

Unmixing Guided Unsupervised Network for RGB Spectral Super-Resolution

Qiaoying Qu^{ID}, Bin Pan^{ID}, *Member, IEEE*, Xia Xu^{ID}, *Graduate Student Member, IEEE*,
Tao Li^{ID}, and Zhenwei Shi^{ID}, *Member, IEEE*

Abstract—Spectral super-resolution has attracted research attention recently, which aims to generate hyperspectral images from RGB images. However, most of the existing spectral super-resolution algorithms work in a supervised manner, requiring pairwise data for training, which is difficult to obtain. In this paper, we propose an Unmixing Guided Unsupervised Network (UnGUN), which does not require pairwise imagery to achieve unsupervised spectral super-resolution. In addition, UnGUN utilizes arbitrary other hyperspectral imagery as the guidance image to guide the reconstruction of spectral information. The UnGUN mainly includes three branches: two unmixing branches and a reconstruction branch. Hyperspectral unmixing branch and RGB unmixing branch decompose the guidance and RGB images into corresponding endmembers and abundances respectively, from which the spectral and spatial priors are extracted. Meanwhile, the reconstruction branch integrates the above spectral-spatial priors to generate a coarse hyperspectral image and then refined it. Besides, we design a discriminator to ensure that the distribution of generated image is close to the guidance hyperspectral imagery, so that the reconstructed image follows the characteristics of a real hyperspectral image. The major contribution is that we develop an unsupervised framework based on spectral unmixing, which realizes spectral super-resolution without paired hyperspectral-RGB images. Experiments demonstrate the superiority of UnGUN when compared with some SOTA methods.

Index Terms—Unsupervised spectral super-resolution, spectral unmixing, adversarial learning.

I. INTRODUCTION

HYPERSPECTRAL image (HSI) contains abundant spectral information, thus it has been applied in quite a few fields, such as scene classification [1], [2], [3], target

detection [4], [5] and segmentation [6]. However, due to the limitations of hardwares, it is hard to obtain HSI with high resolution in spatial domain, which leads to a new research direction: hyperspectral super-resolution. Hyperspectral super-resolution is designed to generate high resolution HSIs from low resolution HSIs or high resolution RGB images. Based on the inputs, the recent methods can be divided into three categories: spatial super-resolution of HSI, spectral super-resolution of RGB image, and fusion-based super-resolution.

Spatial super-resolution of HSI aims at obtaining high resolution HSI by reconstructing spatial information from low resolution HSI. Most existing methods are based on spectral basis functions [7], sparse dictionary representation learning [8], maximum a posteriori (MAP) [9], [10] and deep learning [11], [12], [13], [14] etc. Spatial-spectral prior and consistency regularization are common research focuses [15], [16], [17]. For example, Wang et al. [17] designed a regularization network to enforce spatial-spectral correlations. Moreover, many manually designed priors are also utilized to guide the recovery of spatial information [18].

Spectral super-resolution algorithms reconstruct spectral channels from high resolution RGB image to generate high resolution HSI. Based on the degradation formula, many traditional methods explore the map between RGB image and HSI [19], [20], [21]. For example, Parmar et al. [22] exploit the sparsity of image to jointly reconstruct the spatial-spectral characteristics. Arad et al. [23] created a sparse dictionary of HSI and utilized hyperspectral prior to obtain the RGB projections. In recent years, deep learning methods, especially CNN, have shown promising performance in spectral super-resolution tasks [24], [25], [26]. For example, spectral response function is utilized to group the bands with spectral relevance [31]. Networks based on implicit neural representations are proposed to model the imaging noise and multiple reflections of the object spectrum. [32]. Transformer are applied to spectral super-resolution to capture the self-similarity prior and long-range dependencies [33], [34]. Moreover, some researchers incorporate priors into neural networks to avoid designing parameters manually [27], [28], [29], [30].

Fusion based hyperspectral super-resolution methods have also been widely utilized. These algorithms refer to generating high resolution HSI with low resolution HSI and registered high resolution multispectral

Manuscript received 24 August 2022; revised 2 May 2023; accepted 14 July 2023. Date of publication 1 August 2023; date of current version 1 September 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2022ZD0160401 and Grant 2022YFA1003803; in part by the National Natural Science Foundation of China under Grant 62001251, Grant 62001252, Grant 62125102, and Grant 62272248; and in part by the Beijing–Tianjin–Hebei Basic Research Cooperation Project under Grant F2021203109. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Leyuan Fang. (Corresponding author: Xia Xu.)

Qiaoying Qu and Bin Pan are with the School of Statistics and Data Science, KLMDASR, LEBPS, and LPMC, Nankai University, Tianjin 300071, China (e-mail: quqiaoying@mail.nankai.edu.cn; panbin@nankai.edu.cn).

Xia Xu and Tao Li are with the College of Computer Science, Nankai University, Tianjin 300071, China (e-mail: xuxia@nankai.edu.cn; litao@nankai.edu.cn).

Zhenwei Shi is with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: shizhenwei@buaa.edu.cn).

Digital Object Identifier 10.1109/TIP.2023.3299197

1941-0042 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See <https://www.ieee.org/publications/rights/index.html> for more information.

image (MSI)/RGB image. Traditional methods based on matrix factorization [35], [36], [37], [38], tensor factorization [39], [40], [41], [42], [43] and bayesian [44] are proposed firstly. Then deep learning based methods have become the mainstream [45], [46], [47], [48], [49], [50]. Many model-guided network [51], [52], [53] are proposed to optimize the hand-crafted prior in an end-to-end manner. Recently, unsupervised algorithms are proposed to solve the problem of labels acquisition [54], [55], [56]. However, paired hyperspectral-RGB images are usually not available, thus the fusion based super-resolution is difficult to implement.

Although many methods have been proposed, one of the challenges for spectral super-resolution is that paired high and low-resolution images are required during the training process, which makes the hyperspectral super-resolution task work in a supervised manner. However, paired HS-RGB data is non-trivial to obtain, which greatly increases the difficulty of supervised network training. Therefore, many researchers attempt to develop unsupervised super-resolution networks using unpaired data [57], [58]. However, most of the unsupervised spectral super-resolution algorithms do not consider the spectral characteristics of hyperspectral image, which is, each pixel is formed by a linear combination of several endmembers.

In this paper, based on the characteristics of hyperspectral images, we propose an Unmixing Guided Unsupervised Network (UnGUN) for hyperspectral super-resolution, which realizes spectral super-resolution with the help of guidance hyperspectral image and spectral response function. Spectral unmixing refers to the restoration of the spectra of pure substances in the image (endmember) and the proportion of corresponding substances in each pixel (abundance) [59]. In many application fields of hyperspectral images, for example, remote sensing, the distributions of materials corresponding to different scenes are usually similar. Thus their endmembers are often similar, which indicates that the endmembers of other scenes can help reconstruct the endmembers of the super-resolution images. Even if the ground objects of the existing image and super-resolution image are not corresponding, the overall spectral distribution information contained in the guidance image can still provide prior for the super-resolution image. Moreover, motivated by the fact that spectral unmixing can separate and extract the spatial-spectral information of input completely, which is helpful for the subsequent feature fusion, we develop a spectral super-resolution framework based on spectral unmixing with guidance images. Specifically, UnGUN has two inputs: RGB images that require spectral super-resolution and guidance hyperspectral images. It is worth noting that the RGB image and the guidance image do not need to be paired. In addition, the spectral response function is utilized to constrain the relationship between the reconstructed hyperspectral image and the input RGB image to ensure the super-resolution effect.

