Hyperspectral Remote Sensing Image Synthesis based on Implicit Neural Spectral Mixing Models

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

***Abstract*—Hyperspectral image synthesis, as an emerging research topic, is of great value in overcoming sensor limitations and achieving low-cost acquisition of high-resolution remote sensing hyperspectral images. However, the linear spectral mixing model used in recent studies oversimplifies the real-world hyperspectral imaging process, making it difficult to effectively model the imaging noise and multiple reflections of the object spectrum. As a prerequisite for hyperspectral data synthesis, accurate modeling of nonlinear spectral mixtures has long been a challenge. Considering the above difficulties, we propose a novel method for modeling nonlinear spectral mixtures based on implicit neural representations in this paper. The proposed method learns from implicit neural representation and adaptively implements different mixture models for each pixel according to their spectral signature and surrounding environment. Based on the above neural mixing model, we also propose a new method for hyperspectral image synthesis. Given an RGB image as input, our method can generate an accurate and physically meaningful hyperspectral image. As a set of by-products, our method can also generate sub-pixel-level spectral abundance as well as the solar atmosphere signature. The whole framework is trained end-to-end in a self-supervised manner. We constructed a new dataset for hyperspectral image synthesis based on a wide range of AVIRIS data. Our method achieves an MPSNR of 52.36 dB and outperforms other state-of-the-art hyperspectral synthesis methods. Finally, our method shows great benefits to downstream data-driven applications. With the hyperspectral images and abundance directly generated from low-cost RGB images, the proposed method improves the accuracy of hyperspectral image classification tasks by a large margin, particularly for those with limited training samples.**

***Index Terms*—Remote sensing, hyperspectral image synthesis, implicit neural representation, adaptive spectral mixture model**

I. Introduction

HYPERSPECTRAL remote sensing techniques have unique advantages in many applications such as precision agriculture [1], environment monitoring [2], and mineralogy [3]. In recent years, the rapid development of neural networks and deep learning has brought new ideas to Hyperspectral Image (HSI) processing tasks. However, HSI processing is more challenging compared with other tasks such as cloud detection [4], object detection [5], segmentation [6], etc. Existing deep learning-based remote sensing image processing methods benefit from the utilization of high spatial resolution texture information and heavily rely on a large amount of pixel-by-pixel annotated training data. However, due to the limitations of imaging sensors, the cost of obtaining high spatial resolution data from hyperspectral images is very high, and large-scale annotation relies on field surveys by professionals. Therefore, deep learning methods currently face two main challenges in large-scale hyperspectral image processing applications: limited spatial resolution and scarcity of labeled data.

Recent, hyperspectral image synthesis methods [7–9] have received increasing attention as an emerging research direction in remote sensing field. Hyperspectral image synthesis aims at generating hyperspectral images with both high-spectral and high-spatial resolutions from low-resolution input, which helps alleviate the sensor limitations. According to different input-output configurations, recent hyperspectral synthesis methods are mainly divided into the following three categories: 1) hyperspectral spatial super-resolution with low-resolution HSI (LR-HSI) input [7, 10], 2) spectral super-resolution with high-resolution MSI (HR-MSI) input [8], and 3) image fusion with both LR-HSI and HR-MSI as input [9]. These methods generally follow a data-driven approach, using deep learning networks to learn reconstruction mapping relationships directly from input-output data pairs. Although the reconstruction accuracy has been continuously improved, there is still a long way to go before the practical application of synthetic hyperspectral images. The main reason is that the existing hyperspectral image synthesis methods lack the corresponding physical meaning and ignore the causal factors behind the hyperspectral imaging. The ignored factors include solar illumination, atmospheric absorption, the spectral mixture of ground objects, and sensor quantification. As a result, the realisticity and rationality of synthetic hyperspectral data cannot be guaranteed. Furthermore, the images synthesized by these methods also lack ground truth labels, resulting in significantly lower usability.

As a prelude to this paper, our recently proposed physics-driven hyperspectral image synthesis method PDASS [11] offers the possibility to address the above problems. PDASS considers both hyperspectral imaging mechanisms and linear spectral mixing, and introduces a deep generative network to generate hyperspectral image data based on the standard USGS spectral library [12], which ensures the practical physical

The work was supported by the National Natural Science Foundation of China under Grant 62125102, the National Key Research and Development Program of China (Titled "Brain-inspired General Vision Models and Applications"), and the Fundamental Research Funds for the Central Universities. *(Corresponding author: Zhenwei Shi (e-mail: shizhenwei@buaa.edu.cn))*

Liqin Liu and Zhenwei Shi are with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, and with the Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China, and with the State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing 100191, China, and also with Shanghai Artificial Intelligence Laboratory.

Liqin Liu is also with Shen Yuan Honors College of Beihang University, Beijing 100191, China.

Zhengxia Zou is with the Department of Guidance, Navigation and Control, School of Astronautics, Beihang University, Beijing 100191, China, and also with Shanghai Artificial Intelligence Laboratory.

meanings of the synthetic data. Meanwhile, PDASS can restore the proportion (*abundance*) of each object (*endmember*) in the spectral library in the process of HSI generation. However, the Linear Mixture Model (LMM) used in PDASS oversimplifies the real-world hyperspectral imaging process. On one hand, this assumption makes it difficult to accurately model the imaging noise and multiple reflections of the object spectrum. On the other hand, the nonlinear mixture models (NLMMs) have long been a challenge and are difficult to explicitly model since they heavily depend on the interactions between the environment and multiple types of ground features [13–15].

In this paper, we propose an implicit neural spectral mixing model for modeling nonlinear spectral mixtures. Meanwhile, we propose a novel method for hyperspectral remote sensing image synthesis based on the proposed neural mixing model. We refer to the proposed method as "Implicit Neural Spectral Synthesis (INSS)". Given an RGB image as conditional input, the proposed method first predicts sub-pixel-wise abundance and then synthesizes the hyperspectral data according to the neural spectral mixing model and standard spectral library [12]. For the neural spectral mixing model, the proposed formulation is inspired by the recent advances in Implicit Neural Representation (INR) [16–19]. In INR, an object is usually parameterized by a multilayer perceptron (MLP) that maps coordinates to a signal. Since the coordinates can be any continuous values, the INR is naturally suitable for modeling continuous signals. In our method, since the nonlinear mixing model also has a continuous form in the physical world, therefore we represent the nonlinear mixing function by a multi-layer perceptron (MLP) network which takes in the continuous pixel coordinates, as well as the abundance and the spectral library as its inputs. To achieve adaptive spectral mixing, the parameters of the MLP are determined by the interaction between different ground features in a pixel-by-pixel fashion, i.e., adaptively learns a local nonlinear mixing model for every pixel location. In detail, we design a hyper-network to generate weight parameters of the MLP network. The weight parameters may differ from each other at each pixel location. Furthermore, the mixture model follows the physical process of multiple reflections of the ground object spectrum and the order of the nonlinear mixing is represented by the number of MLP response times. In this way, we can control the order of the mixture model. The hyper-network is trained along with the other network components in an end-to-end manner. As a result, the method not only generates high-resolution hyperspectral data but also acquires the spectral abundance and the pixel-wise nonlinear mixture model.

Compared to the LMM-based spectral synthesis, the proposed method has several advantages. First, the causal factors behind the generated spectral data are recovered more completely, which means not only the proportion of each feature (*abundance*) and the solar atmospheric condition are clear, but also the nonlinear mixture model for each location can be obtained simultaneously. Second, the proposed implicit neural spectral mixing model has a clear physical meaning, each response of the MLP corresponds to a spectral reflection process of the ground objects. Third, the spectral reflection characteristics of the mixing model are pixel-wise determined by the ground objects and the surrounding environment, in line with the actual physical meaning.

To verify our method, we build a new wide-range dataset based on the AVIRIS data. Experiments on the newly collected dataset verify the effectiveness of the proposed method. The proposed method outperforms other state-of-the-art methods [1]. We also verify that the synthetic hyperspectral data produced by our method are of great help for real-world downstream hyperspectral processing tasks. The main contributions of the paper are summarized as follows:

1) We propose an implicit neural spectral mixing model for modeling nonlinear spectral mixtures. We learn from the implicit neural representation and adopt coordinate-driven neural networks to represent the spectral mixture model. The response of the network completely simulates the multiple reflection law of the spectrum of physical objects, and the number of responses corresponds to the order of the mixture model.
2) Based on the neural mixing model, we propose a new hyperspectral image synthesis method. The method can generate accurate and physically meaningful hyperspectral images along with the causal factors including the sub-pixel-level abundance, the solar atmosphere signature and the adaptive mixing model.

The rest of the paper is organized as follows. Section II introduces the related work of HSI reconstruction and implicit neural representation. In section III, details of the implicit neural spectral mixing model and the synthesis method are described. Section IV provides experimental evaluations on the effectiveness of the method and the benefit to downstream tasks. Finally, we draw conclusions in Section V.

II. Related Work

In this section, we give a brief review of hyperspectral image reconstruction methods. Meanwhile, we introduce the implicit neural representation and its application to image generation.

