral super-resolution with optimized bands," Remote Sens., vol. 11, no. 14, 2019, Art. no. 1648.

[31] 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.

[32] J. Hu, Y. Li, and W. Xie, "Hyperspectral image super-resolution by spectral difference learning and spatial error correction," IEEE Geosci. Remote Sens. Lett., vol. 14, no. 10, pp. 1825–1829, Oct. 2017.

[33] H. Irmak, G. B. Akar, and S. E. Yuksel, "A map-based approach for hyperspectral imagery super-resolution," IEEE Trans. Image Process., vol. 27, no. 6, pp. 2942–2951, Jun. 2018.

[34] X.-H. Han, B. Shi, and Y. Zheng, "Self-similarity constrained sparse representation for hyperspectral image super-resolution," IEEE Trans. Image Process., vol. 27, no. 11, pp. 5625–5637, Nov. 2018.

[35] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," IEEE Trans. Image Process., vol. 28, no. 6, pp. 3034–3047, Jun. 2019.

[36] Y. Zhao, J. Yang, and J. C.-W. Chan, "Hyperspectral imagery super-resolution by spatial-spectral joint nonlocal similarity," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 6, pp. 2671–2679, Jun. 2014.

[37] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," IEEE Trans. Image Process., vol. 27, no. 8, pp. 4118–4130, Aug. 2018.

[38] 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.

[39] M. A. Veganzones, M. Simões, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," IEEE Trans. Image Process., vol. 25, no. 1, pp. 274–288, Jan. 2016.

[2]Codes of our R2HGAN are available on http://levir.buaa.edu.cn/Code.htm

[40] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.

[41] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang, "Hyperspectral image super-resolution with optimized RGB guidance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11 653–11662.

[42] 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. 2021.

[43] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, and S. Ozair *et al.*, "Generative adversarial networks," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2014, vol. 3, pp. 2672–2680.

[44] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability," *IEEE Trans. Image Process.*, vol. 29, pp. 116–127, 2020.

[45] S. Song *et al.*, "Low-cost hyper-spectral imaging system using a linear variable bandpass filter for agritech applications," *Appl. Opt.*, vol. 59, no. 5, pp. A 167–A175, 2020.

[46] K. V. Mishra, M. Cho, A. Kruger, and W. Xu, "Spectral super-resolution with prior knowledge," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5342–5357, Oct. 2015.

[47] X. Sun, L. Zhang, H. Yang, T. Wu, Y. Cen, and Y. Guo, "Enhancement of spectral resolution for remotely sensed multispectral image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2198–2211, May 2015.

[48] L. Yan, X. Wang, M. Zhao, M. Kaloorazi, J. Chen, and S. Rahardja, "Reconstruction of hyperspectral data from RGB images with prior category information," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1070–1081, 2020.

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

[50] Y. Jia *et al.*, "From RGB to spectrum for natural scenes via manifold-based mapping," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4715–4723.

[51] Y. Yan, L. Zhang, J. Li, W. Wei, and Y. Zhang, "Accurate spectral super-resolution from single RGB image using multi-scale CNN," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis.*, 2018, pp. 206–217.

[52] S. Galliani, C. Lanaras, D. Marmanis, E. Baltsavias, and K. Schindler, "Learned spectral super-resolution," 2017, *arXiv:1703.09470*. [Online]. Available: http://arxiv.org/abs/1703.09470

[53] J. Li *et al.*, "Hybrid 2-D-3-D deep residual attentional network with structure tensor constraints for spectral super-resolution of RGB images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2321–2335, Mar. 2021.

[54] Y. Zhao, L.-M. Po, Q. Yan, W. Liu, and T. Lin, "Hierarchical regression network for spectral reconstruction from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 1695–1704.

[55] S. Nie, L. Gu, Y. Zheng, A. Lam, N. Ono, and I. Sato, "Deeply learned filter response functions for hyperspectral reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4767–4776.

[56] Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu, "HSCNN+: Advanced CNN-based hyperspectral recovery from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 1052–10528.

[57] K. G. Lore, K. K. Reddy, M. Giering, and E. A. Bernal, "Generative adversarial networks for spectral super-resolution and bidirectional RGB-to-multispectral mapping," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 926–933.

[58] A. Alvarez-Gila, J. Van De Weijer, and E. Garrote, "Adversarial networks for spatial context-aware spectral image reconstruction from RGB," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2017, pp. 480–490.