The proposed UnGUN is optimized in a generative adversarial manner, where the generator mainly consists of three autoencoder branches: two unmixing branches and one reconstruction branch. The unmixing branches realize the decomposition of guidance hyperspectral image and RGB image

respectively. After spectral unmixing, we obtain their corresponding endmembers and abundances, which contain the spectral and spatial information. Furthermore, the reconstruction branch utilizes the abundance from RGB unmixing branch and the endmember from hyperspectral unmixing branch to initialize the coarse hyperspectral image, which is then refined to obtain the final hyperspectral image. Therefore, the spectral feature from hyperspectral image and spatial information from RGB image are integrated through the reconstruction branch. Besides, we design a discriminator to guarantee that the distribution of reconstructed image is similar to the real guidance hyperspectral data. In general, the guidance image has two effects: guiding the initialization of hyperspectral image and improving the generation ability of the generator.

The major contributions of our algorithm can be summarized as follows:

- We propose a new unsupervised spectral super-resolution framework, which attempts to tackle the problem of obtaining pairwise data based on the similarity of material distribution in hyperspectral images.
- We design two spectral unmixing branches and a reconstruction branch, which implement the extraction and fusion of spatial-spectral priors to reconstruct super-resolved hyperspectral images.

II. METHODOLOGY

A. Background

1) *Unmixing*: HSI suffers from low resolution in the spatial domain, thus each pixel of an HSI contains more than one different substance, which is a general challenge in HSI, mixed pixels. Hyperspectral unmixing aims at decomposing each pixel into a set of pure spectra (i.e., endmembers) and corresponding proportions (i.e., abundance). One of the common mixing models for hyperspectral unmixing is LMM (linear mixing model).

Note that $X \in \mathcal{R}^{HW \times C}$ is the hyperspectral image, which contains C channels in the spectral domain and HW pixels in the spatial domain. Let $M \in \mathcal{R}^{p \times C}$ be the endmember matrix with p pure spectra, $A = (a_{i,j})_{HW \times p} \in \mathcal{R}^{HW \times p}$ represents the abundance matrix. Based on the LMM, the unmixing model can be expressed as:

$$X = AM + N \quad (1)$$

where N is an additive white Gaussian noise matrix.

Moreover, the abundance map needs to satisfy abundance non-negative constraint (ANS) as Eq.2 and abundance sum-to-one constraint (ASC) as Eq.3.

$$a_{i,j} \geq 0 \quad \forall i, \forall j \quad (2)$$

$$\sum_{j=1}^p a_{i,j} = 1 \quad \forall i \quad (3)$$

2) *Super-Resolution*: Most super-resolution algorithms can be divided into three categories: spatial super-resolution, spectral super-resolution and fusion-based super-resolution.

Hyperspectral spatial super-resolution attempts to recover HR-HSI from LR-HSI. Given the HR-HSI $X \in \mathcal{R}^{HW \times C}$

and the blurring kernel G , the LR-HSI $Y \in \mathcal{R}^{hw \times C}$ can be obtained by the following formula:

$$Y = (G * X) \downarrow + n \quad (4)$$

where \downarrow represents downsampling operation, n is additive white Gaussian noise. Based on the degradation process, the formula of spatial super-resolution is as follows:

$$\hat{X} = f(X; \phi) \quad (5)$$

where f is the super-resolution model and ϕ is the parameter.

For spectral super-resolution, given the HR-HSI $X \in \mathcal{R}^{HW \times C}$ and the spectral response function R , the RGB image $Z \in \mathcal{R}^{HW \times 3}$ can be obtained by the following formula:

$$Z = X \cdot R + n \quad (6)$$

where n is additive white Gaussian noise. Thus the spectral super-resolution process is as follows:

$$\hat{X} = g(Z; \theta) \quad (7)$$

where g is the super-resolution model and θ is the parameter.

For fusion-based super-resolution, given the HR-HSI $X \in \mathcal{R}^{HW \times C}$, the spectral response function R and the blurring kernel G , the RGB image $Z \in \mathcal{R}^{HW \times 3}$ and LR-HSI $Y \in \mathcal{R}^{hw \times C}$ can be obtained by the following formula:

$$Z = X \cdot R + n \quad (8)$$

$$Y = (G * X) \downarrow + n \quad (9)$$

Thus the fusion-based super-resolution process is as follows:

$$\hat{X} = h(Z, Y; \psi) \quad (10)$$

where h is the super-resolution model and ψ is the parameter.

B. Overall Architecture

Some of the terms are denoted as follows: In our model, the inputs are a RGB image $\mathcal{X} \in \mathcal{R}^{H \times W \times 3}$ and a guidance hyperspectral image $\mathcal{Y} \in \mathcal{R}^{h \times w \times C}$. H and W are the height and width of \mathcal{X} , while h , w , and $C \gg 3$ are height, width, and the number of bands of \mathcal{Y} respectively. Our goal is to reconstruct spectra of \mathcal{X} to obtain an image $\mathcal{Z} \in \mathcal{R}^{H \times W \times C}$ with high spectral resolution. For the convenience of subsequent operations, we expand the 3D tensors into 2D matrices, let $X \in \mathcal{R}^{HW \times 3}$ and $Y \in \mathcal{R}^{hw \times C}$ be input RGB image and guidance HSI respectively. $Z \in \mathcal{R}^{HW \times C}$ denotes reconstructed HR-HSI.

Most spectral super-resolution algorithms are supervised with ground truth. However, in the real world, labels with high spectral resolution are often not available, which leads to the failure of the algorithms. Therefore, many unsupervised algorithms have been proposed recently. Compared with the supervised algorithm, the image reconstructed by the unsupervised algorithm often has color deviation and texture distortion due to the lack of real data during training. To tackle this problem, we propose an unsupervised reconstruction model, which is guided by a guidance HSI.

Our algorithm is inspired by the spectral unmixing. Thus the UnGUN mainly consists of three branches, including two unmixing branches and a reconstruction branch. The overall architecture is represented in Fig.1.

We first perform spectral unmixing for RGB images and guidance HSI respectively to obtain their abundances and endmembers, where abundances contain spatial information, and endmembers possess spectral information.

The unmixing of RGB image and guidance HSI is realized with the same structure:

$$X = A_X \cdot M_X \quad (11)$$

$$Y = A_Y \cdot M_Y \quad (12)$$

where $M_X \in \mathcal{R}^{L \times 3}$ and $M_Y \in \mathcal{R}^{L \times C}$ are endmembers of RGB image and guidance hyperspectral image, $A_X \in \mathcal{R}^{HW \times L}$ and $A_Y \in \mathcal{R}^{hw \times L}$ are abundances respectively. L represents the number of spectra endmembers.

For spectral reconstruction, we utilize the abundance of RGB images and the endmember of guidance HSI to synthesize a coarse HSI \hat{Z} as Eq. 13. Based on the similarity of endmembers corresponding to different hyperspectral datasets, we attempt to initialize spectral information with the guidance of M_Y . Since the M_Y contains information of real spectrum distributions, it can guide the reconstruction branch to generate spectra that conform to real-world distribution, even if the materials that do not exist in guidance image. In this way, with the help of hyperspectral unmixing branch, the reconstruction branch can optimize the endmember matrix contained in the RGB image, which ensures the accuracy of spectral characteristics of the reconstructed hyperspectral image.

$$\hat{Z} = A_X \cdot M_Y \quad (13)$$

Finally, after the adjustment module, the coarse HSI \hat{Z} is refined into the final HSI Z as Eq. 14.

$$Z = Adj(\hat{Z}) \quad (14)$$

Moreover, in order to further ensure the authenticity of the reconstructed image Z , we design a discriminator to identify Z and the guidance image, finally making sure the reconstructed image is close to the real-world hyperspectral data. Specifically, when there are ground objects in RGB image that do not exist in the guidance image, the subsequent adjustment module will generate a more reasonable endmember distribution under the constraint of discrimination loss and reconstruction loss.