A. Hyperspectral Image Reconstruction

Hyperspectral image reconstruction aims at recovering spatial or spectral information from low-resolution input images. Early hyperspectral reconstruction methods are mainly based on manual features and optimization [7, 20–28]. These methods target at preserving the spectral or spatial information of the input image [22, 25–27] while using regularization constraints such as sparsity [21, 23], low-rank [20, 23], correlation between bands [7] and non-local similarity [24] to reconstruct the missing details. Since the same spatial sparse encoding is shared by the input and the synthesized output, dictionary learning and tensor factorization are widely used in HSI reconstruction [9, 29–41]. Dictionary learning is mainly used to refine the spectra after optimization [9, 29–31] and tensor decomposition methods are usually combined with the non-local similarity of HSIs [35–38]. With the development of convolution neural networks (CNNs), fully-data-driven HSI

[1]The dataset and our code are publically available at http://levir.buaa.edu.cn/Code.htm.

generation methods take advantage of the powerful CNN structure [42–66]. There are mainly four categories of the CNN-based HSI generation methods: attention of the band correlation [42–47], simulation of the degradation process [48–52], improvement of the structure [53–61] and utilization of the imaging prior [62–64]. For example, the MS3 method provides a cluster-based multi-branch backpropagation neural network based on super-pixel segmentation [65].

These methods achieve continuous improvement in the accuracy of HSI generation. However, they still fail to alleviate the problem of limited HSI annotation. In this paper, we propose a hyperspectral data synthesis method based on a nonlinear mixture imaging model, which generates hyperspectral data along with the corresponding label (abundance map).

B. Implicit Neural Representation and its Application on Image Synthesis

Implicit Neural Representation (INR), also known as coordinate-based representations [19], provides a new way to parameterize signals based on neural networks. Different from the traditional rasterized representation, INR realizes a continuous representation of the signal. The input of INR is usually the position in the signal, such as the coordinates of the pixels in an image, and the output is the value of the position, such as the RGB value of the pixel [17]. The mapping between the position and the value of the signal is continuous and cannot be expressed explicitly. INR represents the implicit mapping with neural networks, mainly by multi-layer perceptrons (MLPs) [16]. INR has the following advantages due to the continuity of representation: first, INR is resolution free since the input coordinates can take arbitrary decimals [16]. Second, INR can convert non-differentiable problems into differentiable which can be solved by back-propagation [67]. Finally, the parameters of the neural network can be given by hyper-networks or meta-learning, making the mapping flexible [68].

INR is originally proposed for 3D scene representation and novel view synthesis [16, 19]. To achieve an accurate representation of high-frequency signals, methods such as position encoding [69] and SIREN (SInusoidal REpresentation Networks) [18] have been proposed successively. Thanks to the property of resolution independence, INR is naturally suitable for image generation [70, 71] and super-resolution tasks [17, 72]. Local Implicit Image Function (LIIF) inputs the latent code related to the image content as well as the distance between the current location and the center of the latent code to obtain the RGB value of the position [17]. To solve the shape distortion of high-frequency prediction in LIIF, UltraSR designs residual MLP to replace MLP in LIIF [72]. Image generation methods based on INR often combine with styleGAN [73], generating images from latent code [68, 70, 71]. ALIS (Aligning Latent and Image Spaces) obtains the modulation parameters of the AdaIN (Adaptive Instance Normalization) [74] module of each coordinate through MLP and achieves spatial continuous generation by the movement of the anchor [70]. Similarly, CIPS (Conditionally-Independent Pixel Synthesis) uses MLP to learn modulation vectors while allowing pixel independence via additional position encoding input [71]. Different from ALIS and CIPS, INR-GAN uses a hyper-network to predict the parameters of the MLP, directly adjusting the generator weights [68]. ASAPNet (A Spatially-Adaptive Pixelwise Network) provides a fast image translation method, processing each pixel individually [75]. Moreover, Functa transforms the image data into function space and performs classification tasks in the function space [76].

In our paper, we also use INR to represent a continuous nonlinear mixture model. We represent the mixture model with MLP and learn from INR to set pixel-independent parameters for the model.

III. Proposed Method

In this section, we start with the proposed implicit neural spectral mixing model and then introduce the new method of image synthesis, loss functions, implementation details, etc.

A. Implicit Neural Spectral Mixing Model

In this paper, we follow the earth surface reflection model in [11] and assume the spectrum of the ground object after primary reflection can be expressed as follows:

$$\boldsymbol{y} = \boldsymbol{\varphi} \cdot \boldsymbol{r}, \tag{1}$$

where $\boldsymbol{\varphi}$ represents the solar atmospheric absorption signature and $\boldsymbol{r}$ represents the spectral reflectance of the ground object. The notation $\cdot$ represents element-wise multiplication of the two vectors. If there is no multiple reflection or interaction between distinct endmembers, the energy received by a single pixel from the sensor can be represented by the Linear Mixture Model [77]:

$$\begin{aligned} \boldsymbol{l} &= \boldsymbol{t} \cdot \sum_{i=1}^{N_g} \alpha_i \boldsymbol{r}_i + \boldsymbol{n}, \\ \alpha_i \geq 0;\ \boldsymbol{\alpha}^\top \mathbf{1} &= 1. \end{aligned} \tag{2}$$

In Eq. 2, N_g represent the number of spectra. $\boldsymbol{r}_i$ represents a single spectrum (*endmember*) in the spectral library $\mathcal{R} \in \mathbb{R}^{K \times N_g} = [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{N_g}]$. K represents the band number of the spectra. $\boldsymbol{t} = q\boldsymbol{\varphi}$ denotes the atmospheric absorption factors $\boldsymbol{\varphi}$ with sensor quantification correction q.

In practice, the light travels with multiple interactions among distinct endmembers during the imaging of complex remote sensing scenes. In this case, complex nonlinear mixing is involved, which is difficult to express explicitly. Therefore, We introduce an implicit neural spectral mixing model. In detail, we apply multi-layer perceptron $\mathcal{M}$ to represent the complex response of each order (reflection one time) as shown in Fig. 1. The parameters of $\mathcal{M}$ are pixel-wise adaptive, thus realizing an adaptive spectral mixing model $\mathcal{N}$.

To fully simulate the reflection effect of surrounding objects, we propose the following two design ideas. First, each point-wise mixture model $\mathcal{M}_p$ takes both the pixel's coordinates p and its abundance A_p as input. Second, the neural mixing models are parameterized with spatially varying parameters ϕ_p transformed from the object features. Specifically, each $\mathcal{M}_p$ defines the following mapping:

$$\mathcal{M}_p(A_p, p) = \mathcal{M}_p(A_p, p|\phi_p, \mathcal{R}) =: \boldsymbol{x}_p. \tag{3}$$

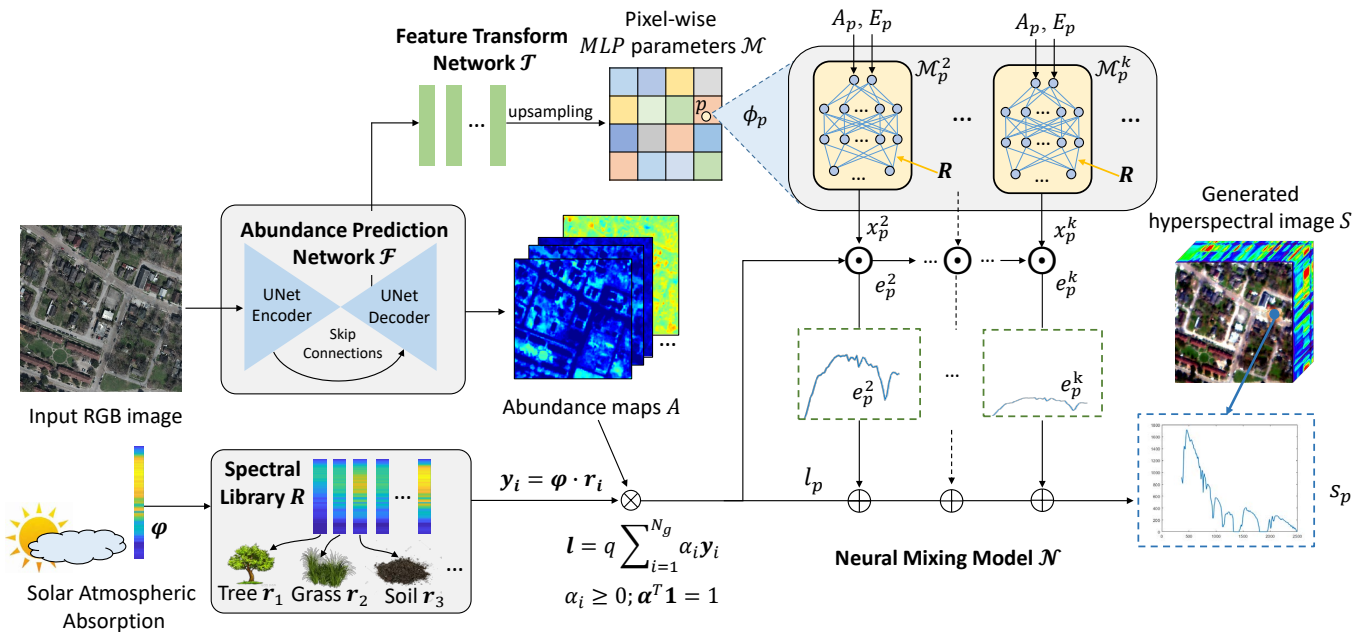


Fig. 1. An overview of the proposed method. Given an input RGB image, we introduce a U-Net based abundance prediction network $\mathcal{F}$ to generate pixel-wise spectral abundance maps. Then the hyperspectral image is synthesized with the abundance map and the standard spectral library following the implicit neural spectral mixing model. The model is pixel-wise adaptive and represented by MLPs whose parameters are from the feature transform network $\mathcal{T}$. All networks can be trained in an end-to-end manner with self-supervised reconstruction losses.

