Physics-informed Hyperspectral Remote Sensing Image Synthesis with Deep Conditional Generative Adversarial Networks

Liqin Liu, Wenyuan Li, Zhenwei Shi, *Member, IEEE*, and Zhengxia Zou*

***Abstract*—High-resolution hyperspectral remote sensing images are of great significance to agricultural, urban, and military applications. However, collecting and labeling hyperspectral images is time-consuming, expensive and usually heavily relies on domain knowledge. In this paper, we propose a new method for generating high-resolution hyperspectral images as well as sub-pixel groundtruth annotations from RGB images. Given a single high-resolution RGB image as its conditional input, unlike previous methods directly predict spectral reflectance that ignores the physics behind, we consider both imaging mechanism and spectral mixing, and introduce a deep generative network that first recovers the spectral abundance for each pixel, and then generate the final spectral data cube with the standard USGS spectral library. In this way, our method not only synthesizes high-quality spectral data existing in real-world but also generates sub-pixel-level spectral abundance with well-defined spectral reflectance characteristics. We also introduce a spatial discriminative network and a spectral discriminative network to improve the fidelity of the synthetic output from both spatial and spectral perspectives. The whole framework can be trained end-to-end in an adversarial training paradigm. We refer to our method as "Physics-informed Deep Adversarial Spectral Synthesis (PDASS)". On the IEEE *grss_dfc_2018* dataset, our method achieves an MPSNR of 47.56 on spectral reconstruction accuracy and outperforms other state-of-the-art methods. As latent variables, the generated spectral abundance and the atmospheric absorption coefficients of sunlight also suggest the effectiveness of our method.**

***Index Terms*—Hyperspectral image, remote sensing, Generation Adversarial Networks (GAN), spectral super-resolution (SSR), imaging model**

I. Introduction

HYPERSPECTRAL remote sensing imagery (HSI) endows its unique advantages in object recognition [1], and is widely used in many fields such as urban planning [2–5], precision agriculture [6], and environmental monitoring [7]. However, high-spatial resolution hyperspectral remote sensing images are very hard to obtain. Usually, for airborne spectral sensors, the spatial resolution of hyperspectral imagery is usually lower than 1m/pixel [8, 9]. For spaceborne sensors, the resolution is even as low as 30m/pixel [10]. The ground truth annotation is time-consuming, expensive, and may heavily require extensive fieldwork. The mixture of subpixel spectral data also brings additional difficulties.

In recent years, efforts have been made in developing post-processing methods to overcome hardware limitations and acquire remote sensing images with both high spatial and spectral resolution. There are mainly three groups of approaches: hyperspectral image spatial super-resolution [11], spectral super-resolution [12] and image fusion [13]. Although efforts have been made in the above directions, there is still a huge gap between current methods and practical applications of hyperspectral images. The key reason is that the hyperspectral data synthesis process is highly ill-posed and the data from the above methods lacks necessary physical meaning. For example, most recent super-resolution based methods [10, 14–18] frame the spatial/spectral data generation as a pair-wise regression process. At the inference stage, these methods directly predict band-wise spectral reflectance of each input pixel location while ignoring the physics behind. This will bring a problem that the predicted spectral data may not truly exist in the real world. Although some recent approaches [19, 20] introduce adversarial training to improve the visual fidelity of the generated data, the imaging mechanism, and sub-pixel spectral mixing are simply ignored.

In this paper, we propose a deep conditional generative model for high-resolution hyperspectral image synthesis and refer to it as "Physics-informed Deep Adversarial Spectral Synthesis (PDASS)". We start from the remote sensing imaging model that fully considers imaging mechanisms including spectral mixing, the influence of sunlight intensity, atmospheric absorption, and quantification. Given a single input high-resolution RGB image, the method first recovers the abundance for each type of ground feature at each pixel location, then generates the final spectral data cube by utilizing the standard USGS spectral library [21] and the imaging model. As a result, the method not only produces high-quality spectral data but also generates sub-pixel-level spectral abundance with well-defined spectral reflectance characteristics. To further improve the visual fidelity of the synthetic hyperspectral data cube, we design our learning framework based on the recent success of deep generative adversarial networks [22–24]. We introduce a spatial discriminator and a spectral discriminator to model the true distribution of real-world hyperspectral data from both spatial and spectral dimensions, and help the generative network produce better results.

The work was supported by the National Natural Science Foundation of China under the Grant 62125102. *(Corresponding author: Zhengxia Zou (e-mail: zhengxiazou@buaa.edu.cn))*

Liqin Liu, Wenyuan Li and Zhenwei Shi are with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, and with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China, and also with State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing 100191, China. Zhengxia Zou is with Department of Guidance, Navigation and Control, School of Astronautics, Beihang University, Beijing 100191, China.

The proposed method has several advantages. First, the causal factors behind the generated data can be nicely recovered. This means that we can not only generate synthetic hyperspectral data itself but also clearly know which types of ground objects (spectrum identity in the library) are contained at each pixel and how much proportion (spectral abundance) they have. Second, the training of our method is conducted in a self-supervised fashion, neither relies on the pairwise RGB-hyperspectral images, nor the groundtruth of abundance. Third, the effect of solar radiation after atmosphere absorption at different wavelengths and quantification can be easily eliminated since it can be optimized all together with the spectral mixing model under a unified framework.

Extensive experiments are conducted to verify the effectiveness of the method. Our method generates physically and visually meaningful results in terms of both band-wise spatial images and pixel-wise spectral curves. On the IEEE *grss_dfc_2018* dataset, our method outperforms other state-of-the-art methods on spectral reconstruction accuracy. Since our method can synthesize high-quality hyperspectral data based on RGB data, at the same time generate sub-pixel labels, it may be of great help to real-world problems such as hyperspectral target detection and ground object classification. Our code is publically available at http://levir.buaa.edu.cn/Code.htm.

The contributions of this paper are summarized as follows:

1) We propose a new method for remote sensing hyperspectral image synthesis. Given an input high-resolution RGB image, the proposed method can not only produce realistic hyperspectral image data but also recover the causal factors behind each pixel location, including the ground object spectral signature as well as the abundance. As a comparison, most previous methods ignore the physics and can only produce spectral data.
2) Starting from the imaging mechanism and linear mixing model, the proposed method fully considers the effects of solar illumination and atmospheric absorption. The atmospheric absorption factors and the abundance map are solved as implicit variables without using per-image ground truth annotation. Although there are no auxiliary supervised signals added to the training process, experiments show that the estimated atmospheric absorption factors are consistent with the true measurements [25], which strongly suggests the rationality of our design.

The rest of the paper is organized as follows. Section II introduces the related works of HSI reconstruction. In section III, details of the proposed method are introduced. Section IV provides experiments on the effectiveness and rationality of the method. Finally, we conclude the method in Section V.

II. Related Work

In this section, we briefly review three groups of methods for hyperspectral image reconstruction, including hyperspectral image (spatial) super-resolution, image fusion, and spectral super-resolution.

A. Hyperspectral Image Super-Resolution

Hyperspectral image super-resolution aims at improving the spatial resolution of HSI while keeping the spectral data unchanged [26]. Many manual prior based methods use sparsity and image neighborhood dependence [11, 27] for spatial super-resolution, but they have limited model capacity and fail to recover more details. With the development of Convolution Neural Networks (CNNs), many super-resolution methods directly learn a mapping from low-resolution input to high-resolution output with CNNs [26, 28–32]. In addition to CNNs, some classical image analysis methods are also introduced to hyperspectral image super-resolution, such as Non-negative Matrix Factorization (NMF) [33] and Mutual Dirichlet Net [34]. Besides, single-image super-resolution methods mining the internal characteristics of HSI to restore spatial details with CNN [32] or Bayes energy minimization (EM) [35].

B. Image Fusion

Different from image spatial super-resolution, image fusion approaches aim at improving spatial resolution hyperspectral images with well-registered auxiliary high-resolution RGB/multispectral images (MSI) [36, 37]. These methods are also known as a variant group of super-resolution methods in recent literature [38–42]. Image fusion methods can be roughly divided into four categories: CNN based [36, 39, 41, 43], dictionary learning based [13, 37, 44], tensor factorization (TF) based [42, 45–49], and optimization based [38, 40, 50–52]. In CNN based methods, deep networks are used to model the degradation of hyperspectral images [36, 39, 43]. Dictionary learning based methods assume that the low-resolution HSI and high-resolution RGB/MSI share the same spatial sparse codes [37, 44]. Similarly, tensor factorization based methods decompose the images into different components to establish the relationship between HSI and RGB [42, 45, 46, 48, 49]. In optimization based methods, local-global similarity measurement is usually applied for the reconstruction process [38, 40]. In addition, for fusion-based methods, the high-resolution RGB and low-resolution hyperspectral image pairs are typically required, which are often difficult and expensive to obtain due to sensor limitations and image registration issues.

C. Spectral Super-resolution

Spectral super-resolution (SSR) reconstructs hyperspectral images by improving spectral information of high-resolution RGB/MSI while preserving the spatial size and details [53, 54]. Recent spectral super-resolution methods mainly utilize powerful CNN structures [55–60] and are trained to learn from an inverse mapping of the imaging degradation model [61]. Spectral Response Function (SRF) is used to guide band grouping and model deep spatial-spectral prior with an optimization-driven CNN [61]. Besides, coupled dictionary learning with different regularization terms are adopted to reconstruct spectral information [54, 62]. Apart from remote sensing hyperspectral applications, spectral super-resolution

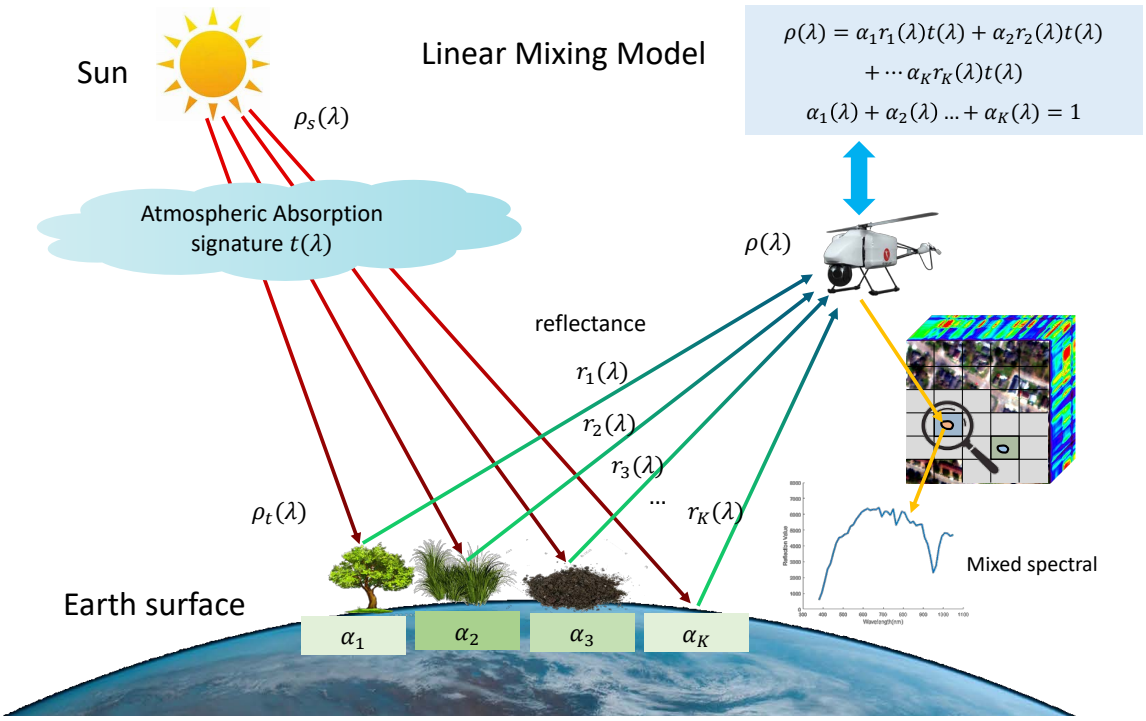


Fig. 1. Hyperspectral remote sensing imaging model. The sunlight $\rho_s(\lambda)$ passes through the atmospheric absorption $t(\lambda)$ and incidents to the ground $\rho_t(\lambda)$. Reflectance of different ground objects are represented as $r_1(\lambda)$, $r_2(\lambda)$, ... and their abundance in pixel area is represented as α_1, α_2, ... Finally, the sensor receives the reflected spectra and the spectral image is obtained according to the Linear Mixing Model.

has also increased broad attention in computer vision community. Many spectral super-resolution methods have been proposed and verified on nature scenes [14–18, 53, 63].