C. Unmixing Branches

The three branches share the same autoencoder unmixing structure as Fig.2. After encoder, the input image is decomposed into abundance. Since the last layer of the encoder is Softmax, the generated abundance satisfies ANC and ASC naturally. And then the decoder acts as the endmember of the input image with latent abundance.

1) *HSI Unmixing Branch*: The HSI unmixing branch can be divided into two parts: Encoder1: $En_1: \mathcal{R}^{hw \times C} \rightarrow \mathcal{R}^{hw \times L}$, which maps the guidance HSI to the latent representation A_Y , and Decoder1: $De_1: \mathcal{R}^{hw \times L} \rightarrow \mathcal{R}^{hw \times C}$, which reconstructs the input HSI with A_Y as the endmember M_Y . The HSI unmixing branch can be expressed as:

$$A_Y = En_1(Y) \quad (15)$$

$$\hat{Y} = De_1(A_Y) \quad (16)$$

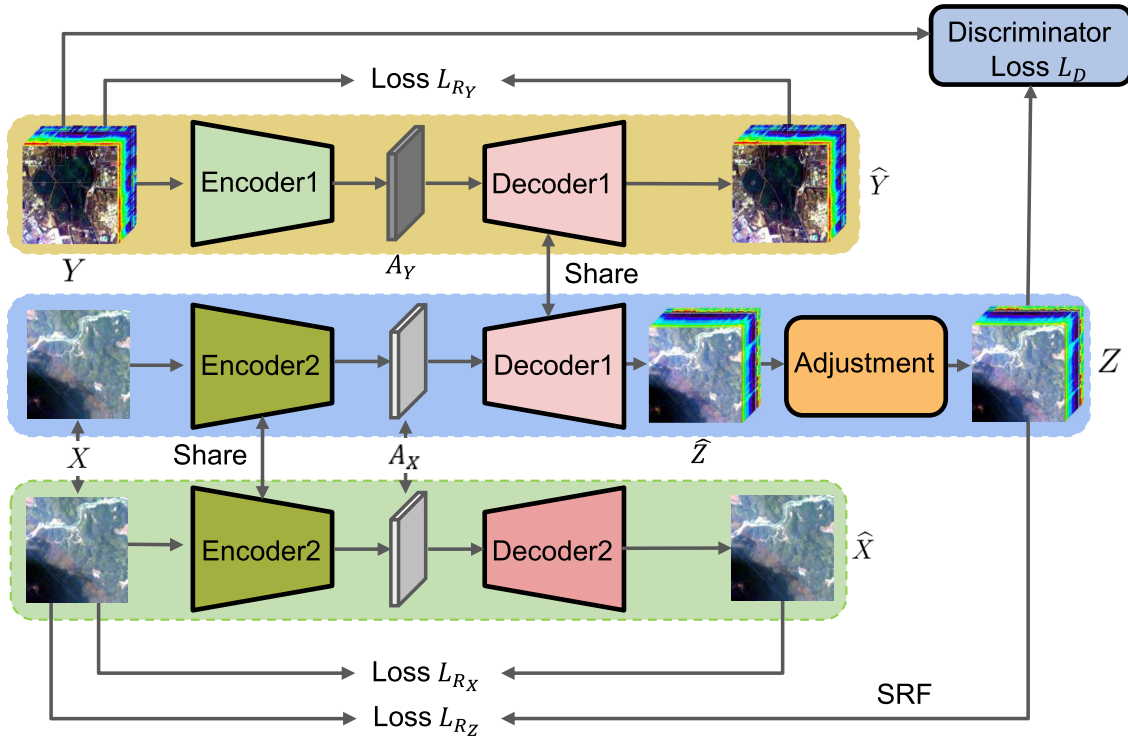


Fig. 1. Illustration of UnGUN. This framework contains three main branches: the hyperspectral unmixing branch (yellow box), the RGB unmixing branch (green box), and the reconstruction branch (blue box).

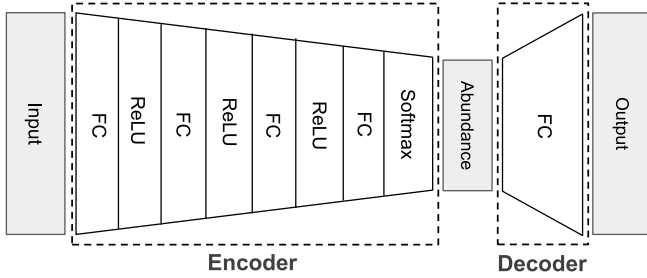


Fig. 2. Illustration of the spectral unmixing structure. The encoder is in the green box and decoder is in the pink box.

To ensure the reconstruction effect, we utilize L2 loss to minimize the pixel-wise error between Y and \hat{Y} as Eq.17:

$$L_{RY} = \|Y - \hat{Y}\|_2 \quad (17)$$

Moreover, since each pixel usually contains only several pure spectra, in reality, the abundance should be sparse. Thus we adopt the sparsity loss as Eq.18 based on KL divergence between abundance and a tiny constant ϵ .

$$\begin{aligned} L_{SY} &= \sum_{j=1}^{hw} KL(\epsilon \| A_{Y_j}) \\ &= \sum_{j=1}^{hw} \left(\epsilon \log \frac{\epsilon}{A_{Y_j}} + (1 - \epsilon) \log \frac{1 - \epsilon}{1 - A_{Y_j}} \right) \end{aligned} \quad (18)$$

where KL is the standard divergence:

$$KL(p \| q) = p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q} \quad (19)$$

Thus, the total loss of the HSI unmixing branch is the weighted sum of Eq.17-Eq.18.

$$L_Y = \alpha_Y L_{RY} + \beta_Y L_{SY} \quad (20)$$

where α_Y and β_Y are trade-off weights.

It is worth noting that in the experiment, the decoder of the HSI unmixing branch, namely the spectral endmember matrix, is initialized from the Salinas dataset by the VCA method.

2) *RGB Unmixing Branch*: Same as HSI unmixing branch, RGB unmixing branch consists of two parts: Encoder2: $En_2 : \mathcal{R}^{HW \times 3} \rightarrow \mathcal{R}^{HW \times L}$, which maps the RGB image to the latent representation A_X , and Decoder2: $De_2 : \mathcal{R}^{HW \times L} \rightarrow \mathcal{R}^{HW \times 3}$, which reconstructs the input RGB image with A_X as the endmember M_X . The RGB unmixing branch can be expressed as:

$$A_X = En_2(X) \quad (21)$$

$$\hat{X} = De_2(A_X) \quad (22)$$

Similarly, the reconstruction loss and sparsity loss are represented by Eq.23 and Eq.24:

$$L_{RX} = \|X - \hat{X}\|_2 \quad (23)$$

$$\begin{aligned} L_{SX} &= \sum_{i=1}^{HW} KL(\epsilon \| A_{X_i}) \\ &= \sum_{j=1}^{HW} \left(\epsilon \log \frac{\epsilon}{A_{X_j}} + (1 - \epsilon) \log \frac{1 - \epsilon}{1 - A_{X_j}} \right) \end{aligned} \quad (24)$$

Then the total loss of the RGB unmixing branch is the weighted sum of Eq.23-Eq.24.

$$L_X = \alpha_X L_{RX} + \beta_X L_{SX} \quad (25)$$

where α_X and β_X are trade-off weights.