In the above equation, all mixture models ($\mathcal{M}$) share the output layer which takes the spectral library $\mathcal{R}$ as layer parameters. The feature before the output layer represents the respective reflection of N_g ground objects and the output $\boldsymbol{x}_p$ denotes the band response in current order to the previous one.

Here, the output at position p of the kth order $\boldsymbol{e}_p^k$ can be represent with the kth order response $\boldsymbol{x}_p^k$:

$$\begin{aligned} \boldsymbol{e}_p^k &= \boldsymbol{e}_p^{k-1} \cdot \boldsymbol{x}_p^k \\ \boldsymbol{x}_p^k &= \mathcal{M}_p^k(\mathcal{A}_p, p). \end{aligned} \tag{4}$$

$\mathcal{M}_p^k$ represent the mixture model of kth order at location p. In this way, the full representation of the implicit neural spectral mixing model can be written as follows:

$$\begin{aligned} \mathcal{N}_p(A_p, p|\phi_p, \mathcal{R}) &= \boldsymbol{l}_p + \boldsymbol{e}_p^2 + \ldots + \boldsymbol{e}_p^k + \ldots \\ &= \boldsymbol{l}_p + \mathcal{M}_p^2 \cdot \boldsymbol{l}_p + \ldots \\ &= \boldsymbol{l}_p + \sum_{k=2}^{N} \left(\boldsymbol{l}_p \cdot \prod_{i=2}^{k} \mathcal{M}_p^i \right), \end{aligned} \tag{5}$$

where $\prod_{i=2}^{k}$ denotes element-wise multiplication, $\boldsymbol{l}_p = \boldsymbol{e}_p^1$ denotes the linear mixture at p as shown in Eq. 2, and $\mathcal{N}_p$ is the summation of each order output. Under model $\mathcal{N}$, different abundance yield different sets of per-pixel parameters. This in turn means each pixel has an adaptive mixing model varying across space.

B. Generating Pixel-wise Parameters for the Neural Mixing Model

Given the neural spectral mixing model $\mathcal{N}$ and the ground object feature surroundings, we predict ϕ_p based on an abundance prediction network $\mathcal{F}$. Predicting a parameter vector ϕ_p for each pixel independently is prohibitive. Instead, we predict the parameter with a convolutional network $\mathcal{T}$, operating on a much lower resolution feature f_l from the U-Net decoder in $\mathcal{F}$. In practice, the feature transform network $\mathcal{T}$ predicts a grid of parameters at $D\times$ smaller resolution than the RGB image x, which has the same resolution as the input f_l. The grid is then D times upsampled to the same resolution of the image x using nearest interpolation, thus obtaining a parameter vector ϕ_p for the adaptive neural mixing model $\mathcal{N}_p$:

$$\phi_p = [upsample(\mathcal{T}(f_l), D)]_p. \tag{6}$$

The adaptive neural mixing model needs pixel-wise different parameters, while ϕ_p is predicted at a low resolution and then upsampled by nearest interpolation. This limits the ability of the mixing model to synthesize high-frequency details. We avoid this limitation by augmenting $\mathcal{M}_p$ to take an encoding of the coordinate p as additional input [69]. We encode each component of the 2D pixel position $p = (p_x, p_y)$ as a vector of sinusoids with frequencies according to the upsampling factor D. Specifically, in addition to the abundance A_p, each MLP takes the following additional inputs:

$$\begin{aligned} E(p_x) &= \left(\sin\left(2\pi p_x / 2^k\right), \cos\left(2\pi p_x / 2^k\right)\right) \\ E(p_y) &= \left(\sin\left(2\pi p_y / 2^k\right), \cos\left(2\pi p_y / 2^k\right)\right) \\ k &= 1, 2, \ldots, \log_2(D). \end{aligned} \tag{7}$$

Here, (p_x, p_y) denotes p's relative position to the center of ϕ_p, which means the same ϕ_p shares by a local patch similar to [17]. With different positional encode as input, a pixel-wise adaptive neural mixing model is realized.

C. Image Synthesis with the Neural Spectral Mixing Model

To synthesize the hyperspectral images, instead of using the linear mixture model, we start from the neural spectral mixing model (Eq. 5), and consider multiple factors including solar atmospheric absorption, the abundance map, and the standard USGS spectral library. The synthesis process mainly involves the abundance prediction network $\mathcal{F}$, the feature transform network $\mathcal{T}$, the mixing model $\mathcal{N}$ and the spectral library $\mathcal{R}$. An overview of this process is shown in Fig. 1.

Given an RGB image x as conditional input, the abundance prediction network $\mathcal{F}$ recovers the abundance maps A for each object in $\mathcal{R}$ at each pixel:

$$A = \mathcal{F}(x|\boldsymbol{\theta}_f), \tag{8}$$

where $\boldsymbol{\theta}_f$ denotes the trainable parameters in $\mathcal{F}$. $W \times H$ is the spatial size of x. $A \in \mathbb{R}^{N_g \times W \times H}$. Meanwhile, we take the feature map $\boldsymbol{f}_m$ from the decoder of $\mathcal{F}$ as input of the feature transform network $\mathcal{T}$. The parameters of the mixture model $\mathcal{N}$ can be generated as follows:

$$\phi = \mathcal{T}(\boldsymbol{f}_m|\boldsymbol{\theta}_t), \tag{9}$$

where $\boldsymbol{\theta}_t$ denotes the trainable parameters in $\mathcal{T}$. Finally, the spectrum at pixel location p can be synthesized with the spectral library $\mathcal{R}$, the abundance A and the neural mixing model $\mathcal{N}$:

$$\begin{aligned} \boldsymbol{s}_p &= \mathcal{N}_p(A_p, E_p, \boldsymbol{t}|\phi_p, \mathcal{R}) \\ &= \mathcal{N}_p(\mathcal{F}(x|\boldsymbol{\theta}_f)_p, E_p, \boldsymbol{t}|\mathcal{T}(\boldsymbol{f}_m|\boldsymbol{\theta}_t)_p, \mathcal{R}), \end{aligned} \tag{10}$$

where $E_p = (E(p_x), E(p_y))$ represents the position encoding at p. $\boldsymbol{t}$ denotes the solar atmospheric factor with sensor quantification, which is optimized along with the $\boldsymbol{\theta}_f, \boldsymbol{\theta}_t$ during the training process.

D. Loss Functions

The proposed method is trained in a self-supervised learning framework. Given a hyperspectral image S, we compose a spectral down-sampled version $x(S)$ with only R,G,B channels and put it into $\mathcal{F}$ for the reconstructed hyperspectral image S_r. The goal of the training process is to make S_r similar to the original image S. Here we introduce five groups of loss functions:1) pixel similarity, 2) spectral angle mapping, 3) HSV color similarity loss, 4) regularization loss on multiple reflections, and 5) adversarial losses. Let $\boldsymbol{s}_l$ denotes the spectrum in S at location l and $\boldsymbol{s}_{rl}$ denotes the spectrum in S_r at that location.

1) pixel similarity: The pixel similarity loss is defined as the pixel-wise L1 distance between S_r and S. Meanwhile, we find the synthesis errors are different for each band, so we design band-wise weight L1 loss. In practice, we compute the variance of each band and weighted the L1 loss of each band according to the variance:

$$\begin{aligned} \mathcal{L}_{pxl} &= \mathbb{E}_{S\sim\mathcal{D}_S}\{\|S - S_r\|_1 \cdot \boldsymbol{w}\}, \\ &= \mathbb{E}_{I\sim\mathcal{D}_I, l\sim\mathcal{I}_l}\{\|\boldsymbol{s}_l - \boldsymbol{s}_{rl}\|_1 \cdot \boldsymbol{w}\} \end{aligned} \tag{11}$$

where $\mathcal{D}_S$ is the training dataset of hyperspectral images. $\mathcal{I}_l$ is the total pixels in image S. $\boldsymbol{w}$ is the variance vector of each band, which is normalized to $\|\boldsymbol{w}\|_1 = 1$ through $\boldsymbol{w} = \boldsymbol{w}/\|\boldsymbol{w}\|_1$.

2) Spectral angle mapping: The spectral angle mapping loss is defined as the cosine similarity between the real spectrum vector and the synthesis one. The loss is written as follows:

$$\begin{aligned} \mathcal{L}_{\cos} &= \mathbb{E}_{S\sim\mathcal{D}_S}\{\cos < S, S_r >\}, \\ &= \mathbb{E}_{I\sim\mathcal{D}_I, l\sim I_l}\{\cos < \boldsymbol{s}_l, \boldsymbol{s}_{rl} >\}, \end{aligned} \tag{12}$$

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

$$\cos < \boldsymbol{s}_1, \boldsymbol{s}_2 >= \frac{\boldsymbol{s}_1^T \boldsymbol{s}_2}{\sqrt{\|\boldsymbol{s}_1\|_2^2 \|\boldsymbol{s}_2\|_2^2}}. \tag{13}$$

3) HSV color similarity: The HSV color similarity loss is designed to avoid color distortion of the synthesis spectra. The loss constrains the distance between the HSV color space of the images:

$$\mathcal{L}_{hsv} = \mathbb{E}_{S\sim\mathcal{D}_S}\{\|hsv(x(S)) - hsv(x(S_r))\|_1\}, \tag{14}$$

where $hsv(\cdot)$ denotes the transformation from an RGB image to an HSV color image.