III. Proposed Method

In this section, we start from the hyperspectral remote sensing imaging model, and then introduce the details of abundance prediction, hyperspectral image reconstruction, and loss functions.

A. Imaging Model

Fig. 1 shows an overview of the hyperspectral remote sensing imaging model used in our method. In this paper, we follow [64] and assume the incident light hitting the ground objects is determined by the intensity of the sunlight and the atmospheric absorption:

$$\rho_t(\lambda) = \rho_r(\lambda) + \rho_{as}(\lambda) + t(\lambda)\rho_s(\lambda), \tag{1}$$

where, $\rho_t(\lambda)$ represents the spectral intensity after transmission at wavelength λ. $\rho_r(\lambda)$, $\rho_{as}(\lambda)$, $\rho_s(\lambda)$ represent the intensity contributions from the molecules (Rayleigh scattering), aerosols (including Rayleighaerosol interactions), and the sunlight before absorbed by the atmosphere. $t(\lambda)$ denotes the proportion of sunlight transmitted after atmospheric absorption.

For airborne hyperspectral sensors, the spectrum intensity $\rho_f(\lambda)$ received by the sensors can be written as a product of ground object reflectance and the incident light intensity at the ground surface [65]:

$$\rho_f(\lambda) = \rho_t(\lambda) r(\lambda), \tag{2}$$

where $\rho_f(\lambda)$ represents the spectrum intensity received by the sensor, $r(\lambda)$ represents the spectral reflectance of ground objects, and $\rho_t(\lambda)$ represents the sunlight intensity after atmospheric absorption. In vector form, the above equation can be written as follows:

$$\boldsymbol{y} = \boldsymbol{\varphi} \cdot \boldsymbol{r}, \tag{3}$$

where $\boldsymbol{y}$, $\boldsymbol{\varphi}$, and $\boldsymbol{r}$ are the vector representation of $\rho_f(\lambda)$, $\rho_t(\lambda)$, and $r(\lambda)$, respectively, at different wavelength. The notation $\cdot$ represents element wise multiplication of two vectors.

Due to the nature of hyperspectral imaging, remote sensing hyperspectral images usually have a limited spatial resolution. Therefore, spectral mixing is very common, which means the spectrum of a single-pixel may contain multiple ground objects. In this paper, we assume the spectral signature of a single-pixel follows the Linear Mixing Model [66]:

$$\begin{aligned} \boldsymbol{s}_e &= \sum_{i=1}^{N_g} \alpha_i \boldsymbol{r}_i + \boldsymbol{n}, \\ \alpha_i &\geq 0; \ \boldsymbol{\alpha}^\top \mathbf{1} = 1, \end{aligned} \tag{4}$$

where $\boldsymbol{s}_e \in \mathbb{R}^{K\times 1}$ represents spectrum after linear mixing. $\boldsymbol{R} \in \mathbb{R}^{K\times N_g} = [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{N_g}]$ are the spectral library of N_g objects, also known as endmember matrix [66]. α_i represents the proportion of i-th objects in the spectrum $\boldsymbol{s}_e$, also known as spectral abundance. $\boldsymbol{n} \in \mathbb{R}^{K\times 1}$ represents the perturbation including the noise and modeling errors.

Since the spectral library is usually measured in a laboratory environment, the Linear Mixing Model [66] is inaccurate

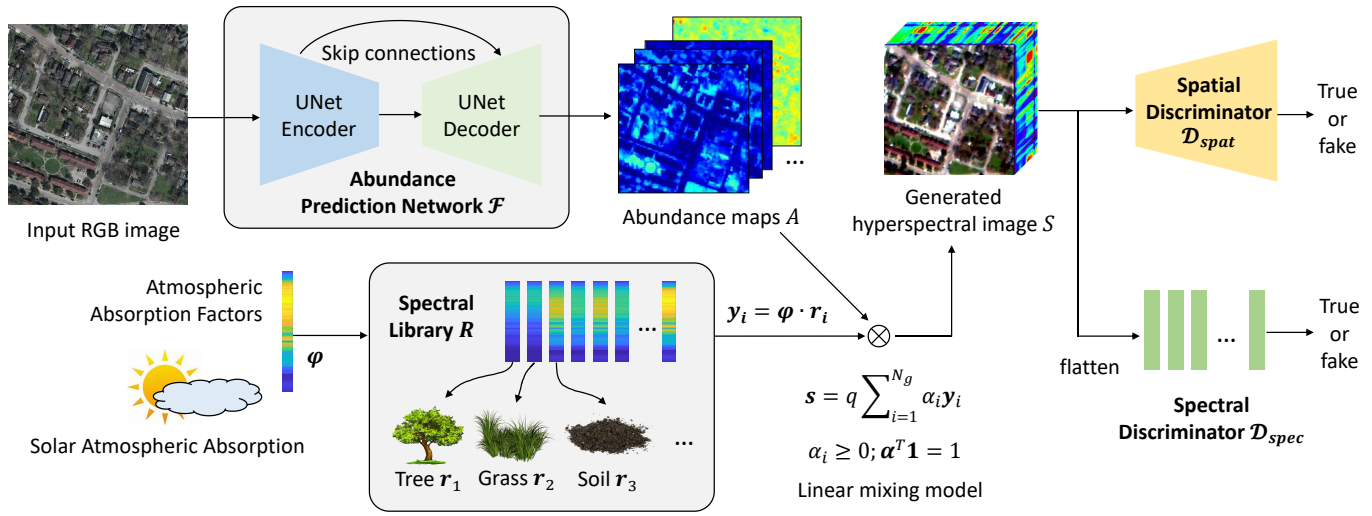


Fig. 2. An overview of the proposed method. Given an input RGB image, we introduce a U-Net based abundance prediction network to generate pixel-wise spectral abundance maps. With a spectral library and linear mixing model, the hyperspectral image can be constructed. We also introduce a spatial discriminator and a spectral discriminator to improve the realisticity and visual quality of the generated hyperspectral image. All networks can be trained in an end-to-end fashion with self-supervised reconstruction losses.

for airborne hyperspectral imaging tasks most of the time. Therefore, we consider the solar radiation and atmospheric absorption, and take the quantification into account. By combining Eq. 3 and Eq. 4, the final spectral intensity received by the sensor can be written as follows:

$$\begin{aligned} \boldsymbol{s} &= q\sum_{i=1}^{N_g}\alpha_i \boldsymbol{y}_i + \boldsymbol{n} \\ &= \boldsymbol{t}\cdot\sum_{i=1}^{N_g}\alpha_i \boldsymbol{r}_i + \boldsymbol{n}, \\ \alpha_i \geq 0;\ \boldsymbol{\alpha}^\top &\mathbf{1} = 1, \end{aligned} \tag{5}$$

where q is the quantitative coefficient to brige the gap between laboratory spectrum and remote sensing image, $\boldsymbol{t} = q\boldsymbol{\varphi}$ represents the atmospheric absorption factors with quantification correction. $\boldsymbol{y}_i = \boldsymbol{\varphi}\cdot\boldsymbol{r}_i$ denotes the reflectance spectrum of ith object.

B. Spectral Abundance Prediction and Image Reconstruction

In this paper, we focus on hyperspectral remote sensing image synthesis based on a single RGB image input. Instead of directly predicting the spectral reflectance, we start from an imaging model and predict the causal factors during the imaging processing, i.e., abundance map associated with the spectral library and the solar atmospheric absorption spectrum of sunlight. Finally, the spectral image can be constructed based on the linear mixing model after atmospheric absorption correction (Eq. 5). Fig. 2 shows an overview of the proposed method.

In our method, we introduce a conditional generative network for spectral abundance prediction. Given an input RGB image x, the generative network $\mathcal{F}$ is trained to recover the abundance maps A for each pixel location and each object:

$$A = \mathcal{F}(I|\boldsymbol{\theta}_F), \tag{6}$$

where $\boldsymbol{\theta}_f$ are trainable network parameters of $\mathcal{F}$, and $A \in \mathbb{R}^{N_g\times W\times H}$, where $W\times H$ are the spatial size of the hyperspectral image.

After we have the predicted abundance maps, the spectral data at pixel location l can be constructed with the spectral library and the abundance:

$$\begin{aligned} \boldsymbol{s}_l &= \boldsymbol{t}\cdot\sum_{i=1}^{N_g}\alpha_{l,i}\boldsymbol{r}_i \\ &= \boldsymbol{t}\cdot\sum_{i=1}^{N_g}\mathcal{F}(x|\boldsymbol{\theta}_F)_{l,i}\boldsymbol{r}_i, \end{aligned} \tag{7}$$

where $\alpha_{l,i} = \mathcal{F}(x|\boldsymbol{\theta}_F)_{l,i}$ represents the abundance value at the pixel location l and object number i. The noise term $\boldsymbol{n}$ is ignored during the reconstruction process.

C. Loss Function

The generative network $\mathcal{F}$ is trained in a self-supervised manner. Given a hyperspectral image S, we first sample from its channel dimension and compose a "spectral down-sampled" version $S(I)$ with only R, G, B channels. We then input I to $\mathcal{F}$ and enforce the reconstructed hyperspectral image S_r to be similar to the original image S as much as possible. To measure the similarity between S and S_r, we introduce three groups of loss functions: 1) pixel similarity loss, 2) spectral angle similarity loss, and 3) adversarial losses.