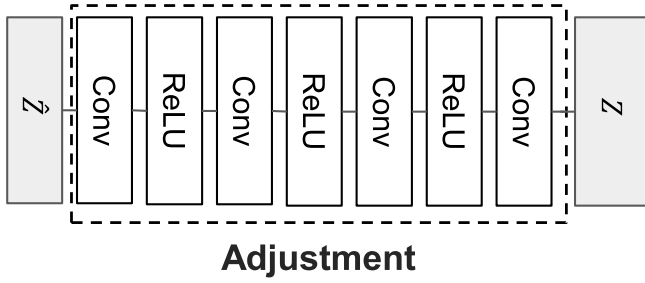


Fig. 3. Illustration of the adjustment module, which consists of four convolution operators and three ReLU operators.

D. Reconstruction Branch

In this branch, we utilize the spatial information of RGB images and spectral information of guidance HSI to reconstruct high-resolution HSI Z . A coarse HSI \hat{Z} is generated first with unmixing structure. The encoder of reconstruction branch shares parameters with RGB unmixing branch, which is Encoder2: $En_2: \mathcal{R}^{HW \times 3} \rightarrow \mathcal{R}^{HW \times L}$. After Encoder2, the RGB input is decomposed to abundance A_X , which contains the texture characteristics of X . Then the coarse image \hat{Z} is obtained through the decoder, which shares parameters with the HSI unmixing branch De_1 . Since the decoder represents the spectral endmember of Y , we generate an initialized HSI \hat{Z} with the help of guidance HSI. The reconstruction of coarse \hat{Z} can be expressed as:

$$A_X = En_2(X) \quad (26)$$

$$\hat{Z} = De_1(A_X) \quad (27)$$

With the help of guidance image, \hat{Z} obtains the spectral information of real world hyperspectral image. However, there are still differences between the overall distribution of real hyperspectral data and \hat{Z} . Therefore, we designed an adjustment module to refine the texture and color information. In order to obtain a more accurate HSI, \hat{Z} is input into the adjustment module $Adj: \mathcal{R}^{HW \times C} \rightarrow \mathcal{R}^{HW \times C}$. Then the refined HSI Z is reconstructed:

$$Z = Adj(\hat{Z}) \quad (28)$$

For reconstruction branch, we adopt two loss function: reconstruction loss L_{RZ} and TV loss L_{TV} . Z is the HSI obtained by X-spectral super-resolution. Therefore, the RGB image generated from Z after the spectral response function should be the same as X . Thus we utilize L2 loss to minimize the error between X and $R \cdot Z$ as Eq.29:

$$L_{RZ} = \|X - R \cdot Z\|_2 \quad (29)$$

where R is the spectral response function.

Another physical property of HSI is that for a pixel, the difference between adjacent channels is not too large. Based on the characteristic, we introduce total variation loss L_{TV} :

$$L_{TV} = \sum_{i=1}^{HW} \sum_{j=1}^{C-1} (Z_{i,j} - Z_{i,j+1}) \quad (30)$$

Algorithm 1 The Pseudocode for UnGUN

for number of training iterations **do**
 for k steps **do**
 Sample minibatch of m RGB image samples $\mathbf{X} = \{X^{(1)}, \dots, X^{(m)}\}$ from RGB dataset.
 Sample minibatch of m guidance HSI samples $\mathbf{Y} = \{Y^{(1)}, \dots, Y^{(m)}\}$ from HSI dataset.
 Update the discriminator by ascending its stochastic gradient.
 end for
 Sample minibatch of m RGB image samples $\mathbf{X} = \{X^{(1)}, \dots, X^{(m)}\}$ from RGB dataset.
 Sample minibatch of m guidance HSI samples $\mathbf{Y} = \{Y^{(1)}, \dots, Y^{(m)}\}$ from HSI dataset.
 Update the generator (including HSI unmixing branch, RGB unmixing branch and reconstruction branch) by descending its stochastic gradient.
end for

Then the total loss of the reconstruction branch is the weighted sum of Eq.29-Eq.30.

$$L_Z = \alpha_Z L_{RZ} + \beta_Z L_{TV} \quad (31)$$

where α_Z and β_Z are trade-off weights.

E. Discriminator

To make the super-resolved HSI Z more realistic, we design a discriminator to make the distribution of Z more similar to the real hyperspectral image Y . The objective of the discriminator is to force the distribution of final high-resolution hyperspectral images Z to resemble the real-world hyperspectral images.

The loss function of the generator (including the HSI unmixing branch, RGB unmixing branch, and reconstruction branch) is expressed as Eq.32:

$$L_G = L_g + L \quad (32)$$

$$L_g = BCE(D(Z), 1) \quad (33)$$

$$L = \alpha L_X + \beta L_Y + \lambda L_Z \quad (34)$$

where $BCE(\cdot)$ is binary cross entropy loss, $D(\cdot)$ is the discriminator, α , β and γ are hyperparameters to balance the different losses.

The loss function of the discriminator is:

$$L_D = BCE(D(Z), 0) + BCE(D(Y), 1) \quad (35)$$

In the training process, we optimize the generator and discriminator alternately and finally reach convergence. Besides, we adopt the gradient clipping method to ensure the stability of adversarial learning and avoid the problem of gradient explosion.

III. EXPERIMENTS

During this section, we represent the results and analysis of comparison experiments on four real datasets: Salinas dataset, ICVL dataset, TG1HRSSC dataset, and DFC2018 Houston

TABLE I
THE QUANTITATIVE RESULTS OF DIFFERENT METHODS ON ICVL DATASET

dataset		BI	DBF [57]	Semantic-embedded [58]	UnGUN	HSCNN-D (supervised) [60]
ICVL	PSNR	24.2597	27.2339	33.8222	33.9775	40.2824
	SSIM	0.9549	0.9562	0.9839	0.9854	0.9915
	SAM	13.0948	10.2609	4.9741	4.6195	1.3813
DFC2018 Houston	PSNR	20.9552	23.6756	26.9625	28.5009	37.4834
	SSIM	0.8109	0.8283	0.8355	0.8737	0.9400
	SAM	25.6224	26.1969	23.0454	21.0743	6.9617
TG1HRSSC	PSNR	22.1813	24.9149	27.4564	28.0471	36.6521
	SSIM	0.8221	0.8386	0.9100	0.8620	0.9145
	SAM	11.3575	15.5577	9.8538	9.3065	4.0112