4) Regularization loss on multiple reflections: Since the spectrum is mainly from the linear mixing part, we regularize the non-linear part by adding losses on multiple reflections. In practice, we use the L1 norm of the residual image $S_r - L_r$ as loss, where $L_r = \{\boldsymbol{l}_p | p = (p_x, p_y), x \in (0, H), y \in (0, W)\}$ denotes the linear mixture part in S_r.

$$\mathcal{L}_{reg} = \mathbb{E}_{S\sim\mathcal{D}_S}\|S_r - L_r\|_1 \tag{15}$$

5) Adversarial losses: To overcome the ill-posedness of the abundance prediction and mixture model, we follow the conditional adversarial training framework [78, 79]. Once used the adversarial loss, we may alleviate the blur of the synthesis image and improve the visual reality. The joint discriminative learning in our previous work [80] is adopt, the conditional spatial discriminator $\mathcal{D}_{spat}$ and the spectral discriminator $\mathcal{D}_{spec}$ follow the same design in [80]. The adversarial losses are defined as follows:

$$\begin{aligned} \mathcal{L}_{adv}^{spat} &= \mathbb{E}_{S\sim\mathcal{D}} \log \mathcal{D}_{spat}(S) \\ &\quad + \mathbb{E}_{S\sim\mathcal{D}} \log(1 - \mathcal{D}_{spat}(S_r)) \\ \mathcal{L}_{adv}^{spec} &= \mathbb{E}_{S\sim\mathcal{D}} \log \mathcal{D}_{spec}(S) \\ &\quad + \mathbb{E}_{S\sim\mathcal{D}} \log(1 - \mathcal{D}_{spec}(S_r)) \\ \mathcal{L}_{adv} &= \mathcal{L}_{adv}^{spat} + \mathcal{L}_{adv}^{spec}. \end{aligned} \tag{16}$$

The above losses are trained with a minimax optimization process, where the $\mathcal{F}, \mathcal{T}$ try to minimize this objective while the two discriminators try to maximize it:

$$\min_{\mathcal{F},\mathcal{T},\boldsymbol{t}} \max_{\mathcal{D}_{spat},\mathcal{D}_{spec}} \mathcal{L}_{adv}. \tag{17}$$

6) Total loss: The proposed method can be trained in an end-to-end manner since all the components are differentiable. The total loss is defined as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{pxl} + \lambda_1 \mathcal{L}_{\cos} + \lambda_2 \mathcal{L}_{hsv} + \lambda_3 \mathcal{L}_{reg} + \lambda_4 \mathcal{L}_{adv}, \tag{18}$$

where λ_1, λ_2, λ_3, and λ_4 are the pre-defined weights for balancing different loss terms. We set the solar atmospheric

absorption $\boldsymbol{t}$ as all trainable variables. The final loss functions are trained by solving the optimization problem below:

$$\boldsymbol{\theta}_f^*, \boldsymbol{\theta}_t^*, \boldsymbol{t}^* = \arg \min_{\boldsymbol{\theta}_f, \boldsymbol{\theta}_t, \boldsymbol{t}} \max_{\boldsymbol{\theta}_D^{spat}, \boldsymbol{\theta}_D^{spec}} \mathcal{L}_{total}. \tag{19}$$

Although there are no losses or constraints attached to $\boldsymbol{t}$ and the abundance A, they are trained as implicit variables all together with other network parameters.

E. Implementation Details

1) Spectral Library: We use the same spectral library as [11] for spectra synthesis. The spectral data is from the standard USGS Spectral Library [12]. In 2014, the library released its version 7 and we choose the AVIRIS 2014 subset in USGS-v7, which consists of 7 types of objects' spectra, including artificial materials, coatings, liquids, minerals, organic compounds, soils and mixtures, and vegetation. We remove all the spectra from minerals, organic compounds, and chemical reagents of artificial materials. The final spectral library used for experiment consists 345 spectra, i.e., $N_g = 345$.

2) Network Architecture: We design the abundance prediction network $\mathcal{F}$ with a U-Net [81] backbone. We use the residual blocks [82] to replace the convolution layer in the encoder for deep spatial-spectral feature extraction. Meanwhile, skip-connections are adopted to fuse features of different semantic depths with element-wise addition. The backbone network consists of 6+6 residual blocks and the upsampling of the features in the decoder is conducted by bilinear interpolation upsampling followed by a convolution layer, which avoids checkerboard artifacts. The output layer adopts softmax normalization along the channel dimension for the N_g abundance to meet the non-negative and sum to 1 constraint.

The feature transform network $\mathcal{T}$ is designed with 2 convolutional layers without changing the feature resolution. In practice, we choose the feature $2^4 = 16$ times down-sampling from the original resolution in the decoder of $\mathcal{F}$. The first layer is designed with 64 3×3 convolution followed by a reflection padding layer with padsize=1. The second layer is a 1×1 convolution layer. The dimension of the output of $\mathcal{T}$ depends on the number of parameters in the mixture model $\mathcal{N}$.

Each MLP in $\mathcal{N}$ consists two hidden layers, the dimension of each MLP is set to $(N_g + 4 \times \log_2^D, 32, N_g, \mathrm{K})$, where $D = 16$. The parameters of the output layer come from the spectral library $\mathcal{R}$ and others come from the feature transform network $\mathcal{T}$. We choose $0.1 \times tanh$ as the activation function for the output layer and leaky ReLU for others. Especially, we finally choose the order of the neural mixing model as 3, which means $\mathcal{N}$ consists of two MLPs.

3) Training details: During the training phase, we randomly select 256×256 patches from the training images. We set the loss weights $\lambda_1 = 10$, $\lambda_2 = 0.1$, $\lambda_3 = 1$ and $\lambda_4 = 0.01$ and optimize the two discriminators once after each 3 iters of the other networks. The whole framework is trained with Adam optimizer [83] and the cosine learning rate [84] after 20 initial epochs with the max-iteration number 80. The initial learning rate is 10^{-5} for the discriminators, and 10^{-4} for other parameters.

IV. Experiment

A. Datasets and Experimental Setup

We collect a new dataset from the NASA Jet propulsion laboratory (JPL), which is collected by the well-known AVIRIS sensors [2]. The dataset contains 6 scenes of hyperspectral images with a spatial resolution of approximately $3m$, with 224 channels. To eliminate the influence of water absorption bands, we remove the bands [104-114, 152-168] during the training and evaluating processes. The RGB images are conducted from bands 35, 18 and 8 from the hyperspectral images. We choose one scene of image for validation, one scene for testing and the rest for training, with all the images cropped into 512×512 patches (train: 330 patches, validation: 66 patches, test: 76 patches). Moreover, the test patches are divided into 256×256 to meet the GPU memory limitation.

We compared our method with four state-of-the-art hyperspectral synthesis methods, including FMNet [54], R2HGAN [80], HSRNet [57], HSCNN+ [66] and PDASS [11]. FMNet [54] uses multiple branches to learn spatial information of adaptive receptive field, while HSRNet [54] conducts group reconstruction according to the spectral response function (SRF). HSCNN+ [66] designs efficient networks with residual blocks. R2HGAN [80] synthesizes hyperspectral data with joint discriminative learning and PDASS [11] synthesizes HSI with the linear mixture model. All the experiments are conducted on a desktop PC with an Intel (R) Core (TM) i7-7700K CPU @ 4.20GHz and an NVIDIA GeForce GTX 1080 GPU. For a fair competition, all methods are optimized adequately and five criteria are used for evaluation, including RMSE (Root Mean Squared Error) [27, 39], MRAE (Mean Relative Absolute Error) [64], SAM (Spectral Angle Mapper) [53], MSSIM (Mean Structural SIMmilarity) [27, 61] and MPSNR (Mean Peak Signal-to-Noise Ratio) [27, 61]. The MPSNR is the mean of PSNR in each hyperspectral band. The PSNR value of each band is in direct proportion to the negative logarithm of the RMSE of that band. Detailed calculation of the indicators can be found in [80].

B. Comparison with Other Methods

The synthesized images by different methods are shown in Fig. 2. To better illustrate, the false-color images (bands 35, 18, and 8) and the MPSNR score are compared. We can see that HSCNN+ [66] has obvious structural damage and color deviation in building areas. FMNet [54] reconstructs images with undesired color deviation. The same phenomenon happens with PDASS [11], which is caused by the incompleteness of the linear mixture model. The hyperspectral images generated by R2HGAN [80], HSRNet [57] and our method are visually indistinguishable from real images. In addition, our method achieves a higher MPSNR than other methods, which can be also found in Table I.

As shown in Table I, our method achieves the best reconstruction accuracy except for the RMSE metric on the AVIRIS dataset. Our method makes a significant improvement on MRAE from 0.2165 (PDASS [11]) to 0.1913. For RMSE,

[2]The data is collected from https://aviris.jpl.nasa.gov/dataportal.