1) Pixel similarity losses: The pixel similarity loss is defined as the pixel-wise L1 distance between the input and the reconstructed hyperspectral image:

$$\begin{aligned} \mathcal{L}_{pxl} &= \mathbb{E}_{S\sim\mathcal{D}_S}\{\|S - S_r\|_1\}, \\ &= \mathbb{E}_{I\sim\mathcal{D}_I, l\sim\mathcal{I}_l}\{\|\boldsymbol{s}_l - \boldsymbol{t}\cdot\sum_{i=1}^{N_g}\mathcal{F}(S(I)|\boldsymbol{\theta}_F)_{l,i}\boldsymbol{r}_i\|_1\} \end{aligned} \tag{8}$$

where $\mathcal{D}_S$ is the training dataset of hyperspectral images. $\mathcal{I}_l$ means the total pixels in image S.

2) Spectral angle similarity loss: The spectral angle similarity loss is defined as the cosine similarity between the original spectral vector and the reconstructed one. The loss is written as follows:

$$\begin{aligned}\mathcal{L}_{\cos} &= \mathbb{E}_{S\sim\mathcal{D}_S}\{\cos < S, S_r >\}, \\ &= \mathbb{E}_{I\sim\mathcal{D}_I, l\sim I_l}\{\cos < \boldsymbol{s}_l, \boldsymbol{t}\cdot\sum_{i=1}^{N_g}\mathcal{F}(S(I)|\boldsymbol{\theta}_F)_{l,i}\boldsymbol{r}_i >\},\end{aligned} \tag{9}$$

where the cosine distance between two vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ is defined as follows:

$$\theta\left(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\right) = \arccos\frac{\boldsymbol{\beta}_1^T\boldsymbol{\beta}_2}{(\sqrt{\|\boldsymbol{\beta}_1\|_2^2\|\boldsymbol{\beta}_2\|_2^2}}. \tag{10}$$

3) Adversarial losses: The spectral abundance prediction we faced is naturally an ill-posed problem. Given one input RGB pixel, there could be multiple solutions for the underlying spectral abundance. With only the above pair-wise losses, we may have a blurring effect on the reconstructed images. To tackle this problem and improve the visual fidelity, we propose to train our networks under a conditional adversarial training framework [23, 24]. The joint discriminative learning introduced in our previous work [20] is adopted to improve the spatial-spectral realisticity and visual quality of the generated hyperspectral images.

We design two discriminators, a conditional spatial discriminator $\mathcal{D}_{spat}$, and a spectral discriminator $\mathcal{D}_{spec}$. The conditional spatial discriminator takes in the reconstructed hyperspectral image S_r or a real hyperspectral image S, and is trained to tell whether the input is generated (fake) or not (true). The spectral down-sampled RGB image $S(I)$ is also used as the conditional input and is concatenated with the spectral images. The spectral discriminator $\mathcal{D}_{spec}$ takes in the spectral vectors from S or S_r, and is also trained to tell whether they are authentically generated. The abundance prediction network $\mathcal{F}$ is also trained to fool the two discriminators and make the generated hyperspectral images as real as possible either from spatial or spectral dimensions. The adversarial losses for $\mathcal{D}_{spat}$ are defined as follows:

$$\begin{aligned}\mathcal{L}_{adv}^{spat} &= \mathbb{E}_{S\sim\mathcal{D}}\log\mathcal{D}_{spat}(S) \\ &\quad + \mathbb{E}_{S\sim\mathcal{D}}\log(1-\mathcal{D}_{spat}(S_r)),\end{aligned} \tag{11}$$

where

$$S_r(l) = \boldsymbol{t}\cdot\sum_{i=1}^{N_g}\mathcal{F}(x|\boldsymbol{\theta}_f)_{l,i}\boldsymbol{r}_i, \tag{12}$$

Similarly, for the spectral discriminator $\mathcal{D}_{spec}$, we have the following adversarial loss:

$$\begin{aligned}\mathcal{L}_{adv}^{spec} &= \mathbb{E}_{S\sim\mathcal{D}}\log\mathcal{D}_{spec}(S(l)) \\ &\quad + \mathbb{E}_{S\sim\mathcal{D}}\log(1-\mathcal{D}_{spec}(S_r(l))).\end{aligned} \tag{13}$$

The above adversarial losses can be trained with a minimax optimization process, where the generator tries to minimize this objective while the discriminators try to maximize it:

$$\min_{\mathcal{F}}\max_{\mathcal{D}_{spat},\mathcal{D}_{spec}}(\mathcal{L}_{adv}^{spat}+\mathcal{L}_{adv}^{spec}). \tag{14}$$

4) Total loss: Since all components of our networks are differentiable, the whole framework can be trained in an end-to-end fashion. The total loss function is written as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{pxl} + \lambda_1\mathcal{L}_{\cos} + \lambda_2\mathcal{L}_{adv}^{spat} + \lambda_3\mathcal{L}_{adv}^{spec}, \tag{15}$$

where λ_1, λ_2, and λ_3 are the pre-defined weights for balancing different loss terms. We set the solar atmospheric absorption as all trainable variables. The final loss functions are trained by solving the optimization problem below:

$$\boldsymbol{\theta}_F^*, \boldsymbol{t}^* = \arg\min_{\boldsymbol{\theta}_F,\boldsymbol{t}}\max_{\boldsymbol{\theta}_D^{spat},\boldsymbol{\theta}_D^{spec}}\mathcal{L}_{total}. \tag{16}$$

Note that although there are no losses or constraints attached on $\boldsymbol{t}$, it is trained as implicit variables all together with other network parameters.

D. Implementation Details

1) Spectral Library: We construct our spectral library based on the USGS Spectral Library Version 7 [21]. The library contains measured spectra, the spectra convolved to other spectrometer or imaging spectrometer characteristics, the spectra resampled to broad band multispectral sensors, and that oversampled to finer wavelength spacing [21]. Since the AVIRIS sensor has a high spectral resolution ($10nm$) and wide coverage of wavelength (0.4-2.5μm), we select the AVIRIS 2014 of the convolved spectra as our base spectral library where spectra captured by many different sensors are convolved to the AVIRIS sensor. The base spectral library consists of 7 types of object spectra including artificial materials, coatings, liquids, minerals, organic compounds, soils and mixtures, and vegetation. After removing those abnormal spectra, we further remove the mineral and organic categories since we mainly focus on urban scenes. In artificial materials, the spectra of some chemical reagents are also removed. There are 345 spectra left in our library. Detailed information of the spectral library is shown in Table I and the comparison of using different subsets of the library can be found in section IV-C.

TABLE I
NUMBER OF SPECTRA IN OUR SPECTRAL LIBRARY

Category	USGS-v7	Normal	Removed	Selected
Artificial Materials	290	263	263	84
Coatings	12	11	11	11
Liquids	24	14	14	14
Minerals	1276	877	–	0
Organic Compounds	360	142	–	0
Soils and Mixtures	209	164	164	164
Vegetation	286	72	72	72
Total	**2457**	**1543**	**524**	**345**

Considering the discrepancy between the USGS library and the data to synthesis, we calibrate all the spectra data by aligning their wavelength to the sensor of the synthesized data with linear interpolation. Different from [67] using the nearest neighbor method, which may cause local spectral distortion, linear interpolation can better preserve the spectral information and improve the accuracy of abundance inversion.

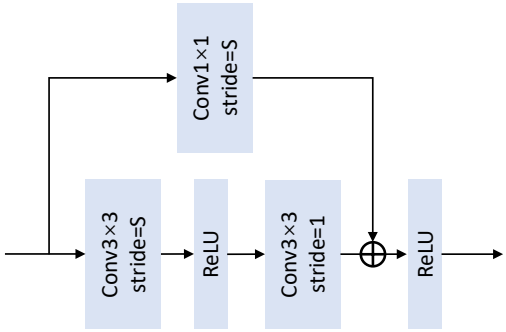


Fig. 3. Structure of resblocks in our abundance prediction network.

2) Abundance Prediction: To generate fine-grained abundance maps, we adopt U-Net [68] as the backbone architecture of our generative network $\mathcal{F}$. Skip-connections between different layers are adopted to fuse features of different semantic depths. The features are fused with element-wise addition. The backbone network consists of 6+6 residual blocks. We set the stride size to 2 for every layer consisting of two blocks, where the input image is downsampled $2^6 = 64$ times on its spatial dimension. The configuration of each residual block is shown in Fig. 3. When upsampling the feature maps, bilinear interpolation upsampling followed by a convolution layer is used instead of deconvolution or pixel-shuffle to avoid checkerboard artifacts. The final abundance maps are generated with a convolutional layer. Since the object abundance has non-negative values and the sum of the values for different objects in each pixel is 1, we add a softmax layer at the output-end of the network. The softmax normalization is performed along the channel dimension.

3) Training details: We randomly select 128×128 patches on the training image pairs for training. In our loss function, we set $\lambda_1 = 10$ and $\lambda_2 = \lambda_3 = 0.01$. To avoid model collapse, we update the discriminators after every 3 updates of other parameters. The networks are trained with the Adam optimizer [69]. We adopt cosine learning rate drop [70] after training 400 epochs using the initial learning rate and set the max-iteration number to 800. The initial learning rate is set to 10^{-4} for the abundance prediction network and 10^{-5} for the discriminators. The learning rate for solar atmospheric absorption $\boldsymbol{t}$ is set to 10^{-4}.

IV. Experiment

A. Datasets and Experimental Setup

We evaluate our method on the IEEE *grss_dfc_2018* [8] and GF5 datasets [20]. The IEEE *grss_dfc_2018* is collected by NCALM (National Center for Airborne Laser Mapping) from Houston University on February 16, 2017 [8, 9]. The hyperspectral data is acquired by an ITRES CASI 1500 (a VNIR sensor of ITRES company, which offers 1500 pixels across its field of view), covering a 380-1050nm spectral range with 48 bands. We use the hyperspectral data for HSI generation, the data has a spatial size of 4172×1202 pixels. We chose bands 23, 12, 5 from the hyperspectral images to construct RGB image input. The original hyperspectral images and downsampled RGB images are cropped into 27 paired patches of 512×512 pixels, with 24 pairs for training and 3 for testing. There are no overlaps between training and testing patches. Moreover, 3 test patches were divided into 12 256×256 patches for inference due to GPU memory limitation.

The GF5 dataset [20] has 6 scenes of HSI captured by the GF5 visible and near-infrared sensor (VNIR). The HSIs have 150 bands covering the wavelength range 390-1035nm and are cropped to 120 512×512 patches, where 115 for training and 5 for testing. Same experiment settings can be find in [20].

The experiment is conducted on a desktop PC with an Intel (R) Core (TM) i7-7700K CPU @ 4.20GHz and an NVIDIA GeForce GTX 1080 GPU card. The training process of PDASS takes about 5 hours and the testing process of a 256×256 image only takes 0.0317s. We compare our method with four state-of-the-art spectral super-resolution methods, including HSRNet [61], HSCNN+ [71], FMNet [15], and R2HGAN [20]. HSCNN+ [71] learns to map a RGB image directly to a hyperspectral image with Densely Connected Networks [72]. FMNet [15] uses the pixel-aware receptive field to integrate multi-layer features for the spectral super-resolution. HSRNet [61] reconstructs hyperspectral image with Spectral Response Functions (SRF). R2HGAN [20] recovers hyperspectral image under the GAN framework with joint discriminative learning. For fair competition, all these methods are optimized adequately and parameters for the best results are selected.