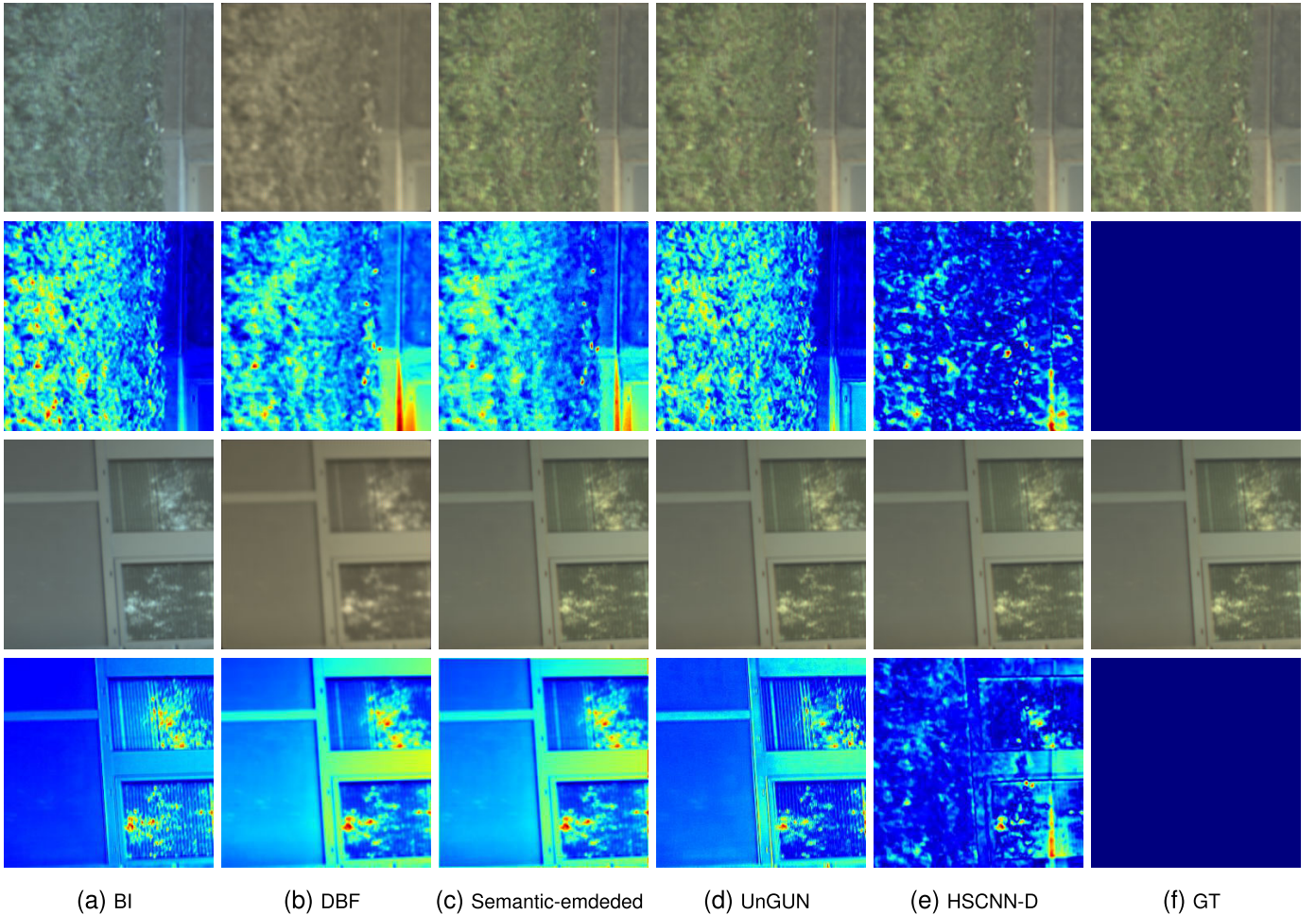


Fig. 4. Visual results of different methods on the ICVL dataset. The first and third rows are the pseudo-color images generated by the hyperspectral images after the super-resolution, and the second and fourth rows are their corresponding error maps with the real images.

dataset, where Salinas dataset is utilized as guidance data, and the remaining are used for super-resolution. Moreover, to verify the effectiveness of each module, we perform ablation experiments on the ICVL dataset.

A. Compared Algorithms and Quantitative Assessments

We select four comparison methods, which are Bilinear Interpolation (BI), DBF [57], Semantic-embedded method [58] and HSCNN-D [60]. Bilinear interpolation is one of the most common up-sampling methods. It generates new pixels

by the adjacent pixels to realize the super-resolution. DBF and Semantic-embedded method are unsupervised strategies, which do not need paired hyperspectral-RGB data. However, since the Semantic-embedded method requires semantic segmentation information, which is difficult to obtain for the dataset we use, we remove the semantic-embedded regularization when conducting the comparison experiment. HSCNN-D is a supervised spectral super-resolution algorithm, thus its results represent the upper bound of unsupervised algorithms to some extent.

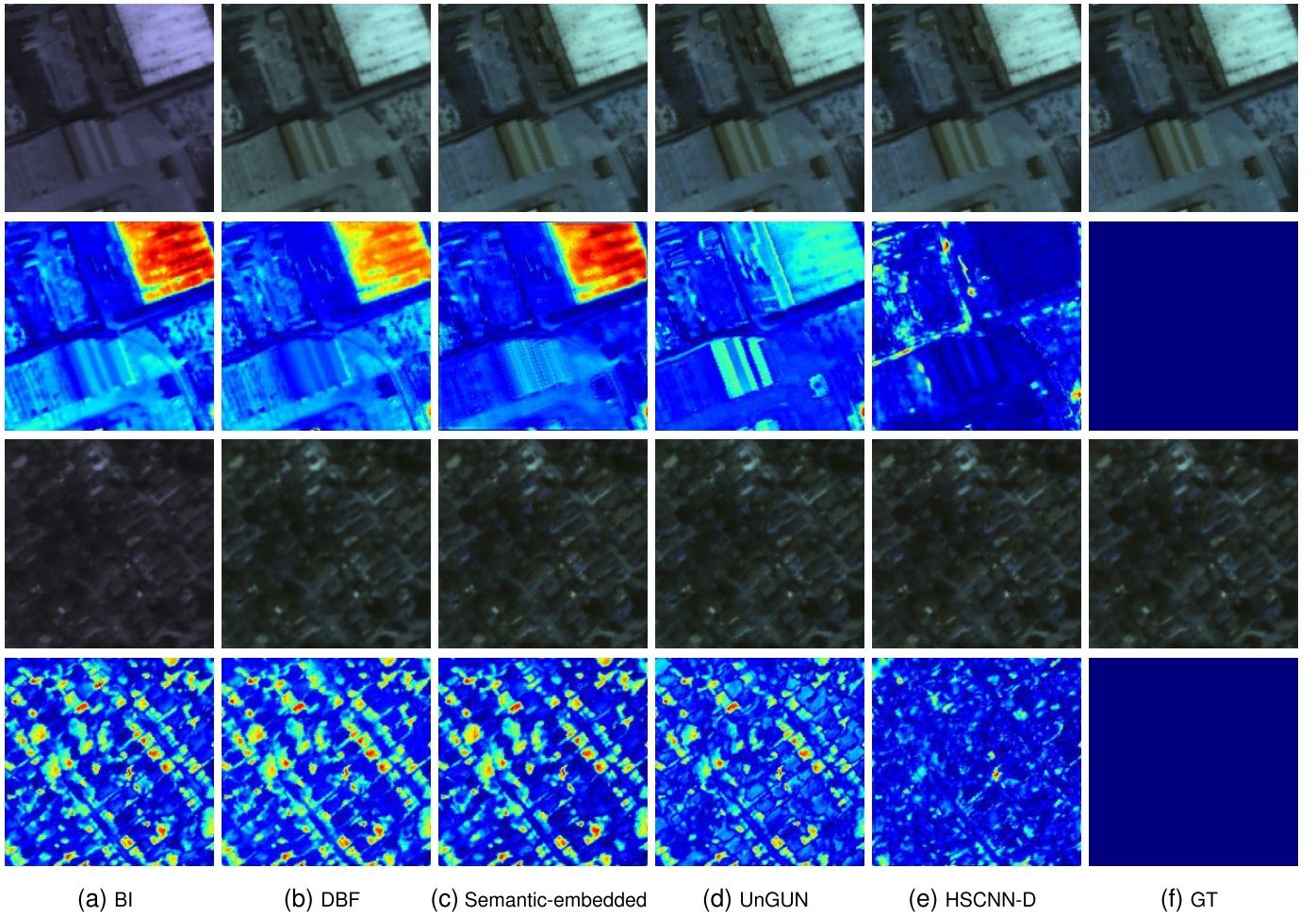


Fig. 5. Visual results of different methods on the DFC2018 Houston dataset. The first and third rows are the pseudo-color images generated by the hyperspectral images after the super-resolution, and the second and fourth rows are their corresponding error maps with the real images.