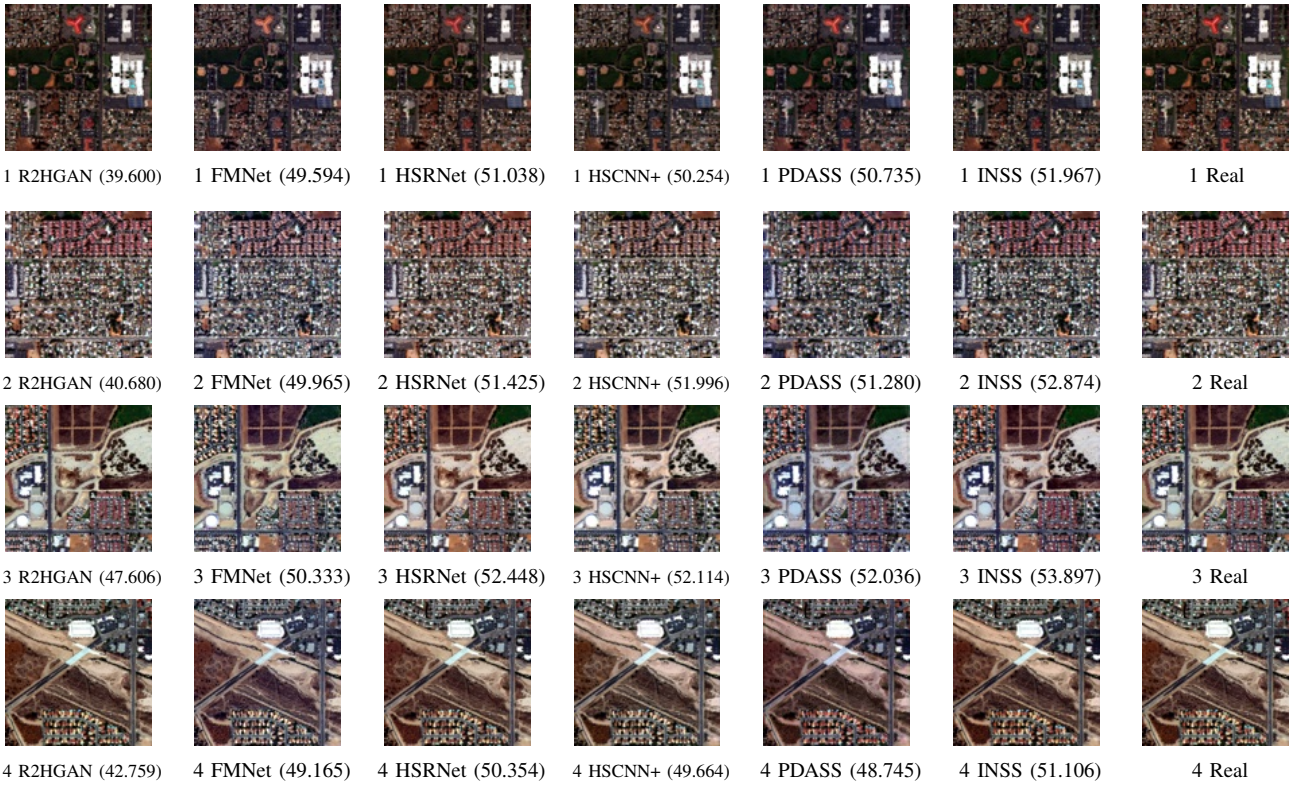
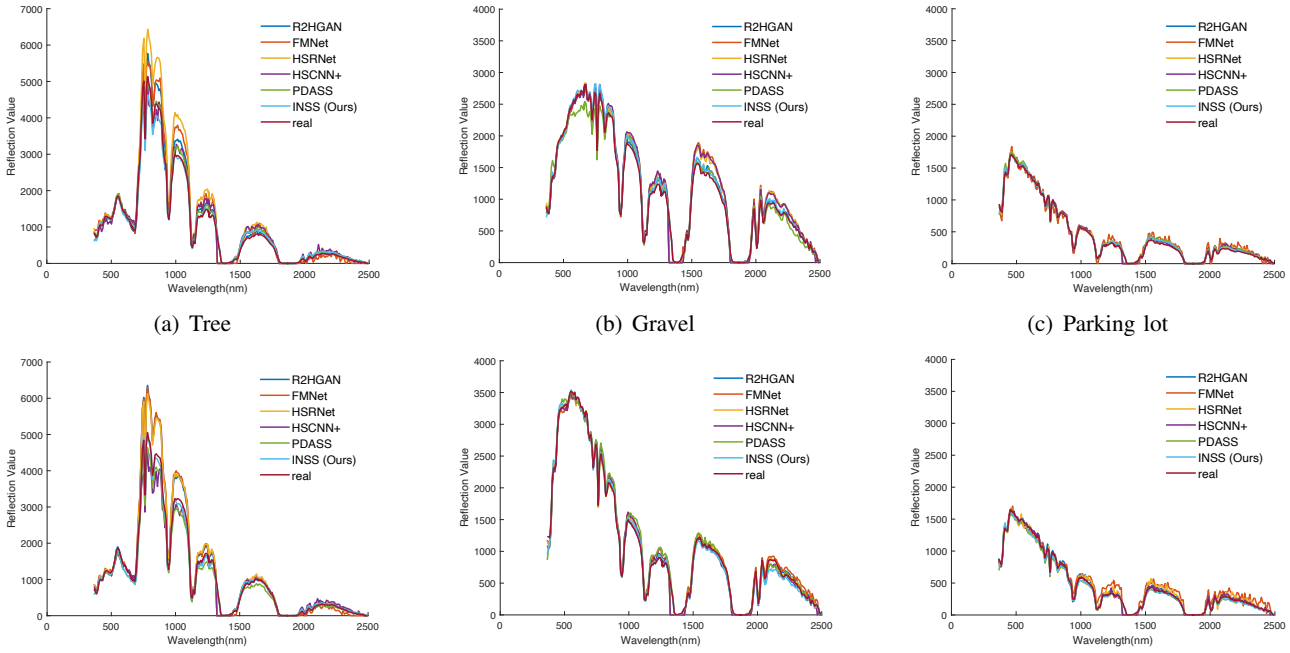


Fig. 2. False-color visualization (band No. 35, 18, and 8) of the synthesis hyperspectral image with different methods: R2HGAN [80], FMNet [54], HSRNet [57], HSCNN+ [66] PDASS [11], and INSS (Ours). The reconstruction MPSNR is given along with the image ID. For example, 1 R2HGAN (39.600) means the result of R2HGAN [80] on test image #1 with MPSNR equals 39.600.



(a) Tree (b) Gravel (c) Parking lot

(d) Grass (e) Building (f) Asphalt road

Fig. 3. Spectral curves on six objects generated by different methods: R2HGAN [80], FMNet [54], HSRNet [57], HSCNN+ [66], PDASS [11], and INSS (Ours).

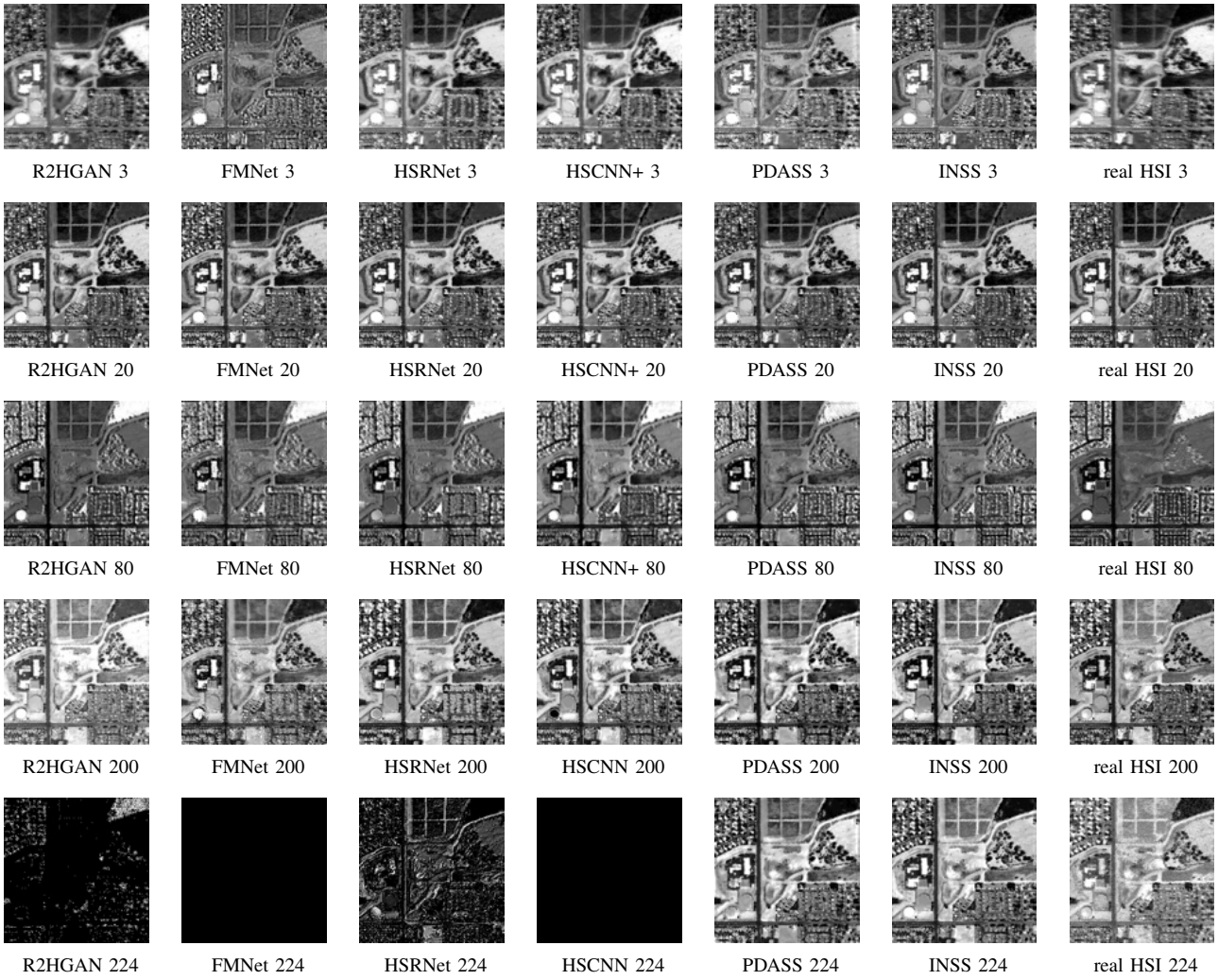


Fig. 4. Band compare of the generated hyperspectral images. Each row contains a particular band generated by different methods: R2HGAN [80], FMNet [54], HSRNet [57], HSCNN+ [66], PDASS [11], and INSS (Ours). Each column shows different bands of one method.