We choose multiple criteria, including RMSE (Root Mean Squared Error) [10, 53, 73], MRAE (Mean Relative Absolute Error) [14, 73], SAM (Spectral Angle Mapper) [55, 74], MSSIM (Mean Structural SIMmilarity) [10, 60, 75] and MPSNR (Mean Peak Signal-to-Noise Ratio) [10, 60, 75] as evaluation metrics. RMSE, MRAE and MPSNR are pixel-wise measures widely used for image super-resolution [43, 76]. SAM measures the shape similarity of the generated spectra and the real ones and has been widely used in hyperspectral image processing methods. MSSIM is a structural similarity index, which measures the mean SSIM of each band between generated HSIs and real ones. The detailed calculation of the indicators can be found in [20].

B. Comparison with Other Methods

Fig. 4 shows the comparison results between different methods on IEEE *grss_dfc_2018* [8] dataset. The false-color image (band 23, 12, and 5) of the generated hyperspectral images as well as their MPSNR are shown. We can see that the spectral details of the image generated by HSRNet [61] are completely lost and only part of the spatial structure is retained. HSCNN+ [71] generates hyperspectral images with a slight color deviation but most spatial information is consistent with the real one. FMNet [15] correctly restores RGB color information with a slight spatial distortion. The false-color images generated by R2HGAN [20] and PDASS (ours) are visually indistinguishable from the real images. However, our method has a higher reconstruction accuracy (MPSNR) than R2HGAN [20].

The generation results on other bands are shown in Fig. 5. HSRNet [61] loses most spatial information in most spectral

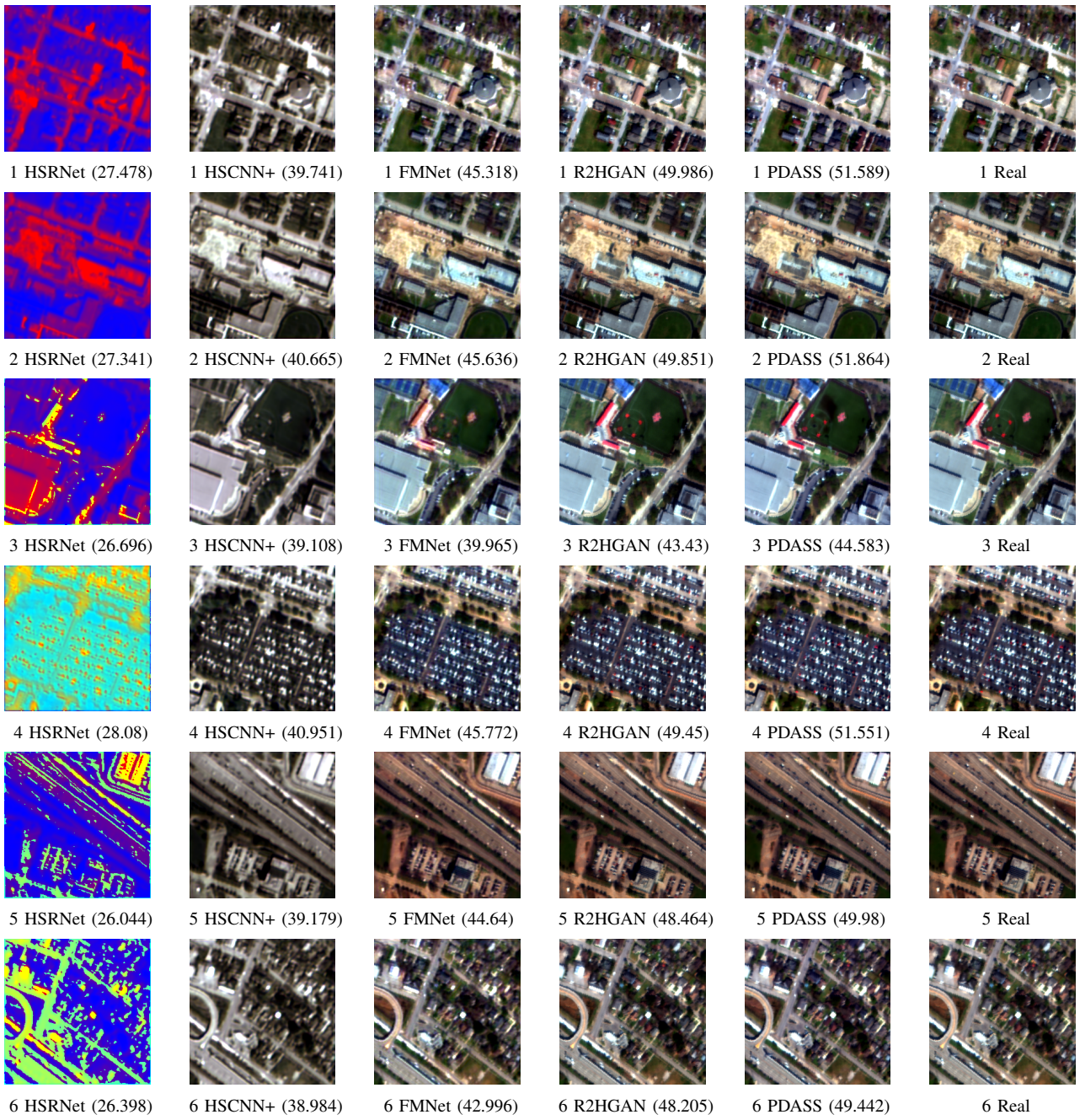


Fig. 4. False-color visualization (band No. 23, 12, and 5) of the generated hyperspectral image with different methods: HSRNet [61], HSCNN+ [71], FMNet [15], R2HGAN [20], and PDASS (ours). The test image ID and reconstruction MPSNR are also given. For example, 1 HSRNet (27.478) means the result of HSRNet on test image #1 with MPSNR = 27.478.

TABLE II
SPECTRAL RECONSTRUCTION ACCURACY OF DIFFERENT METHODS ON IEEE *grss_dfc_2018* [8] DATASET. FOR RMSE, MRAE, AND SAM, A LOWER SCORE INDICATES BETTER, WHILE FOR MSSIM AND MPSNR A HIGHER SCORE INDICATES BETTER.

Method	RMSE ↓	MRAE ↓	SAM ↓	MSSIM ↑	MPSNR ↑
HSRNet	19899.24	3.4615	1.1257	0.765	29.7306
HSCNN+	986.6544	0.1737	0.1515	0.9455	38.9456
FMNet	697.9392	0.1177	0.0875	0.9729	42.8291
R2HGAN	466.7432	**0.075**	0.0596	0.9861	46.8614
PDASS (Ours)	**406.3703**	0.076	**0.0553**	**0.9879**	**47.5641**

bands. For the recovery of the first band, HSCNN+ [71], FMNet [15], and R2HGAN [20] produce noisy band images and only recover part of the spatial structure. PDASS (Ours) produces results with the best visual quality and has less noise than all the other methods. Table II shows the performance of different comparison methods on different metrics, including MRAE, RMSE, SAM, MSSIM and MPSNR. The spectral curves generated by the methods are shown in Fig. 6. From Table II and Fig. 6, we can see that HSRNet [61] fails to produce reasonable results where the pixel-wise error is even three times higher than the real spectral reflectance

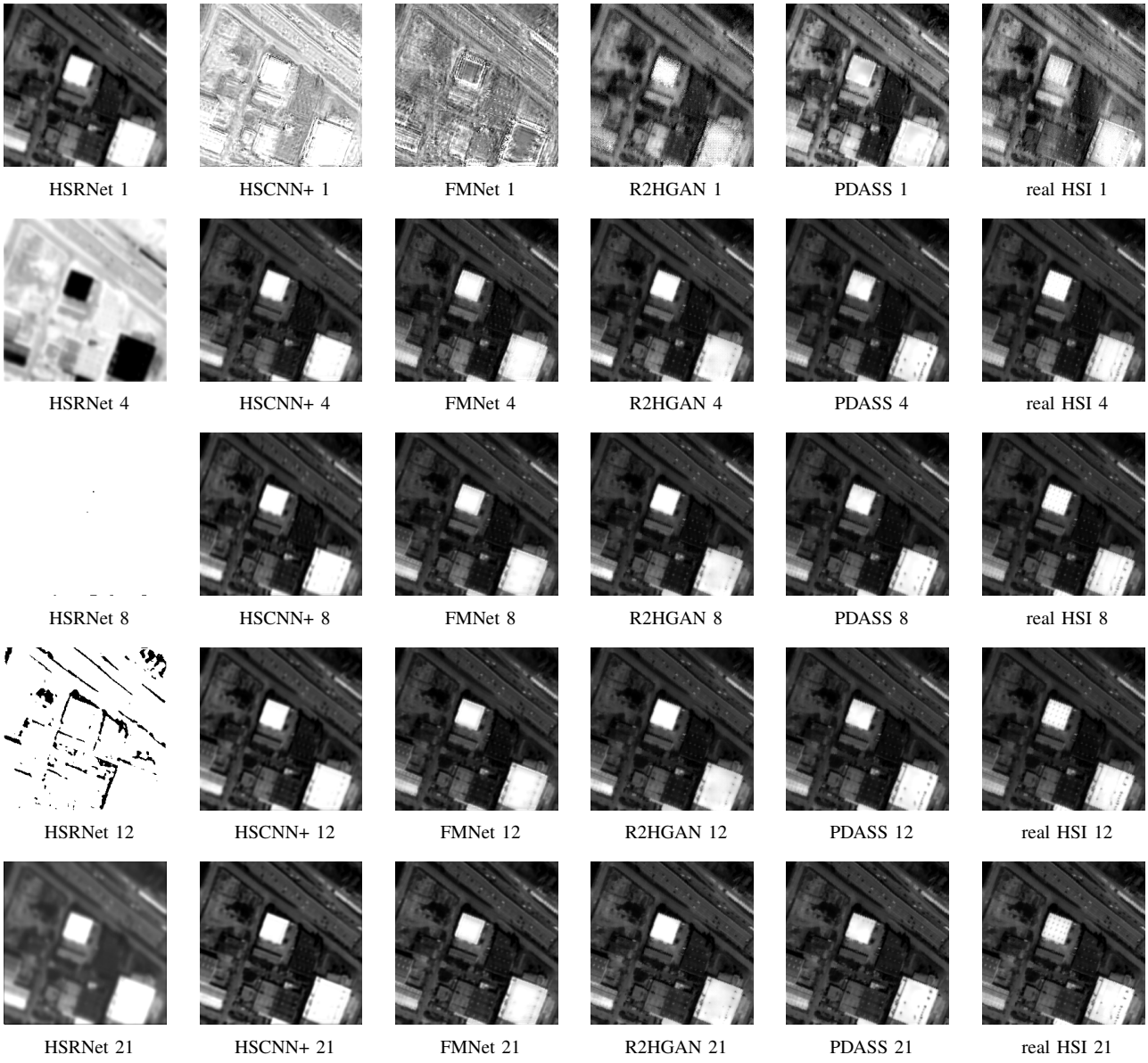
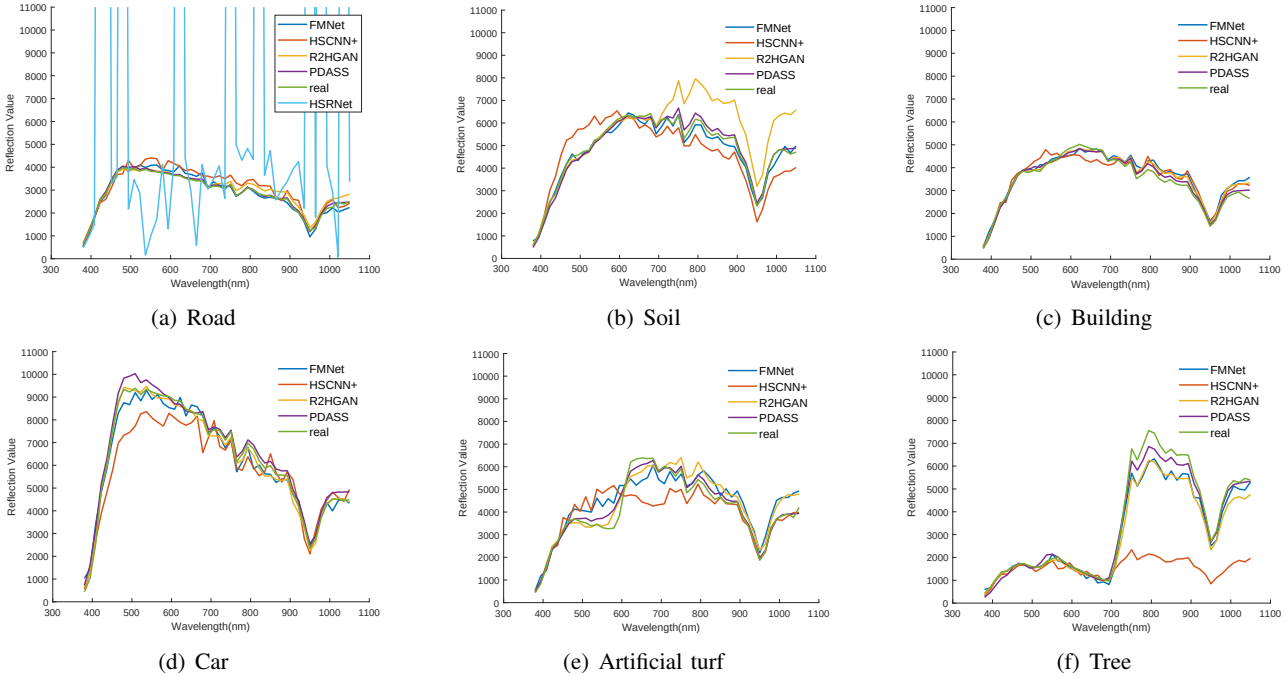