We adopt three quantitative assessments to evaluate the effectiveness of super-resolution: PSNR, SSIM and SAM. They are respectively defined as follows:

$$\text{MSE} = \frac{1}{N} \sum_{i=0}^N \|x_i - y_i\|^2 \quad (36)$$

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) = 20 \cdot \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right) \quad (37)$$

where MAX_I is the maximum value of the pixel point, for example, if each sample point is expressed in 8 bits, MAX_I is 255. x and y are two images. N is the number of pixels of x and y .

$$\text{SSIM} = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (38)$$

where μ_x and μ_y represent the means of x and y , σ_x and σ_y are variances, σ_{xy} is the covariance, and c_1 , c_2 are manual constants.

$$\text{SAM} = \frac{1}{N} \sum_{i=1}^N \arccos \left(\frac{x_i^T y_i}{\|x_i\| \|y_i\|} \right) \quad (39)$$

where N is the number of pixels of x and y .

B. Dataset and Experiment Setting

1) *Guidance Hyperspectral Dataset*: Since the endmember matrix of guidance HSI needs to be utilized to provide prior for RGB image spectral reconstruction, the number of channels should be larger than that of the RGB image requiring spectral super resolution. Thus we adopt Salinas dataset, Indian Pines dataset and Washington DC dataset as guidance HSI dataset.

Salinas dataset informs the Salinas Valley in California, USA. The corrected image contains 204 channels, which cover the 0.4-2.5 μm spectral range.

The Washington DC dataset is an aerial hyperspectral image obtained by Hydice sensor. The data contained 191 bands in the visible and near-infrared bands ranging from 0.4 to 2.4 μm , with data sizes of 1208×307 .

Indian Pines dataset is gathered by AVIRIS sensor over the Indian Pines test site in North-western Indiana and consists of 145×145 pixels and 224 spectral reflectance bands in the wavelength range 0.4-2.5 μm spectral range.

The guidance image should be down-sampled according to the spectral wavelength, so that the down-sampled image has the same number of channels as the hyperspectral image

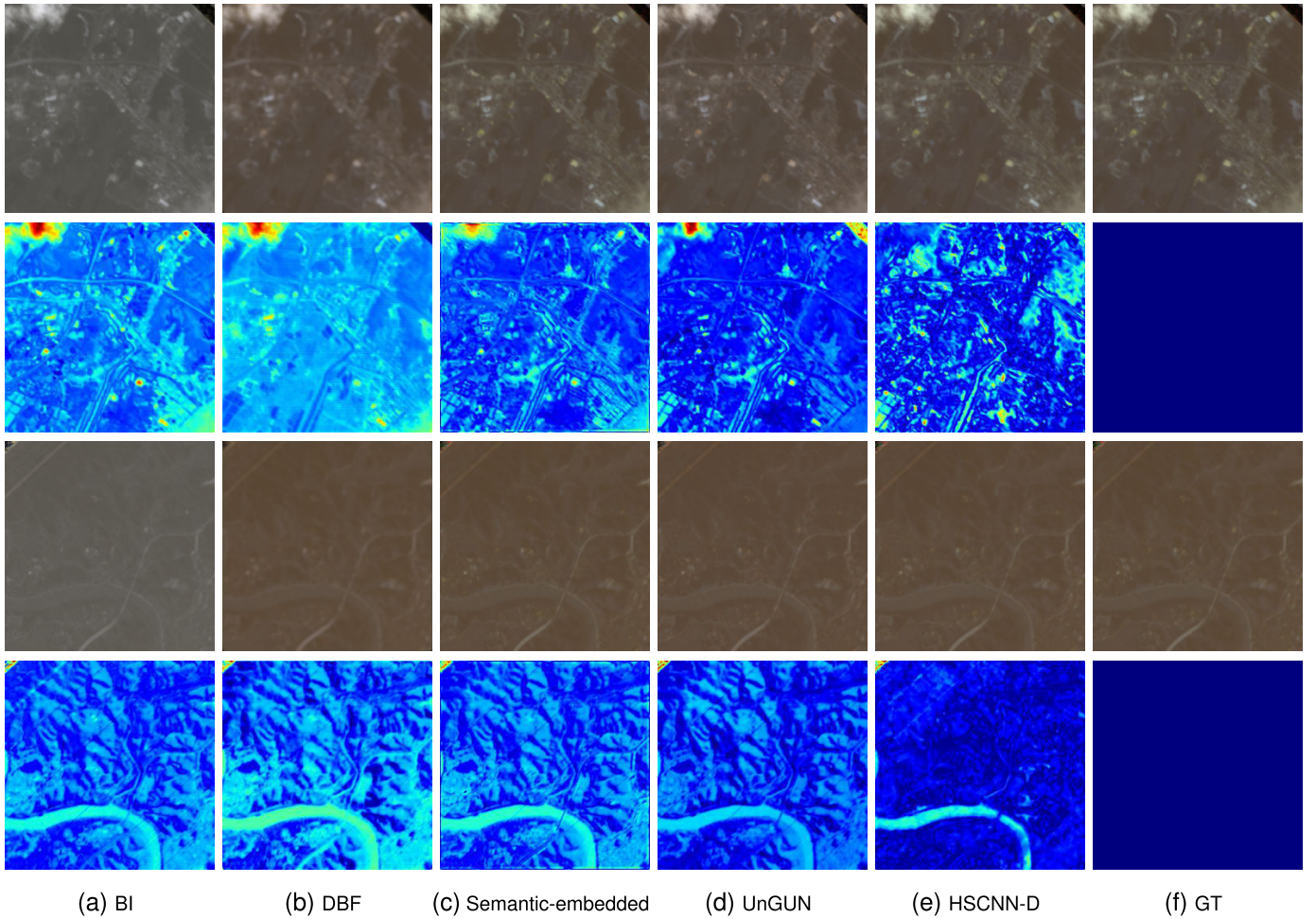


Fig. 6. Visual results of different methods on the TG1HRSSC dataset. The first and third rows are the pseudo-color images generated by the hyperspectral images after the super-resolution, and the second and fourth rows are their corresponding error maps with the real images.

obtained from the spectral super-resolution. Moreover, the three datasets are divided into 64×31 patches before experiments.

2) *ICVL Dataset*: ICVL dataset is made up of 201 images containing 31 spectral channels. We crop the complete image into $31 \times 256 \times 256$ subimages for training. RGB images are generated by SRF and corresponding hyperspectral images. 64 images are randomly selected as the training set and 8 images are selected as the test set. During training, the synthesized RGB images are used as the inputs, while the paired hyperspectral images do not need to be fed into the network. Moreover, downsampled Salinas image with 31 channels will be simultaneously entered into the network as a guidance image.

3) *DFC2018 Houston Dataset*: DFC2018 Houston dataset was released by NCALM. It contains images with 50 channels ranging from 380-1050 nm. 70 images are randomly selected as the training set and 10 images are selected as the test set. Same as the ICVL dataset, RGB images after SRF and down-sampled Salinas data with 50 channels are used as the inputs of the network.