TABLE I
RECONSTRUCTION ACCURACY OF DIFFERENT METHODS ON AVIRIS DATASET. FOR RMSE, MRAE, AND SAM, A LOWER SCORE INDICATES BETTER, WHILE FOR MSSIM AND MPSNR A HIGHER SCORE IS BETTER.

Method	RMSE ↓	MRAE ↓	SAM ↓	MSSIM ↑	MPSNR ↑
R2HGAN	2051.48	14.745	0.2621	0.9612	41.7964
FMNet	320.33	0.3055	0.1533	0.9759	49.7225
HSRNet	**301.31**	0.2327	0.1344	0.9796	51.8475
HSCNN+	317.15	0.2150	0.1404	0.9787	51.3637
PDASS	334.39	0.2165	0.1317	0.9762	50.9376
INSS (Ours)	304.04	**0.1913**	**0.1238**	**0.9801**	**52.3584**

our method ranks second only higher than HSRNet [57], where the MRAE and SAM are much higher than ours.

The comparison on spectral curves and different bands are shown in Figs 3, 4. We can see that FMNet [54], HSRNet [57] and HSCNN+ [66] tend to generate spectral curves which are not as smooth as the real ones. For example, the generated spectra of gravel by the three methods has spectral deviation at wavelength 1500-1800nm and 2000-2500nm as shown in Fig. 3(b). The spectra of grass have the same phenomenon at wavelength 700-900nm (Fig. 3 (d)). The spectral curves of R2HGAN [80] are visually similar to that of FMNet [54], HSRNet [57] and HSCNN+ [66], but there are some abnormal spectra, in which the spectral reflection values at some bands are close to the maximum value. The abnormal values can be found in band 224 in Fig. 4, with unexpected highlight pixels. Therefore, the indicators of R2HGAN [80] perform poorly even with good false-color images and some fine spectra. PDASS [11] generates spectra similar to the real ones except for certain bands of some objects. As shown in Fig. 3 (d), PDASS generates spectra with deviation at wavelength 700-1100nm. INSS compensates for these deviations through the neural mixing model and generates spectra closer to the real ones.

In Fig. 4, five typical bands of the generated hyperspectral images by all methods are shown. It can be found that R2HGAN [80], FMNet [54], HSRNet [57] and HSCNN+ [66] lost spatial information in band 224. Moreover, FMNet [54] loses most spatial information in almost all bands except band 20. PDASS [11] and INSS (Ours) recover more spatial information, especially on some bands where many sensor

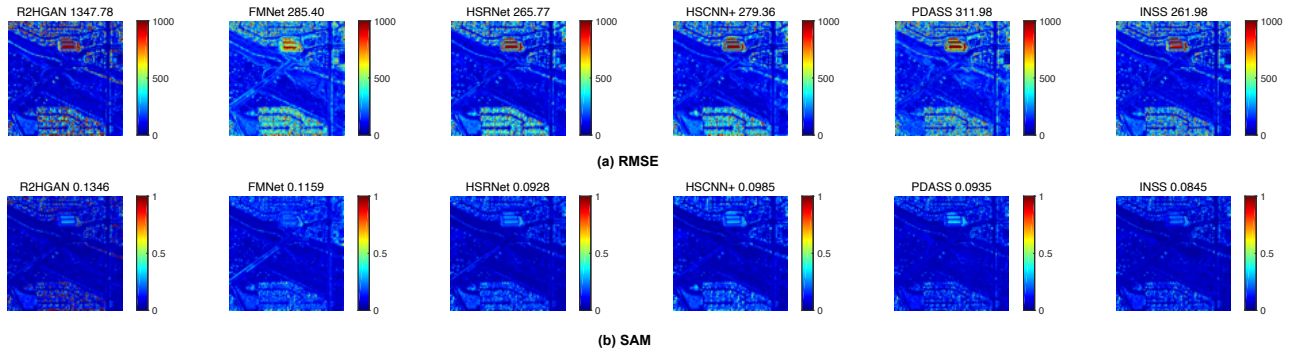


Fig. 5. Error images of different methods in RMSE and SAM metrics, where the lower value denotes the higher reconstruction accuracy. (a) the RMSE visualization, (b) the SAM visualization.

noises exist, such as band 3 and band 224. The rich spatial information is due to the band recovery with the help of the standard spectral library, which avoids the noise in the real image brought by the sensors.

The maps of different methods on RMSE and SAM of the fourth test image in Fig. 2 (test image #4) are shown in Fig. 5. From the RMSE map in Fig. 5 (a), it can be found the value of INSS (Ours) is obviously lower than others, especially in the areas of highlighted and residential buildings. The visualization is consistent with the RMSE values of the image, where INSS (Ours) gets the lowest RMSE of 261.98. From Fig. 5 (b), INSS (Ours) and HSRNet [57] have visually better performance on SAM than other methods. When we carefully compare the highlight building in the upper center, the area of vegetation in the lower left corner and the left road area of the image, we can find INSS has lower SAM values than HSRNet, which is corresponding to the SAM value of the image (0.0845 for INSS and 0.0928 for HSRNet [57]). Therefore, INSS (Ours) has the lowest RMSE and SAM on test image #4, which denotes the highest reconstruction accuracy.

C. Ablation Studies

We conduct ablation studies on different technical components of our method, including the order of the spectral mixing model, the design of the mixing model, the downscale factor of the feature transform network $\mathcal{T}$, whether detach gradient on $\mathcal{T}$, and the loss functions. All the ablation experimental results are shown in Table II.

1) Order of the neural spectral mixing model: In experiments 6-9 of Table II, we study the order of the neural spectral mixing model. We find the reconstruction accuracies of 2-order and 3-order models reach a high level (experiments 7 and 8). The hyperspectral images generated by the linear mixture model (1-order) decrease sharply in accuracy (experiment 6). For example, the MPSNR decreases from 52.2778 to 50.9376. When the order of the model is raised to 4 (experiment 9), the reconstruction accuracy is slightly lower than that of the 3-order model. Therefore, we choose the 3-order model as our final setting.

2) Design of the neural spectral mixing model: In experiments 10-12 of Table II, we study the design of MLPs in the neural mixing model. $N_{h-layer}$ denotes the number of hidden layers and N_{neuron} denotes the number of neurons in the hidden layer. From experiments 8 and 10, we find that increasing the number of hidden layers (from 2 to 3) cannot lead to improved accuracy. Moreover, when we set the hidden layer number as 1, the number of MLP parameters increased largely and the model cannot get reasonable results. Therefore, the hidden layer number is set to 2. For experiments 11 and 12, the number of neurons in the hidden layer is changed to 16 and 64 respectively and the synthesis accuracy is barely affected by it. In conclusion, we chose $N_{h-layer} = 2$ and $N_{neuron} = 32$ for balance of the number of parameters and the reconstruction accuracy.

3) Downscale factor of $\mathcal{T}$: We experiment on different downscale factors of the feature transform network $\mathcal{T}$ by changing its input. Since there is no spatial resolution change in $\mathcal{T}$, the different input feature denotes a different downscale factor D in Eq. 6. As shown in experiments 4 and 8 in Table II, the downscale factor is set to $2^3 = 8$ and $2^4 = 16$, respectively. There is not much difference in accuracy between the two experiments, which indicates the robustness of the downsampling factor to the mixing model.

4) Whether detach on the input of $\mathcal{T}$: We evaluate the effect whether detach gradient of the input of $\mathcal{T}$ in Table II experiments 3 and 5. When the gradient of $\mathcal{T}$ is detached, the reconstruction accuracy has a slight decrease. Therefore, we input the feature $2^4 = 16$ times downscaled to $\mathcal{T}$ without detaching the gradient as summarized from experiments 3, 4, 5, and 8.

5) Loss functions: For the loss functions, we experiment on the weighted L1 loss and the HSV color loss. We compare the pixel-wise L1 loss whether using weights on different bands or not. As shown in experiment 1 of Table II, without weighted L1 loss $\mathcal{L}_{band}$, the focus of the optimization is shifted to errors where the change is not obvious. All the indicators are getting worse especially the MPSNR, which decreases from 52.3584 to 52.095. Therefore, the weighted L1 loss is critical for the reconstruction of the AVIRIS dataset since there are many bands to generate. Meanwhile, we attempt to remove the HSV color loss and find a slight decrement in the accuracy as shown in Table II experiment 2.