Fig. 5. Different bands of the generated hyperspectral images. Each row shows a particular band generated by different methods: HSRNet [61], HSCNN+ [71], FMNet [15], R2HGAN [20], and PDASS (ours). Each column show different bands. The test image ID and reconstruction PSNR are also given.

and a numerical overflow has occurred on those generated abnormal spectral peaks. HSCNN+ [71] uses DenseNet [72] as its network backbone to learn the mapping between input RGB images and the output hyperspectral images, and gets reliable MPSNR (38.95) and MSSIM (0.9455). The spectral curve generated by HSCNN+ [71] has a similar shape to the real spectrum, but the response values are far from the real one. FMNet [15] designs pixel-aware receptive field and improves all the five indicators compared to HSCNN+ [71]. Particularly, the SAM of the FMNet [15] decreases 42% from that of HSCNN+ [71] and the spectra in Fig. 6 are much more closer to the real ones than HSCNN+ [71]. However, for some ground objects such as road in (a) and artificial turf in (e), the spectra of FMNet [15] of them still have clear shape errors at some wavelength compared with real ones. R2HGAN [20] and PDASS (Ours) have better reconstruction accuracy, compared to FMNet [15], the MPSNR of R2HGAN [20] has been improved from 42.8291 to 46.8614. The spectra generated by R2HGAN [20] are closer to the real ones than HSCNN+ [71] and FMNet [15] as shown in Fig. 6. There are still some cases that R2HGAN [20] behaves not very well. For example, the spectrum of soil in (b) generated by R2HGAN is quite different from the true spectrum at wavelength 700-1050nm. Compared to R2HGAN, the proposed PDASS recovers spectra through the abundance inversion of objects in the spectral library, so that the spectra have more actual physical meaning. The reconstruction accuracy outperforms R2HGAN [20] and other methods except on MRAE. The MRAE of PDASS (Ours) is 0.076, which is quite close to the best 0.075 of R2HGAN [20]. The spectra recovered are closest to the real ones as shown in Fig. 6.

The reconstruction accuracy of the GF5 dataset is shown



(a) Road (b) Soil (c) Building

(d) Car (e) Artificial turf (f) Tree

Fig. 6. Spectral curves generated by different methods: HSRNet [61], HSCNN+ [71], FMNet [15], R2HGAN [20], and PDASS (ours). Because HSRNet has a numerical overflow problem, we only visualize its result in the first sub-figure for easy comparison.

in Table III. HSCNN+ [71], HSRNet [61] and FMNet [15] have unreliable generation with MRAE>5. R2HGAN [20] and PDASS (Ours) have similar synthesis accuracy with MPSNR over 60. Although the performance of PDASS (Ours) is slightly lower than that of R2HGAN [20] except RMSE, PDASS (Ours) achieves a significant improvement on the RMSE accuracy, from 178.3 to 100.84 (lower is better). Most importantly, PDASS (Ours) recovers the per-pixel feature abundance while achieving comparable results to R2HGAN, while R2HGAN can not.

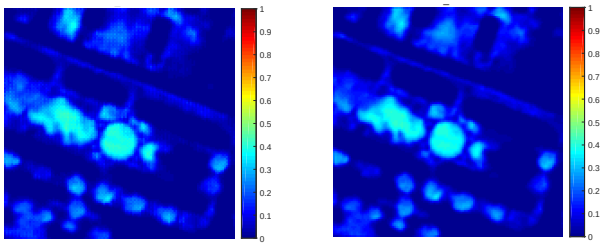
TABLE III
SPECTRAL RECONSTRUCTION ACCURACY OF DIFFERENT METHODS ON GF5 DATASET [20].

Method	RMSE ↓	MRAE ↓	SAM ↓	MSSIM ↑	MPSNR ↑
HSCNN+	2241.99	6.8294	0.192	0.9432	46.273
HSRNet	2625.23	5.9974	0.4832	0.9426	48.166
FMNet	7793.36	40.7865	0.4842	0.8741	42.792
R2HGAN	178.3	**0.126**	**0.0435**	**0.9972**	**61.479**
PDASS (Ours)	**100.84**	0.1342	0.0608	0.9956	60.218

C. Ablation Studies

We conduct ablation studies on different technical components of our method, including the cosine similarity loss, discriminative, network architecture design, and latent solar atmospheric absorption with quantification factor. Table IV shows the result of all ablation experimental results. The spectra generated by the proposed method with different configurations are shown in Fig. 7 and 8.

1) Similarity Loss: In experiment 4 of Table IV, we remove $\mathcal{L}_{\cos}$ from the loss function. The reconstruction accuracy decreases sharply after removing $\mathcal{L}_{\cos}$. For example, the SAM error increased by 10% (0.0553 to 0.0605). In Fig. 7, we show the spectra generated without $\mathcal{L}_{\cos}$ loss (marked by **wo-ls**). Comparing with our full implementation, the **wo-ls** has much larger spectral variation compared to the real one particularly at wavelength 1000-1050nm.

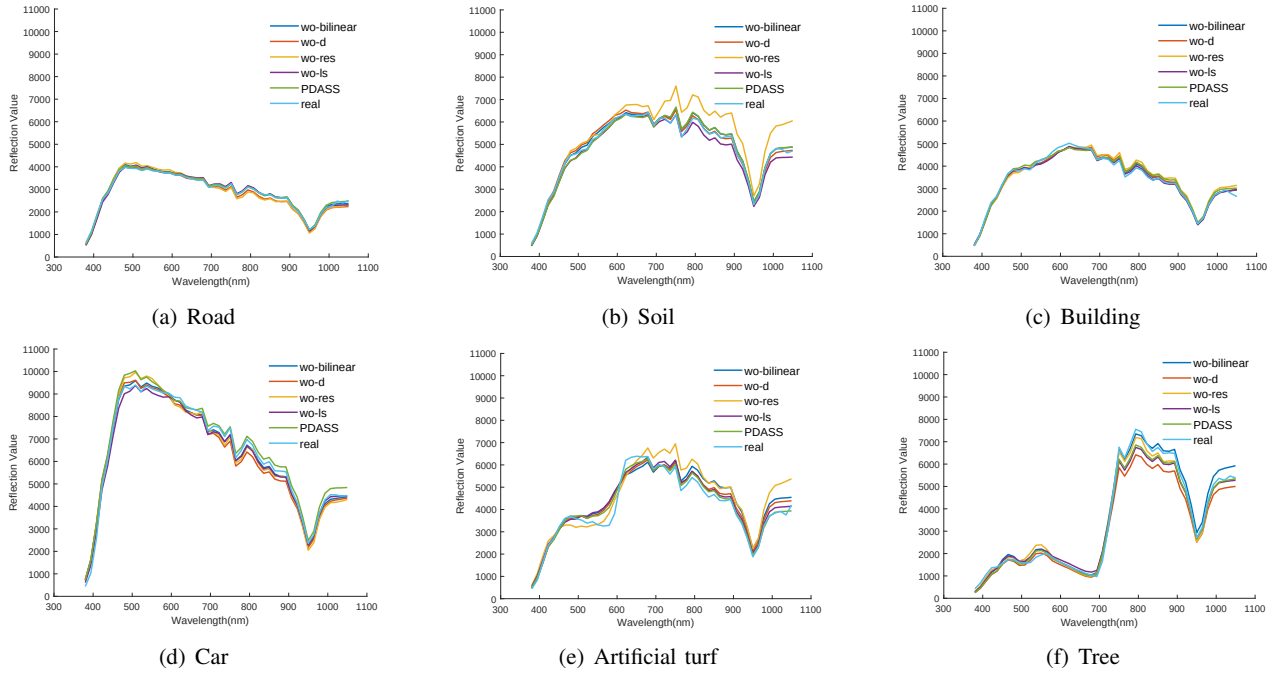


(a) Deconv (b) Bilinear+Conv

Fig. 9. Abundance maps ("Lawn_Grass") generated by $\mathcal{F}$ under two upsampling configurations: (a) deconvolution, and (b) bilinear upsampling + convolution. There is a noticeable checkerboard artifact produced by the deconvolution layers.

2) Network Architecture: We design residual blocks for feature extraction in our abundance prediction networks. In experiment 5 of Table IV, we replace the resblocks with 4×4, $stride = 2$ standard convolution layers. After the replacement, the MPSNR drops sharply, which suggests that fine-grained feature extraction from RGB images is crucial for pixel-wise and ill-posed hyperspectral image generation process. In Fig. 7, the curves marked with **wo-res** show the result of removal of the resblocks. We can see the removal causes a clear deviation in the recovered spectral curves.