4) *TG1HRSSC Dataset*: TG1HRSSC dataset contains three types of hyperspectral data. Visible near-infrared spectral data is selected for experiments, which contains 54 channels

covering 400-900 nm. 40 images are randomly selected as the training set and 7 images are selected as the test set. Similarly, RGB images after SRF and downsampled 54 channels-Salinas image are input into the network.

C. Comparison Experiments

1) *Performance on ICVL Dataset*: Table I shows the quantitative results of the UnGUN and comparison methods. Bilinear interpolation is the simplest operation, thus its quantization performance is worse than the other four methods. It can also be seen that UnGUN performs much better than bilinear interpolation and unsupervised algorithm DBF. In addition, compared with the Semantic-embedded method, although the quantitative results of our algorithm are close to those of the semantic embedding method, our three measurements are still better, which indicates the superiority of our method. PSNR measures absolute error between reconstructed image and ground truth. SSIM and SAM represent structure difference and spectral difference. It is observed that our algorithm outperforms other unsupervised algorithms in many ways, especially spectral difference. Since our network takes spectral characteristics into account, we get the best result in SAM. HSCNN-D achieves the best results, which is a reasonable

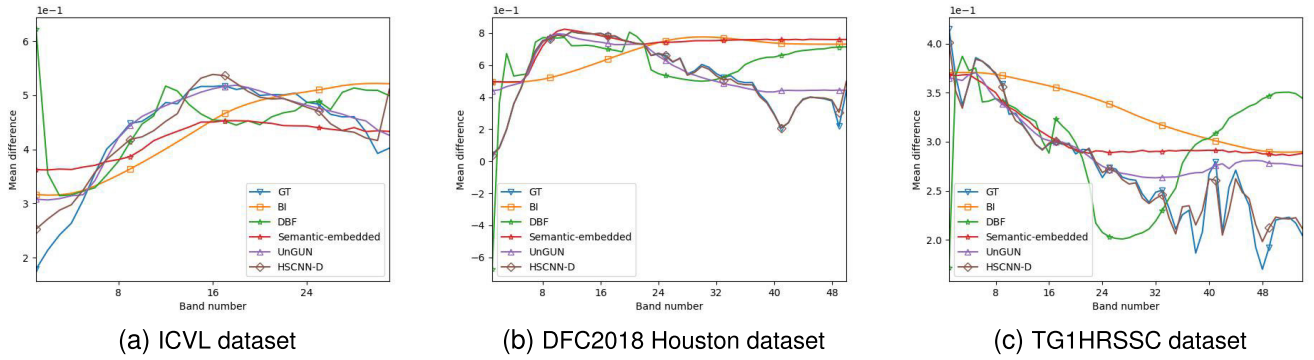


Fig. 7. Pixel value curves of randomly selected pixels on three datasets.

TABLE II
QUANTITATIVE RESULTS OF ABLATION EXPERIMENTS ON THE ICVL DATASET

Options	complete	without unmixing	without adjustment	without discriminator
unmixing structure	✓		✓	✓
adjustment	✓	✓		✓
discriminator	✓	✓	✓	
PSNR	33.9775	26.8966	31.5829	31.7162
SSIM	0.9854	0.9667	0.9824	0.9824
SAM	4.6195	5.0632	5.4890	5.5481

result since the real spectral information are utilized to optimize the super-resolution network.

Figure 4 displays the visualization of the super-resolution effects of each method. HSCNN-D is a supervised algorithm, thus the error between the super-resolved hyperspectral image and the ground truth is the smallest. For the remaining four methods, they lack the guidance of the labels during training, so the error is larger. And it is obvious that there is a color difference between the real images and the images generated by the bilinear interpolation method and DBF method. The other three methods show no such phenomenon, which is consistent with the quantitative results. Moreover, we can see that the error of UnGUN is smaller than the other three unsupervised methods, which means our method preserves more spatial and spectral characteristics.

To further validate the super-resolution performance, one pixel is randomly selected from the super-resolved hyperspectral image. The values of this pixel from the 1st to the 31st channels are plotted in Figure 7a. The curve corresponding to HSCNN-D is the closest to the true value. Bilinear interpolation and the DBF method do not achieve satisfactory results. And there is no obvious superiority between our method and the Semantic-embedded method.

2) *Performance on DFC2018 Houston Dataset:* Table I shows the quantitative results on the DFC2018 Houston dataset. Same as the ICVL dataset, bilinear interpolation performs worst since it has no other operations to extract and reconstruct spatial and spectral features. DBF achieves better effectiveness than bilinear interpolation, but there is still a gap between DBF and the Semantic-embedded method. HSCNN-D still gets far better results than the other four unsupervised

methods. And our network outperforms all the other unsupervised algorithms in terms of absolute error, structural error and spectral error, which validates the effectiveness of unmixing framework and guidance images.

Figure 5 represents the visualization results. For the HSCNN method, the generated hyperspectral image has almost no error from the real image. And for the remaining three methods, UnGUN is superior to bilinear and DBF method. However, the Semantic-embedded method surpasses our network in the performance of the second image but is inferior to our algorithm.

From Figure 7b, we can see that the pixel values of the image obtained by the HSCNN-D method are almost the same as the real values. The curve corresponding to our method is slightly different from the real image at channels 32 to 50, but it is still more accurate than the other three unsupervised methods.

3) *Performance on TG1HRSSC Dataset:* Table I shows the quantitative results of the TG1HRSSC dataset. Similarly, the HSCNN-D method is far superior to the other four methods in PSNR, SSIM, and SAM. Our method has the best effect from the perspective of PSNR and SAM among the remaining four methods, which demonstrates that spatial-spectral information is maintained. However, under the index of SSIM, the Semantic-embedded method is slightly better than ours.

Figure 6 represents the visualization results. The images generated by the bilinear interpolation may have color differences from the real images. And the reconstruction effect of our method is relatively satisfactory.

Figure 7c displays the values of the random pixel. Same as the DFC2018 Houston dataset, the reconstructed hyperspectral

image from HSCNN-D is the closest to the real one, and the image obtained by our method is second only to HSCNN-D. This also verifies the effectiveness of our method.

D. Ablation Experiments

To validate the effectiveness of each module in UnGUN, we conduct ablation experiments on the ICVL dataset. We design three ablation experiments based on the unmixing structure, adjustment module, and discriminator, obtained quantitative results are shown in Table II.

For unmixing structure, we use ResNet to replace the unmixing structure, and the replaced network needs only one input: RGB image. The effect of the network without unmixing structure is worse than that of the complete network under the three assessments, especially the PSNR, which reaches 26.8966, even worse than the DBF method. It indicates that, as the backbone of the network, the unmixing structure is essential.

The function of the adjustment module is to refine the initial coarse hyperspectral image, in this way, the final super-resolution image has accurate spatial and spectral information. After removing the adjustment module, the obtained hyperspectral image information is lost and the reconstruction effect is reduced.

As for the discriminator, it promotes the distribution of the reconstructed hyperspectral image as close as possible to the real hyperspectral data. Since there is no label in the unsupervised strategy, we cannot guarantee that the reconstructed hyperspectral image conforms to the real-world distribution without the discriminator. The results in the table also show that the network effect is reduced when the discriminator is removed.