[59] H. Peng, X. Chen, and J. Zhao, "Residual pixel attention network for spectral reconstruction from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 2012–2020.

[60] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2642–2651.

[61] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4401–4410.

[62] 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. 2242–2251.

[63] A. Bulat, J. Yang, and G. Tzimiropoulos, "To learn image super-resolution, use a GAN to learn how to do image degradation first," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 185–200.

[64] S. Lei and Z. Shi, "Hybrid-scale self-similarity exploitation for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, 2021, to be published, doi: 10.1109/TGRS.2021.3069889.

[65] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, vol. 29, pp. 2234–2242.

[66] S. Nowozin, B. Cseke, and R. Tomioka, "f-GAN: Training generative neural samplers using variational divergence minimization," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, vol. 29, pp. 271–279.

[67] H. Zhang *et al.*, "StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5907–5915.

[68] E. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a Laplacian pyramid of adversarial networks," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 1486–1494.

[69] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Representations*, 2018.

[70] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: http://arxiv.org/abs/1411.1784

[71] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2868–2876.

[72] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 1857–1865.

[73] J. Y. Zhu *et al.*, "Toward multimodal image-to-image translation," in *Proc. 31st Annu. Conf. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 465–476.

[74] X. Huang, M.-Y. Liu, S. J. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 179–196.

[75] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8110–8119.

[76] S. Lei, Z. Shi, and Z. Zou, "Coupled adversarial training for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3633–3643, May 2020.

[77] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Inf. Fusion*, vol. 62, pp. 110–120, 2020.

[78] W. Zhao, X. Chen, X. Ge, and J. Chen, "Using adversarial network for multiple change detection in bitemporal remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, 2020, to be published, doi: 10.1109/LGRS.2020.3035780.

[79] Z. Wu, J. Li, Y. Wang, Z. Hu, and M. Molinier, "Self-attentive generative adversarial network for cloud detection in high resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1792–1796, Oct. 2020.

[80] C. Shi, L. Fang, Z. Lv, and H. Shen, "Improved generative adversarial networks for VHR remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, 2020, to be published, doi: 10.1109/LGRS.2020.3025099.

[81] J. Feng, H. Yu, L. Wang, X. Cao, X. Zhang, and L. Jiao, "Classification of hyperspectral images based on multiclass spatial-spectral generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5329–5343, Aug. 2019.

[82] A. Mehta, H. Sinha, P. Narang, and M. Mandal, "HIDeGan: A hyperspectral-guided image dehazing GAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 846–856.

[83] 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.

[84] L. Zhang *et al.*, "Pixel-aware deep function-mixture network for spectral super-resolution," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 12 821–12828.

[85] J. He, J. Li, Q. Yuan, H. Shen, and L. Zhang, "Spectral response function-guided deep optimization-driven network for spectral super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, Feb. 2021 to be published, doi: 10.1109/TNNLS.2021.3056181.

Liqin Liu received the B.S. degree in probe, guidance, and control (astronautics) in 2018 from Beihang University, Beijing, China, where she is currently working toward the Doctorate degree in pattern recognition and intelligent systems with the Image Processing Center, School of Astronautics.

Her research interests include hyperspectral image processing, machine learning, and deep learning.

Sen Lei received the B.S. degree in probe, guidance, and control (astronautics) in 2015 from Image Processing Center, School of Astronautics, Beihang University, Beijing, China, where he is currently working toward the Doctorate degree in pattern recognition and intelligent systems.

His research interests include deep learning and image super-resolution.

Ning Zhang received the Ph.D. degree in optical engineering from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China, in 2012.

He is currently a Senior Engineer with Shanghai Aerospace Electronic Technology Institute, Shanghai, China. His current research interests include remote sensing image processing and electronic system design.

Xinzhong Zhu received the M.S. degree in mechatronic engineering from Wuhan University of Technology, Wuhan, China, in 2000. He is currently a Research Fellow with the Shanghai Aerospace Electronic Technology Institute, Shanghai, China. His current research interests include high-speed data processing and electronic system design.

Zhenwei Shi (Member, IEEE) received the Ph.D. degree in mathematics from the Dalian University of Technology, Dalian, China, in 2005.

He was a Postdoctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA, from 2013 to 2014. He is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing. He has authored or coauthored more than 100 scientific papers in refereed journals and proceedings, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *IEEE Geoscience and Remote Sensing Letters*, and IEEE Conference on Computer Vision and Pattern Recognition. His current research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Dr. Shi serves as an Associate Editor for the *Infrared Physics and Technology*.