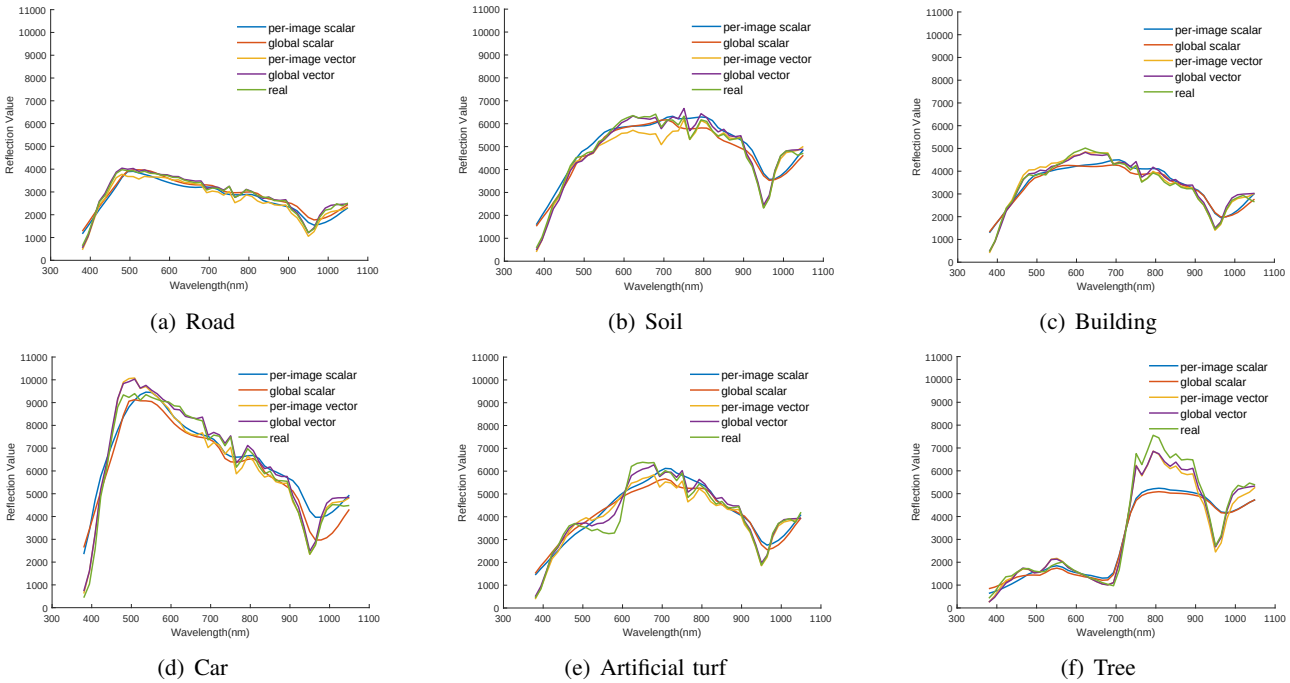
We also replaced the bilinear interpolation with deconvolution layers in U-Net (experiment 6 of Table IV). As a result, in Fig.9, we can see that the abundance map generated by



(a) Road (b) Soil (c) Building

(d) Car (e) Artificial turf (f) Tree

Fig. 7. Comparison of generated spectra by different methods for an ablation study. (a-f) show the generated spectra on different ground objects. "wo-bilinear" represents replacing the bilinear upsample with deconvolution. "wo-d" denotes the removal of the discriminators. "wo-res" denotes removing resblocks in our abundance prediction network and replacing them with standard convolutions. "wo-ls" means removing the $L_{\cos}$ loss term.



(a) Road (b) Soil (c) Building

(d) Car (e) Artificial turf (f) Tree

Fig. 8. Comparison of generated spectra by different $\boldsymbol{t}$ learning strategies. "per-image scalar" means t is a trainable scalar and is optimized separately for every image. "global scalar" means t is a trainable scalar and is optimized as a global variable on the whole dataset. "per-image vector" means $\boldsymbol{t}$ is a trainable scalar and is optimized separately for every image. "global vector" means $\boldsymbol{t}$ is a trainable vector and is optimized as a global variable on the whole dataset (our final implementation).

TABLE IV
ABLATION STUDIES OF DIFFERENT TECHNICAL COMPONENTS OF THE PROPOSED METHOD.

Name	$\boldsymbol{t}$ vector	$\boldsymbol{t}$ global	$\mathcal{L}_{cos}$	resblock	Up+conv	$\mathcal{D}_{spat}, \mathcal{D}_{spec}$	RMSE ↓	MRAE ↓	SAM ↓	MSSIM ↑	MPSNR ↑
1	✗	✗					625.5191	0.1718	0.1222	0.9698	42.5559
2	✗	✓					615.8163	0.1738	0.1239	0.9707	43.0278
3	✓	✗					430.9514	0.093	0.0663	0.9838	46.2568
4			✗				430.4892	0.0837	0.0605	0.9861	46.8193
5				✗			581.1201	0.1099	0.0685	0.9804	43.2282
6					✗		417.9923	0.0781	0.0565	0.9876	**47.7349**
7						✗	419.1033	0.0809	0.0576	0.9867	47.051
PDASS	✓	✓					**406.3703**	**0.076**	**0.0553**	**0.9879**	47.5641

deconvolution layers has clear checkerboard artifacts. The **wo-bilinear** curves in Fig. 7 show the spectra recovered with deconvolution instead of bilinear upsampling. A similar effect can be observed after the replacement and bilinear interpolation has more accurate spectral reconstruction, particularly for the artificial turf and trees (Fig. 7(e,f)) at wavelength 700-1050nm.

3) Adversarial training: In experiment 7 of Table IV, we evaluate the effect of with or without adversarial training. The **wo-d** curves in Fig. 7 show the influence after removing the spatial discriminator $\mathcal{D}_{spat}$ and spectral discriminator $\mathcal{D}_{spec}$. We can find the adversarial training improved the authenticity of the generated spectra. MSSIM decreases from 0.9879 to 0.9867 and MPSNR decreases from 47.564 to 47.051 after removing the discriminators.

4) Solar atmospheric absorption with quantification factor: When learning the atmospheric absorption with quantification correction factor $\boldsymbol{t}$, we can either optimize it as a vector with a factor corresponding to the element of each band, or we can simply set it as a trainable scalar. The column of "$\boldsymbol{t}$ vector" in Table IV shows the results of the above two configurations. We also investigate whether we should estimate a separate $\boldsymbol{t}$ for each image or on the whole dataset as a global trainable vector (column "$\boldsymbol{t}$ global"). In Table IV, experiment 1, 2, 3 and PDASS show the comparison of different design of the solar atmospheric absorption factors. We can see estimating a separate absorption value for each band ($\boldsymbol{t}$ vector) gives to better reconstruction. Meanwhile, our full implementation has a 1.3 improvement on MPSNR compared with experiment 3, which suggests that estimating $\boldsymbol{t}$ as a global vector gives a better result than estimating in an image-by-image manner.

Fig. 8 plots the spectra generated from different $\boldsymbol{t}$ designs. The **per-image scalar**, **global scalar**, **per-image vector**, **global vector** curves correspond to experiment 1,2,3 and our full implementation. We can find that estimating $\boldsymbol{t}$ as a scalar will produce over-smooth spectral curves and can not produce faithful spectra especially near the atmospheric absorption bands (wavelength 900-1050nm).

5) Spectral library: We compare the reconstruction accuracy of different spectral library selection. Note that a higher reconstruction accuracy of PSNR does not necessarily mean a higher accuracy for downstream tasks. Therefore, in addition to the reconstruction accuracy, the accuracy of a downstream segmentation task is also used as the criterion for selecting the spectral library. The Mean Intersection over Union (mIoU) [77] between the segmentation output and the ground truth is used as the criteria, where a higher mIoU means a better spectral library quality. As shown in Table V, N_g=1543 denotes using all the normal spectra in the library and it reaches the best reconstruction accuracy, but causes fallacious abundance and the mIoU of the downstream task is lower than others. N_g=524 represents removing the two categories and reduces the degree of freedom of abundance regression to about 1/3 of the original (1543 to 524). When further removing some chemical reagents and using 345 spectra as the library, the reconstruction accuracy and the mIoU of the downstream task basically remain unchanged as N_g=524 and even have a slight improvement. This may be because the difficulty of abundance regression reduces and avoids illogical abundance after removing unreasonable spectra. To ensure the completeness of the spectral library, we no longer remove spectra for experiments.

TABLE V
COMPARISON OF DIFFERENT SPECTRAL LIBRARY SELECTION

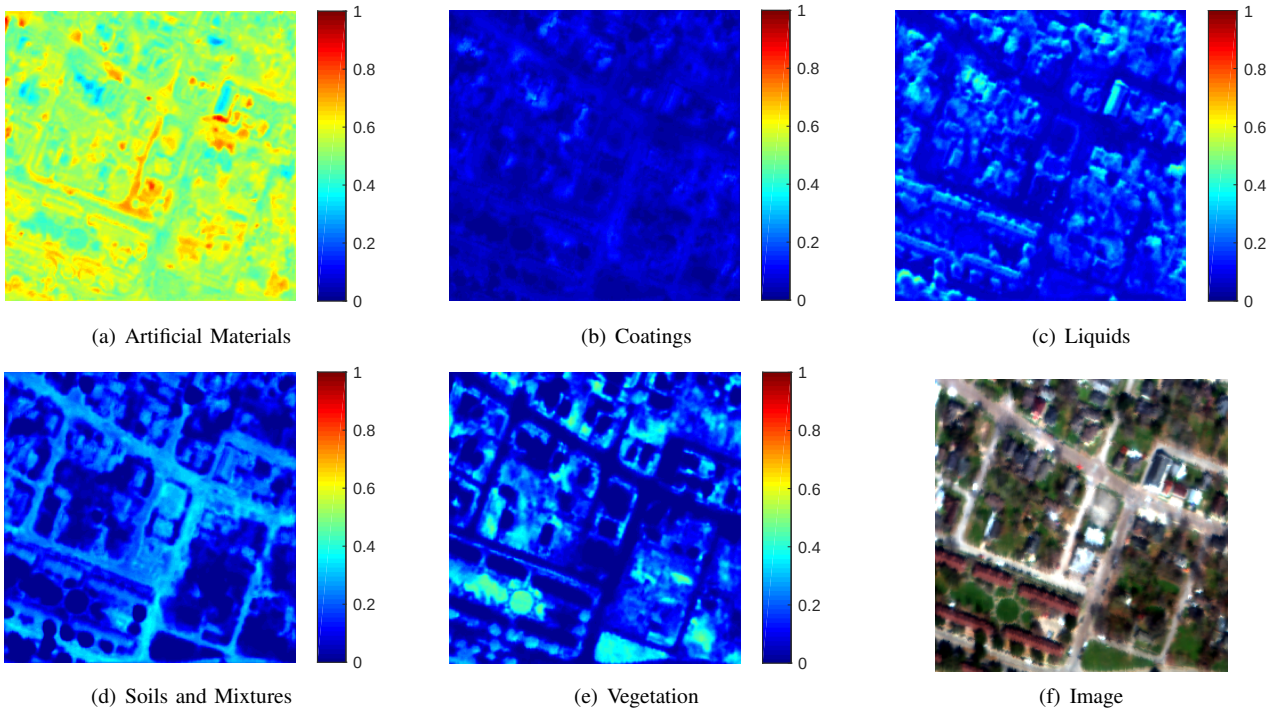
N_g in $\mathcal{R}$	RMSE ↓	SAM ↓	MSSIM ↑	MPSNR ↑	mIoU ↑
1543	404.75	0.0546	0.9882	48.1449	0.2074
524	416.62	0.0568	0.9874	47.4338	0.2098
345	406.37	0.0553	0.9879	47.5641	0.2105

Considering the rationality of abundance, problem difficulty, library completeness, reconstruction accuracy and performance on the downstream task, we chose N_g=345.