In conclusion, the parameter settings have little influence on

TABLE II
ABLATION STUDIES OF DIFFERENT SETTINGS OF THE METHOD, † DENOTES THE FINAL SETTING OF INSS.

Name	Order	$N_{h-layer}$	N_{neuron}	D	Detach	$\mathcal{L}_{band}$	$\mathcal{L}_{hsv}$	RMSE ↓	MRAE ↓	SAM ↓	MSSIM ↑	MPSNR ↑
1						✗		311	0.1963	0.1246	0.9792	52.095
2							✗	309.88	0.193	0.1241	0.9794	52.1829
3				2^3	✓			312.3	0.196	0.1251	0.9791	52.1018
4				2^3	✗			309.41	0.1942	0.1245	0.9795	52.1574
5				2^4	✓			311.57	0.1967	0.1259	0.9791	52.1019
6	1							334.39	0.2165	0.1317	0.9762	50.9376
7	2							306.59	0.1921	0.1237	0.9798	52.2778
8†	3	2	32	2^4	✗			**304.04**	0.1913	**0.1238**	**0.9801**	52.3584
9	4							317.46	0.1994	0.1257	0.9785	52.0093
10		3						309.25	0.1944	0.1243	0.9794	52.0024
11			16					306.24	**0.1865**	0.1242	0.9799	**52.3660**
12			64					306.13	0.1883	0.124	0.9799	52.3231

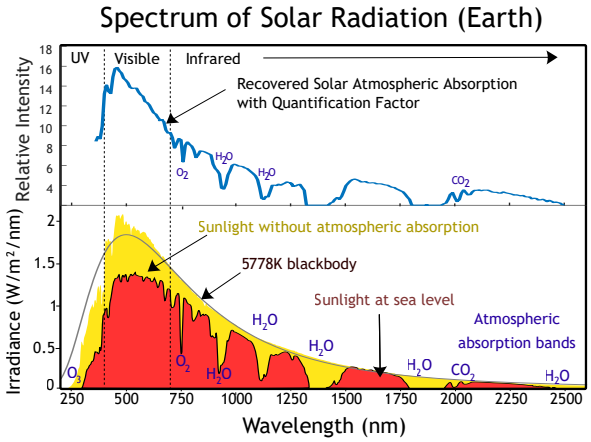


Fig. 6. Estimated solar atmospheric absorption factors and their true values [85].

the reconstruction accuracy except for the order of the model. We choose the experiment settings as experiment 8 in Table II.

D. Analysis of Latent Variables

1) Analysis on the solar atmospheric absorption: The estimated solar atmospheric absorption factors is shown in Fig. 6 as well as its true measurement [85]. The solar atmospheric absorption signature with quantification factors is aligned with its true measurement on wavelength. We can find the recovered and true signatures are consistent. Although there is no groundtruth of the signature and supervised constraints on it, the signature learns the irradiance rises from 380nm to 500nm, and decrements till 2500nm. Meanwhile, we recover the absorption peaks of various substances without omission. As shown in Fig. 6, the absorption peak of oxygen (O_2) at 750nm, the absorption peak of carbon dioxide (CO_2) at around 2000nm and the two peaks of water (H_2O) at 950nm and 1100nm are all visible in the estimated signature.

2) Analysis on abundance maps: The predicted abundance maps are visualized in Fig. 7, demonstrating the recovered abundance is consistent with the true distribution of objects. The abundance of bone black pigment is relatively high in the vegetation areas (shown in Fig. 7(a)) which is mainly because its spectrum has similar absorption peaks as vegetation. Therefore, we can take its abundance as vegetation. The tin roofs in residential areas show a high abundance of iron oxide in Fig. 7(b). Low abundance occurs on natural umber and seawater coast, as these objects are absent from the scene as shown in Fig. 7(c)(d). The abundance of seawater coast is mainly distributed on buildings and roadsides since these areas have gravel similar to the coast. Two buildings have a high abundance of dust debris as shown in Fig. 7(e). From Fig. 7(f), the bare ground has a high-value abundance on soil and mixtures, which suggests the rationality of the abundance map. Meanwhile, the vegetation abundance is high in the areas covered with vegetation, such as the greening of residential, the regular grassland and the vegetation growing on bare ground as shown in Fig. 7(g). Although there are still some predicted abundance maps that have less reasonable distributions, such as the abundance of iron oxide distributes on road, almost all the abundance maps have significant physical meaning.

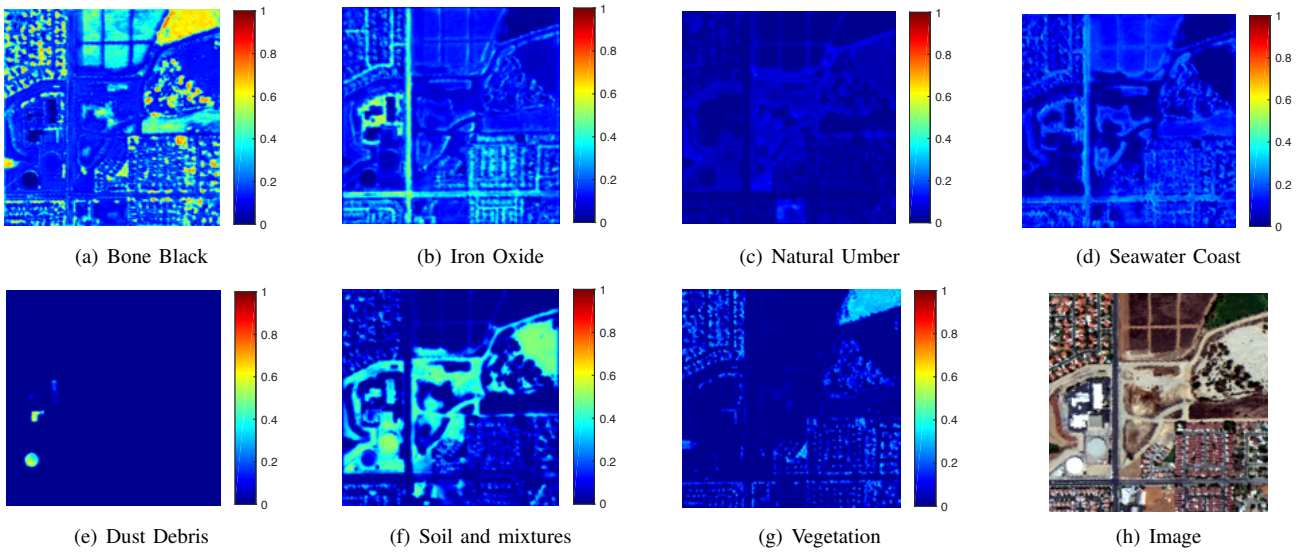
3) Analysis on pixel-wise neural mixing model: For the pixel-wise mixing model, it is determined by the abundance input, the positional encoding and the parameters of the MLP. We pixel-wisely concatenate them and use t-SNE [86] to reduce the dimension and visualize the parameters of each pixel. We plot all the pixels of test image #1 in Fig. 8 (a) and typical pixels manually labeled of five classes in Fig. 8 (b). As shown in Fig. 8, pixels of each class distribute in a different area in the manifold, showing the rationality of the pixel-wise mixing model.

In conclusion, the estimated solar atmospheric absorption, the abundance maps and the pixel-wise mixing model all suggest the rationality and validity of our design.

E. Usage of the Synthesis Data

For the synthesized hyperspectral data and the corresponding abundance map, we design a U-Net based network $\mathcal{A}$ to learn the map from hyperspectral data to its abundance map related to the spectral library:

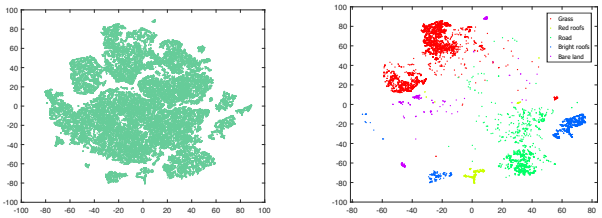
$$A(S) = \mathcal{A}(S|\boldsymbol{\theta}_a), \tag{20}$$



(a) Bone Black (b) Iron Oxide (c) Natural Umber (d) Seawater Coast

(e) Dust Debris (f) Soil and mixtures (g) Vegetation (h) Image

Fig. 7. Visualization of the recovered abundance map on different ground objects: (a) bone black pigment, (b) iron oxide, (c) natural umber pigment, (d) seawater coast, (e) dust debris, (f) soil and mixtures, and (g) vegetation. (h) shows the input image.