D. Analysis of Latent Variables

In the top of Fig 10, we show the estimated solar atmospheric absorption factors by using our method (recovered solar atmospheric absorption with quantification factors). In the bottom of Fig 10, we show the true spectrum of solar radiation [25] as well as the measured absorption factors. We can see that the recovered absorption factors are consistent with the true reference, although there is no supervised loss or constraint attached to these variables during training. Our method learns the irradiance rising from wavelength at 380nm, max out at wavelength around 500nm, and decrement till 1050nm. Also, the absorption peak near wavelength at 750nm for oxygen (O_2) and 950nm for water (H_2O) are visible in the estimated absorption curve.

We add the abundance maps according to the categories of the library and show them in Fig. 11. The recovered abundance



(a) Artificial Materials (b) Coatings (c) Liquids

(d) Soils and Mixtures (e) Vegetation (f) Image

Fig. 11. Visualization of the recovered abundance map on different ground object categories: (a) artificial materials, (b) coatings, (c) liquids, (d) soils and mixtures, and (e) vegetation. (f) shows the input image.

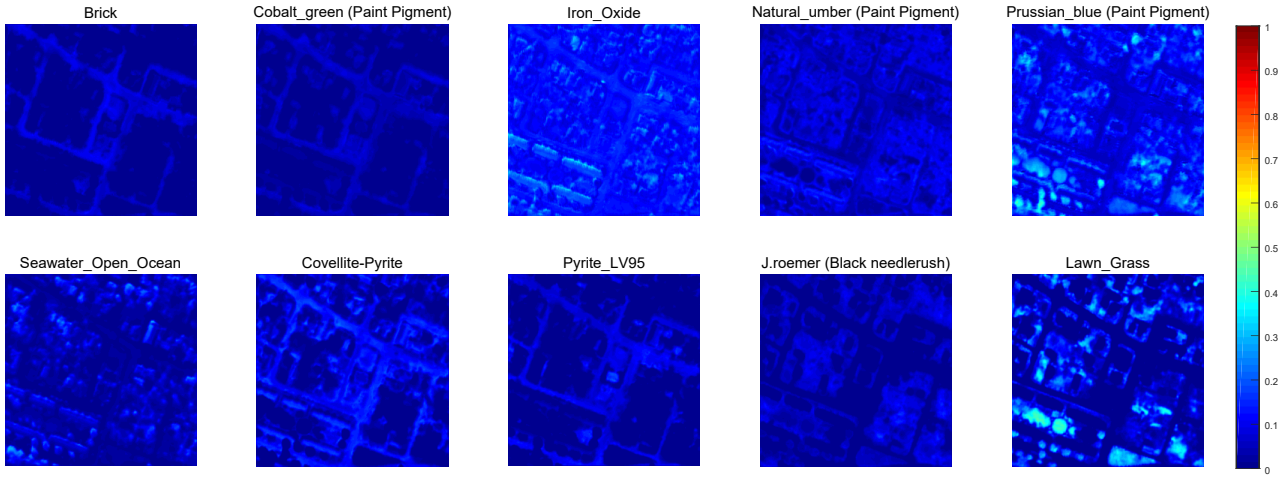


Fig. 12. Visualization of the recovered abundance map on 10 ground objects. We select 10 high abundance maps from the abundance of each object. The name of the abundance map represents the corresponding object.

map is consistent with the distribution of the ground objects. For example, Fig. 11(a) shows the abundance map of all objects belonging to artificial materials. The buildings and cars all have a high-value response (mostly >0.6) since they all belong to artificial materials. The coatings have a very low response because we see less paint from an aerial view. The abundance of soils and mixtures mainly appears on roads, it is mainly because the roads are paved with a mixture of artificial materials and soil mixtures such as gravel, stones, etc. The grass and tree have large responses in the abundance map of vegetation. Since there is almost no water region, the abundance of liquids is very low at most pixels, except the pixels under the shadow. The above visualization suggests that the abundance maps generated by our method are reasonable and have significant physical meaning.

We also provide a per-spectral-abundance shown in Fig. 12, where 10 relatively high abundance of the features is visualized. The abundance of the ‘brick’ is mainly distributed on the road. The abundance map of ‘iron oxide’ is relatively high in some tin roof areas. The ‘covellite pyrite’ and ‘pyrite LV95’ all have high abundance distribution on both buildings and roads since pyrite is the most distributed sulfide in the earth’s crust and is mainly used for buildings and roads. ‘Black needlerush’ is found mainly in afforest areas and ‘lawn grass’

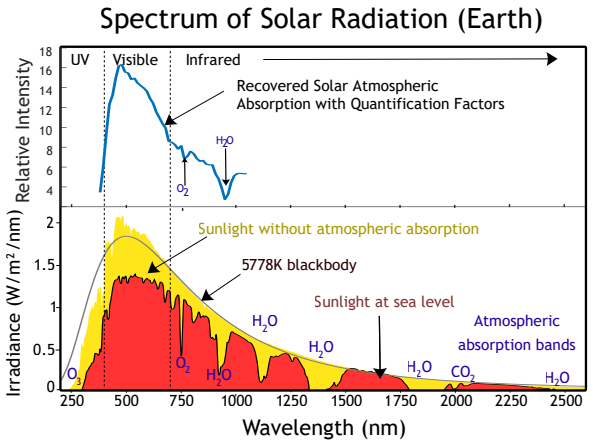


Fig. 10. Estimated solar atmospheric absorption factors and their true values [25].

has the highest abundance on trimmed grass areas. These show the rationality of abundance distribution. Although, there are also some ground features with unreasonable abundance. For example, the paint pigment of 'Prussian blue' has a high abundance on the tree and grass. Simultaneously, the paint pigment of 'natural umber' has a global high abundance distribution.

E. Downstream Task Verification

To verify the effectiveness of the generated spectra for downstream tasks, we re-divide the training and testing set of IEEE *grss_dfc_2018* [8]. All the labeled samples are used for classification training and testing while those un-annotated are used for training the HSI generator. We designed a U-Net [68] with 5+5 residual blocks in Fig. 3 as the backbone for the 20-class object (spectral) classification task. 75% of all the labeled pixels are used for training and the rest for testing.

Table VI shows the accuracy of different input data, where Macro-F1, overall accuracy (OA) and Mean Intersection over Union (mIoU) are chosen as indicators [77, 78]. It can be found that when using input synthesized HSI instead of the original RGB as input, both the classification criteria have a clear improvement. The mIoU increased 0.0122 from 0.1983 to 0.2105. Typically, there is a 3.6% increment on OA even exceeding that of the real HSI. Although less effective than using real HSI, synthesis HSI shows great potential in improving the accuracy of RGB segmentation.

TABLE VI
PERFORMANCE OF DIFFERENT DATA ON THE CLASSIFICATION TASK

Input data	F1 ↑	OA ↑	mIoU ↑
RGB	0.2351	0.8042	0.1983
Synthesis HSI	0.2490	0.8400	0.2105
Real HSI	0.2810	0.8350	0.2327

V. CONCLUSION

Spectral mixing generally exists in remote sensing hyperspectral images due to the low spatial resolution of airborne and spaceborne hyperspectral sensors. We propose a hyperspectral remote sensing image synthesis method based on spectral library and conditional RGB input images. Instead of directly recovering the spectral reflectance, we start from the hyperspectral imaging model and predict the subpixel level abundance map of ground objects. The hyperspectral data thus can be constructed based on the predicted abundance, spectral library, the solar atmospheric absorption factors, and the linear mixing model, with clear physical significance. The following experiments suggest the superiority and rationality of our method. First, on the IEEE *grss_dfc_2018* dataset, our method achieves the best reconstruction accuracy compared to previous state-of-the-art methods with an MPSNR of 47.56. Second, the predicted abundance maps have a clear physical meaning and are consistent with the distribution of ground objects. Third, the estimated solar atmospheric absorption factors are consistent with the true measurement. Finally, an extensive ablation study verifies the effectiveness of our design. Since our method can synthesize high-quality hyperspectral data based on RGB data, at the same time generate sub-pixel labels, our method may be of great help to real-world problems such as hyperspectral target detection and ground object classification. In the future, the direction of our efforts lies in two parts, the first one is beyond the LMM [79] and the second one is more reasonable abundance distribution constraints.

VI. ACKNOWLEDGEMENT

The authors would like to thank the National Center for Airborne Laser Mapping and the Hyperspectral Image Analysis Laboratory at the University of Houston for acquiring and providing the data used in this study, and the IEEE GRSS Image Analysis and Data Fusion Technical Committee.

REFERENCES

[1] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, 2017.

[2] J. Li, P. Gamba, and A. Plaza, "A novel semi-supervised method for obtaining finer resolution urban extents exploiting coarser resolution maps," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 10, pp. 4276–4287, 2014.

[3] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5966–5978, 2021.

[4] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "Spectralformer: Rethinking hyperspectral image classification with transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[5] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4340–4354, 2021.

[6] L. Liang, L. Di, L. Zhang, M. Deng, Z. Qin, S. Zhao, and H. Lin, "Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method," *Remote Sensing of Environment*, vol. 165, pp. 123–134, 2015.

[7] X. Yang and Y. Yu, "Estimating soil salinity under various moisture conditions: An experimental study," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 2525–2533, 2017.

[8] 2018 IEEE GRSS Data Fusion Contest. Online: http://www.grss-ieee.org/community/technical-committees/data-fusion.

[9] Y. Xu, B. Du, L. Zhang, D. Cerra, M. Pato, E. Carmona, S. Prasad, N. Yokoya, R. Hänsch, and B. Le Saux, "Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 6, pp. 1709–1724, 2019.

[10] 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 Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9010–9024, 2019.

[11] T. Akgun, Y. Altunbasak, and R. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1860–1875, 2005.

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

[13] M. A. Veganzones, M. Simes, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 274–288, 2016.

[14] J. Li, C. Wu, R. Song, W. Xie, C. Ge, B. Li, and Y. Li, "Hybrid 2-D-3-D deep residual attentional network with structure tensor constraints for spectral super-resolution of RGB images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2321–2335, 2021.

[15] L. Zhang, Z. Lang, P. Wang, W. Wei, S. Liao, L. Shao, and Y. Zhang, "Pixel-aware deep function-mixture network for spectral super-resolution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 821–12 828.

[16] S. Nie, L. Gu, Y. Zheng, A. Lam, N. Ono, and I. Sato, "Deeply learned filter response functions for hyperspectral reconstruction," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4767–4776.

[17] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang, "Joint camera spectral response selection and hyperspectral image recovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 256–272, 2022.

[18] R. Hang, Q. Liu, and Z. Li, "Spectral super-resolution network guided by intrinsic properties of hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 30, pp. 7256–7265, 2021.

[19] J. Li, R. Cui, B. Li, R. Song, Y. Li, Y. Dai, and Q. Du, "Hyperspectral image super-resolution by band attention through adversarial learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 6, pp. 4304–4318, 2020.

[20] L. Liu, S. Lei, Z. Shi, N. Zhang, and X. Zhu, "Hyperspectral remote sensing imagery generation from RGB images based on joint discrimination," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7624–7636, 2021.

[21] R. F. Kokaly, R. N. Clark, G. A. Swayze, K. E. Livo, T. M. Hoefen, N. C. Pearson, R. A. Wise, W. M. Benzel, H. A. Lowers, R. L. Driscoll, and A. J. Klein, "Usgs spectral library version 7," *Data Series*, 2017.

[22] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[23] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[24] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[25] F. Santos, A. Bühler, N. Filho, and D. Zambra, "A importância da determinação do espectro da radiação local para um correto dimensionamento das tecnologias de conversão," *Avances en Energías Renovables y Medio Ambiente*, vol. 19, pp. 11.43–11.54, 10 2015.

[26] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised adaptation learning for hyperspectral imagery super-resolution," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 3070–3079.

[27] R. C. Patel and M. V. Joshi, "Super-resolution of hyperspectral images: Use of optimum wavelet filter coefficients and sparsity regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 1728–1736, 2015.

[28] J. Hu, Y. Li, and W. Xie, "Hyperspectral image super-resolution by spectral difference learning and spatial error correction," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1825–1829, 2017.

[29] J. Hu, X. Jia, Y. Li, G. He, and M. Zhao, "Hyperspectral image super-resolution via intrafusion network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7459–7471, 2020.

[30] J. Hu, Y. Tang, and S. Fan, "Hyperspectral image super resolution based on multiscale feature fusion and aggregation network with 3-d convolution," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5180–5193, 2020.

[31] X. Wang, J. Ma, and J. Jiang, "Hyperspectral image super-resolution via recurrent feedback embedding and spatial-spectral consistency regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[32] P. V. Arun, K. M. Buddhiraju, A. Porwal, and J. Chanussot, "CNN-based super-resolution of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 9, pp. 6106–6121, 2020.

[33] W. Xie, X. Jia, Y. Li, and J. Lei, "Hyperspectral image super-resolution using deep feature matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 6055–6067, 2019.

[34] Y. Qu, H. Qi, C. Kwan, N. Yokoya, and J. Chanussot, "Unsupervised and unregistered hyperspectral image super-resolution with mutual dirichlet-net," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–18, 2021.

[35] H. Irmak, G. B. Akar, and S. E. Yuksel, "A MAP-based approach for hyperspectral imagery super-resolution," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2942–2951, 2018.

[36] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang, "Hyperspectral image super-resolution with optimized RGB guidance," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11 653–11 662.

[37] H. Kwon and Y.-W. Tai, "RGB-guided hyperspectral image upsampling," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 307–315.

[38] R. Wu, W.-K. Ma, X. Fu, and Q. Li, "Hyperspectral super-resolution via global-local low-rank matrix estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7125–7140, 2020.

[39] K. Zheng, L. Gao, W. Liao, D. Hong, B. Zhang, X. Cui, and J. Chanussot, "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2487–2502, 2021.

[40] X.-H. Han, B. Shi, and Y. Zheng, "Self-similarity constrained sparse representation for hyperspectral image super-resolution," *IEEE Transactions on Image Processing*, vol. 27, no. 11, pp.

5625–5637, 2018.
[41] W. Wei, J. Nie, L. Zhang, and Y. Zhang, "Unsupervised recurrent hyperspectral imagery super-resolution using pixel-aware refinement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.
[42] W. Wan, W. Guo, H. Huang, and J. Liu, "Nonnegative and nonlocal sparse tensor factorization-based hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8384–8394, 2020.
[43] L. Zhang, J. Nie, W. Wei, Y. Li, and Y. Zhang, "Deep blind hyperspectral image super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 6, pp. 2388–2400, 2021.
[44] C. Yi, Y.-Q. Zhao, and J. C.-W. Chan, "Hyperspectral image super-resolution based on spatial and spectral correlation fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 4165–4177, 2018.
[45] Y. Xu, Z. Wu, J. Chanussot, P. Comon, and Z. Wei, "Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 348–362, 2020.
[46] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 3034–3047, 2019.
[47] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4118–4130, 2018.
[48] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2672–2683, 2019.
[49] 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 Transactions on Cybernetics*, vol. 50, no. 10, pp. 4469–4480, 2020.
[50] Y. Zhao, J. Yang, and J. C.-W. Chan, "Hyperspectral imagery super-resolution by spatial-spectral joint nonlocal similarity," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2671–2679, 2014.
[51] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability," *IEEE Transactions on Image Processing*, vol. 29, pp. 116–127, 2020.
[52] 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 Transactions on Image Processing*, vol. 30, pp. 3084–3097, 2021.
[53] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," in *European Conference on Computer Vision*. Springer, 2016, pp. 19–34.
[54] L. Gao, D. Hong, J. Yao, B. Zhang, P. Gamba, and J. Chanussot, "Spectral superresolution of multispectral imagery with joint sparse and low-rank learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2269–2280, 2021.
[55] U. B. Gewali, S. T. Monteiro, and E. Saber, "Spectral super-resolution with optimized bands," *Remote Sensing*, vol. 11, no. 14, p. 1648, 2019.
[56] X. Han, H. Zhang, J.-H. Xue, and W. Sun, "A spectral-spatial jointed spectral super-resolution and its application to HJ-1A satellite images," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2021.
[57] X. Zheng, W. Chen, and X. Lu, "Spectral super-resolution of multispectral images using spatial-spectral residual attention network," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–14, 2021.
[58] T. Li and Y. Gu, "Progressive spatial-spectral joint network for hyperspectral image reconstruction," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–14, 2021.
[59] W. Chen, X. Zheng, and X. Lu, "Semisupervised spectral degradation constrained network for spectral super-resolution," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2021.
[60] S. Mei, R. Jiang, X. Li, and Q. Du, "Spatial and spectral joint super-resolution using convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4590–4603, 2020.
[61] J. He, J. Li, Q. Yuan, H. Shen, and L. Zhang, "Spectral response function-guided deep optimization-driven network for spectral super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
[62] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Spectral super resolution of hyperspectral images via coupled dictionary learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 2777–2797, 2019.
[63] Y. Jia, Y. Zheng, L. Gu, A. Subpa-Asa, A. Lam, Y. Sato, and I. Sato, "From RGB to spectrum for natural scenes via manifold-based mapping," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4715–4723.
[64] H. R. Gordon and M. Wang, "Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm," *Applied Optics*, vol. 33, no. 3, pp. 443–452, 1994.
[65] T. Liang, X. Sun, H. Wang, R. Ti, and C. Shu, "Airborne polarimetric remote sensing for atmospheric correction," *Journal of Sensors*, vol. 2016, pp. 1–7, 2016.
[66] X. Xu, Z. Shi, and B. Pan, "L0-based sparse hyperspectral unmixing using spectral information and a multi-objectives formulation," *ISPRS journal of photogrammetry and remote sensing*, vol. 141, pp. 46–58, 2018.
[67] X. Han, J. Yu, J. Luo, and W. Sun, "Reconstruction from multispectral to hyperspectral image using spectral library-based dictionary learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1325–1335, 2019.
[68] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
[69] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[70] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *ICLR (Poster)*, 2016.
[71] Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu, "HSCNN+: Advanced cnn-based hyperspectral recovery from RGB images," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 1052–10 528.
[72] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
[73] 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 Transactions on Computational Imaging*, vol. 6, pp. 1070–1081, 2020.
[74] Y. Yan, L. Zhang, J. Li, W. Wei, and Y. Zhang, "Accurate spectral super-resolution from single rgb image using multi-scale CNN," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. Springer, 2018, pp. 206–217.
[75] P. Arun, K. M. Buddhiraju, A. Porwal, and J. Chanussot, "Cnn based spectral super-resolution of remote sensing images," *Signal Processing*, vol. 169, p. 107394, 2020.
[76] S. Lei, Z. Shi, and Z. Zou, "Coupled adversarial training for remote sensing image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3633–3643, 2020.
[77] K. Chen, Z. Zou, and Z. Shi, "Building extraction from remote sensing images with sparse token transformers,"

Remote Sensing, vol. 13, no. 21, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/21/4441

[78] D. Yu, H. Duan, J. Fang, and B. Zeng, "Predominant instrument recognition based on deep neural network with auxiliary classification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 852–861, 2020.

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

Liqin Liu received her B.S. degree from Beihang University, Beijing, China in 2018. She is currently working toward her doctorate degree in the Image Processing Center, School of Astronautics, Beihang University. Her research interests include hyperspectral image processing, machine learning and deep learning.

Wenyuan Li received his B.S. degree from HuaDian University, Beijing, China in 2017. He is currently working toward his doctorate degree in the Image Processing Center, School of Astronautics, Beihang University. His research interests include deep learning image processing, and pattern recognition.

Zhenwei Shi (M'13) received his Ph.D. degree in mathematics from Dalian University of Technology, Dalian, China, in 2005. He was a Postdoctoral Researcher in the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He was Visiting Scholar in the Department of Electrical Engineering and Computer Science, Northwestern University, U.S.A., from 2013 to 2014. He is currently a professor and the dean of the Image Processing Center, School of Astronautics, Beihang University. His current research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Dr. Shi serves as an Editor for the Pattern Recognition, the ISPRS Journal of Photogrammetry and Remote Sensing, and the Infrared Physics and Technology, etc. He has authored or co-authored over 200 scientific papers in refereed journals and proceedings, including the IEEE Transactions on Pattern Analysis and Machine Intelligence, the IEEE Transactions on Image Processing, the IEEE Transactions on Geoscience and Remote Sensing, the IEEE Geoscience and Remote Sensing Letters, the IEEE Conference on Computer Vision and Pattern Recognition and the IEEE International Conference on Computer Vision. His personal website is http://levir.buaa.edu.cn/.

Zhengxia Zou received his B.S. degree and his PhD degree from the Image Processing Center, School of Astronautics, Beihang University in 2013 and 2018, respectively. He is currently an Associate Professor at the School of Astronautics, Beihang University. During 2018-2021, he was a postdoc research fellow at the University of Michigan, Ann Arbor. His research interests include computer vision and related problems in remote sensing and autonomous driving. He has published more than 20 peer-reviewed papers in toptier journals and conferences, including TPAMI, TIP, TGRS, CVPR, ICCV, AAAI. His research has been featured in more than 30 global tech media outlets and adopted by multiple application platforms with over 50 million users worldwide. His personal website is https://zhengxiazou.github.io/.