(a) All pixels (b) Typical pixels of 5 classes

Fig. 8. t-SNE [86] visualization of the mixing model, pixels of different classes locate in different areas of the manifold where all pixels are distributed in.

where $\boldsymbol{\theta}_a$ denotes the parameters in $\mathcal{A}$. Given a hyperspectral processing task, we can first use $\mathcal{A}$ to obtain the abundance map, which contains much more information about the ground objects and plays an important role in the task. Then we design the downstream method with the input data as well as the abundance $A(S)$, so that the downstream model comes from Eq. 21 to Eq. 22. Since the model $\mathcal{A}$ has learned effective priors from a large number of seen images, the abundance $A(S)$ brings these priors to the downstream tasks and improves the accuracy.

$$y = \mathcal{D}(S|\boldsymbol{\theta}_d) \tag{21}$$

$$y = \mathcal{D}(S, A(S)|\boldsymbol{\theta}_d) \tag{22}$$

We experiment on a downstream classification task to evaluate the effectiveness of the synthesized data and abundance. We use the Salinas dataset, which was collected by the AVIRIS sensor and has a 512×217 spatial size with a resolution of 3.7m/pixel. It includes 16 classes with a variety of vegetation and soils. In Table III, we show the overall accuracy (OA) of classification with (Eq. 22) or without (Eq. 21) the abundance obtained by Eq. 20.

TABLE III
THE OVERALL CLASSIFICATION ACCURACY ON THE SALINAS DATASET. THE ABUNDANCE OBTAINED BY OUR METHOD IMPROVES THE ACCURACY.

Classification Method	Samples	OA w/o abundance (%)	OA w abundance (%)	Improve (%)
2D-CNN (Our design)	50	85.55	86.76	1.21
	100	88.34	89.12	0.78
	200	90.34	91.09	0.75
3D-CNN [87]	50	69.103	89.775	20.672
	100	88.981	92.509	3.528
	200	88.729	96.417	7.688
3D-FCN [88]	50	83.827	88.001	4.174
	100	90.15	92.393	2.243
	200	89.212	94.846	5.634
Multi-scale 3D-CNN [89]	50	90.349	94.451	4.102
	100	91.079	95.583	4.504
	200	93.54	96.391	2.851

To fully verify the improvement in classification accuracy, we experiment with four classification methods as shown in Table III and chose the number of training samples in each class as 50, 100, and 200. Note we design a 2D-CNN based network $\mathcal{D}$ with six 3×3 convolutional layers with 9×9 patches input. The three other methods are based on 3D-CNN [87], 3D-FCN (Fully Convolutional Network) [88] and Multi-scale 3D-CNN [89], respectively. We use the implementation of the three methods in the DeepHyperX toolbox [90]. We can find the abundance significantly improves the accuracy regardless of the classification method or the sample number.

Meanwhile, for methods 3D-CNN and 3D-FCN [87, 88], without abundance input, when we add the samples from 100 to 200 per class, the classification accuracy even decreases, which denotes that increase of training samples cannot bring accuracy improvement. When we use the synthesis data to help classification, the sample increment from 100 to 200 still brings an accuracy improvement since additional valid information is introduced.

As a result, the synthesis data and abundance map help

improve the downstream tasks and have the potential to reduce sample labeling requirements.

V. Conclusion

Physics-informed hyperspectral image synthesis can generate hyperspectral images along with the corresponding abundance map according to the remote sensing imaging model. In this paper, we propose a hyperspectral image synthesis method based on an implicit neural spectral mixing model. We inspire from the implicit neural representation and design the pixel-wise adaptive mixing model to compensate for the incompleteness of the linear mixture model. We predict the sub-pixel-level abundance of ground objects, estimate the pixel-wise mixing model, and finally synthesize the hyperspectral image. The image is synthesized with the predicted abundance, the solar atmospheric absorption factors and the spectral library according to the neural mixing model. Meanwhile, the experiments suggest the superiority and validity of our method. First, on the new-collected AVIRIS dataset, our method achieves the best reconstruction accuracy with an MPSNR as high as 52.32, which outperforms previous state-of-the-art methods. Second, the visualization of the two implicit variables (the abundance map and the solar atmospheric absorption factors) demonstrate clear physical meanings and consistency to the true measurements. Third, the improvement in the downstream work shows the potential of the synthesis data and the abundance. Finally, an extensive ablation study verifies the robustness of our method. In the future, we will explore a wider range of real-world downstream applications utilizing the synthesized high-quality HSI and its corresponding abundance.

VI. Acknowledgement

The authors would like to thank the NASA Jet Propulsion Laboratory (JPL) for providing the AVIRIS data used in this study. Meanwhile, we would thank the United States Geological Survey (USGS) for the spectral library used in our experiment.

References

[1] 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.

[2] 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.

[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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2426–2442, 2014.

[4] X. Wu, Z. Shi, and Z. Zou, "A geographic information-driven method and a new large scale dataset for remote sensing cloud/snow detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 174, pp. 87–104, 2021.

[5] Z. Zou and Z. Shi, "Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1100–1111, 2017.

[6] K. Chen, Z. Zou, and Z. Shi, "Building extraction from remote sensing images with sparse token transformers," *Remote Sensing*, vol. 13, no. 21, p. 4441, 2021.

[7] 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.

[8] 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.

[9] 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 Transactions on Image Processing*, vol. 25, no. 1, pp. 274–288, 2016.

[10] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Asian conference on computer vision*. Springer, 2014, pp. 111–126.

[11] L. Liu, W. Li, Z. Shi, and Z. Zou, "Physics-informed hyperspectral remote sensing image synthesis with deep conditional generative adversarial networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[12] 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.

[13] J. M. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *Image and Signal Processing for Remote Sensing XV*, vol. 7477. International Society for Optics and Photonics, 2009, p. 74770I.

[14] C. C. Borel and S. A. Gerstl, "Nonlinear spectral mixing models for vegetative and soil surfaces," *Remote sensing of environment*, vol. 47, no. 3, pp. 403–416, 1994.

[15] Y. Itoh, S. Feng, M. F. Duarte, and M. Parente, "Semisupervised endmember identification in nonlinear spectral mixtures via semantic representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3272–3286, 2017.

[16] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *European conference on computer vision*. Springer, 2020, pp. 405–421.

[17] Y. Chen, S. Liu, and X. Wang, "Learning continuous image representation with local implicit image function," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8628–8638.

[18] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7462–7473, 2020.

[19] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 165–174.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] X. Sun, L. Zhang, H. Yang, T. Wu, Y. Cen, and Y. Guo, "Enhancement of spectral resolution for remotely sensed multispectral image," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 5, pp. 2198–2211, 2015.

[27] 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.

[28] 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.

[29] 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.

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

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] W. He, N. Yokoya, and X. Yuan, "Fast hyperspectral image recovery of dual-camera compressive hyperspectral imaging via non-iterative subspace-based fusion," *IEEE Transactions on Image Processing*, vol. 30, pp. 7170–7183, 2021.

[41] J. Aeschbacher, J. Wu, and R. Timofte, "In defense of shallow learned spectral reconstruction from rgb images," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 471–479.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.

[57] 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.

[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] 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.

[60] 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.

[61] 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.

[62] 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.

[63] 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.

[64] 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.

[65] 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*, vol. 19, pp. 1–5, 2021.

[66] Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu, "Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 939–947.

[67] F. Cole, K. Genova, A. Sud, D. Vlasic, and Z. Zhang, "Differentiable surface rendering via non-differentiable sampling," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 6088–6097.

[68] I. Skorokhodov, S. Ignatyev, and M. Elhoseiny, "Adversarial generation of continuous images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 753–10 764.

[69] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7537–7547, 2020.

[70] I. Skorokhodov, G. Sotnikov, and M. Elhoseiny, "Aligning latent and image spaces to connect the unconnectable," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 14 144–14 153.

[71] I. Anokhin, K. Demochkin, T. Khakhulin, G. Sterkin, V. Lempitsky, and D. Korzhenkov, "Image generators with conditionally-independent pixel synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 278–14 287.

[72] X. Xu, Z. Wang, and H. Shi, "Ultrasr: Spatial encoding is a missing key for implicit image function-based arbitrary-scale super-resolution," *arXiv preprint arXiv:2103.12716*, 2021.

[73] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8110–8119.

[74] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1501–1510.

[75] T. R. Shaham, M. Gharbi, R. Zhang, E. Shechtman, and T. Michaeli, "Spatially-adaptive pixelwise networks for fast image translation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 882–14 891.

[76] E. Dupont, H. Kim, S. Eslami, D. Rezende, and D. Rosenbaum, "From data to functa: Your data point is a function and you should treat it like one," *arXiv preprint arXiv:2201.12204*, 2022.

[77] 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.

[78] 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.

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

[80] 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.

[81] 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.

[82] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[83] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[84] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *ICLR (Poster)*, 2016.

[85] 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.

[86] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[87] A. B. Hamida, A. Benoit, P. Lambert, and C. B. Amar, "3-d deep learning approach for remote sensing image classification," *IEEE Transactions on geoscience and remote sensing*, vol. 56, no. 8, pp. 4420–4434, 2018.

[88] H. Lee and H. Kwon, "Contextual deep cnn based hyperspectral classification," in *2016 IEEE international geoscience and remote sensing symposium (IGARSS)*. IEEE, 2016, pp. 3322–3325.

[89] M. He, B. Li, and H. Chen, "Multi-scale 3d deep convolutional neural network for hyperspectral image classification," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 3904–3908.

[90] N. Audebert, B. Le Saux, and S. Lefèvre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE geoscience and remote sensing magazine*, vol. 7, no. 2, pp. 159–173, 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.

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/.

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

He was a Post-Doctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He was 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 Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing. He has authored or coauthored over 200 scientific articles 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 (CVPR) and the IEEE International Conference on Computer Vision (ICCV). 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. His personal website is http://levir.buaa.edu.cn/.