IV. CONCLUSION

Inspired by spectral unmixing, we propose an unsupervised spectral super-resolution algorithm, which does not require paired hyperspectral-RGB images during training. Under the guidance of the guidance hyperspectral image, a coarse hyperspectral image is generated after three unmixing branches. The coarse image will be further refined through the adjustment module to get an accurate image. Moreover, a discriminator is adopted to make sure the distribution of the reconstructed image is realistic. Comparative experiments demonstrate the superiority of UnGUN. And the ablation experiments further validate the effectiveness of each module.

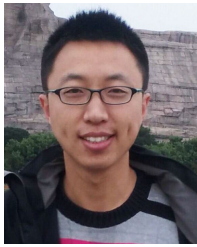
REFERENCES

- [1] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.
- [2] Q. Wang, W. Huang, Z. Xiong, and X. Li, "Looking closer at the scene: Multiscale representation learning for remote sensing image scene classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 4, pp. 1414–1428, Apr. 2022.
- [3] L. Fang, N. He, S. Li, P. Ghamisi, and J. A. Benediktsson, "Extinction profiles fusion for hyperspectral images classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1803–1815, Mar. 2018.
- [4] Z. Zou and Z. Shi, "Hierarchical suppression method for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 330–342, Jan. 2016.
- [5] K. Makantasis, K. Karantzalos, A. Doulamis, and K. Loupos, "Deep learning-based man-made object detection from hyperspectral data," in *Proc. Int. Symp. Vis. Comput.*, vol. 9474, Dec. 2015, pp. 717–727.
- [6] P. Ren, M. Xu, Y. Yu, F. Chen, X. Jiang, and E. Yang, "Energy minimization with one dot fuzzy initialization for marine oil spill segmentation," *IEEE J. Ocean. Eng.*, vol. 44, no. 4, pp. 1102–1115, Oct. 2019.
- [7] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1860–1875, Nov. 2005.
- [8] S. Gou, S. Liu, S. Yang, and L. Jiao, "Remote sensing image super-resolution reconstruction based on nonlocal pairwise dictionaries and double regularization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4784–4792, Dec. 2014.
- [9] H. Zhang, L. Zhang, and H. Shen, "A super-resolution reconstruction algorithm for hyperspectral images," *Signal Process.*, vol. 92, no. 9, pp. 2082–2096, Sep. 2012.
- [10] H. Irmak, G. B. Akar, and S. E. Yuksel, "A MAP-based approach for hyperspectral imagery super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2942–2951, Jun. 2018.
- [11] Z. Tang, Q. Xu, P. Wu, Z. Shi, and B. Pan, "Feedback refined local-global network for super-resolution of hyperspectral imagery," *Remote Sens.*, vol. 14, no. 8, p. 1944, Apr. 2022.
- [12] Q. Li, Q. Wang, and X. Li, "Exploring the relationship between 2D/3D convolution for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8693–8703, Oct. 2021.
- [13] Q. Wang, Q. Li, and X. Li, "Hyperspectral image superresolution using spectrum and feature context," *IEEE Trans. Ind. Electron.*, vol. 68, no. 11, pp. 11276–11285, Nov. 2021.
- [14] Q. Wang, Q. Li, and X. Li, "Spatial-spectral residual network for hyperspectral image super-resolution," 2020, *arXiv:2001.04609*.
- [15] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1082–1096, 2020.
- [16] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3844–3851.
- [17] X. Wang, J. Ma, and J. Jiang, "Hyperspectral image super-resolution via recurrent feedback embedding and spatial-spectral consistency regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503113, doi: [10.1109/TGRS.2021.3064450](https://doi.org/10.1109/TGRS.2021.3064450).
- [18] Y. Mei, Y. Fan, Y. Zhou, L. Huang, T. S. Huang, and H. Shi, "Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5689–5698.
- [19] J. Wu, J. Aeschbacher, and R. Timofte, "In defense of shallow learned spectral reconstruction from RGB images," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 471–479.
- [20] H. Garcia, C. V. Correa, and H. Arguello, "Multi-resolution compressive spectral imaging reconstruction from single pixel measurements," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 6174–6184, Dec. 2018.
- [21] Y. Jia et al., "From RGB to spectrum for natural scenes via manifold-based mapping," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4715–4723.
- [22] M. Parmar, S. Lancel, and B. A. Wandell, "Spatio-spectral reconstruction of the multispectral datacube using sparse recovery," in *Proc. 15th IEEE Int. Conf. Image Process.*, 2008, pp. 473–476.
- [23] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. Comput. Vis. (ECCV)*, 2016, pp. 19–34.
- [24] Z. Meng, J. Ma, and X. Yuan, "End-to-end low cost compressive spectral imaging with spatial-spectral self-attention," in *Proc. Comput. Vis. (ECCV)*, 2020, pp. 187–204.
- [25] X. Miao, X. Yuan, Y. Pu, and V. Athitsos, "Lambda-Net: Reconstruct hyperspectral images from a snapshot measurement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4058–4068.
- [26] X. Hu et al., "HDNet: High-resolution dual-domain learning for spectral compressive imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17521–17530.

- [27] T. Huang, W. Dong, X. Yuan, J. Wu, and G. Shi, "Deep Gaussian scale mixture prior for spectral compressive imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16211–16220.
- [28] L. Wang, C. Sun, M. Zhang, Y. Fu, and H. Huang, "DNU: Deep non-local unrolling for computational spectral imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1658–1668.
- [29] L. Wang, C. Sun, Y. Fu, M. H. Kim, and H. Huang, "Hyperspectral image reconstruction using a deep spatial-spectral prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8024–8033.
- [30] R. Hang, Q. Liu, and Z. Li, "Spectral super-resolution network guided by intrinsic properties of hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 30, pp. 7256–7265, 2021.
- [31] J. He, J. Li, Q. Yuan, H. Shen, and L. Zhang, "Spectral response function-guided deep optimization-driven network for spectral super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4213–4227, Sep. 2022.
- [32] L. Liu, Z. Zou, and Z. Shi, "Hyperspectral remote sensing image synthesis based on implicit neural spectral mixing models," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5500514.
- [33] Y. Cai et al., "Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17481–17490.
- [34] Y. Cai et al., "MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 744–754.
- [35] J. Liu, Z. Wu, L. Xiao, J. Sun, and H. Yan, "A truncated matrix decomposition for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 29, pp. 8028–8042, 2020.
- [36] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.
- [37] J. Xue, Y.-Q. Zhao, Y. Bu, W. Liao, J. C. Chan, and W. Philips, "Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 3084–3097, 2021.
- [38] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability," *IEEE Trans. Image Process.*, vol. 29, pp. 116–127, 2020.
- [39] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, Oct. 2019.
- [40] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.
- [41] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3862–3871.
- [42] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 3034–3047, Jun. 2019.
- [43] R. Dian, S. Li, L. Fang, T. Lu, and J. M. Bioucas-Dias, "Nonlocal sparse tensor factorization for semiblind hyperspectral and multispectral image fusion," *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 4469–4480, Oct. 2020.
- [44] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3631–3640.
- [45] J.-F. Hu, T.-Z. Huang, L.-J. Deng, T.-X. Jiang, G. Vivone, and J. Chanussot, "Hyperspectral image super-resolution via deep spatio-spectral attention convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7251–7265, Dec. 2022, doi: [10.1109/TNNLS.2021.3084682](https://doi.org/10.1109/TNNLS.2021.3084682).
- [46] Z. Zhu, J. Hou, J. Chen, H. Zeng, and J. Zhou, "Hyperspectral image super-resolution via deep progressive zero-centric residual learning," *IEEE Trans. Image Process.*, vol. 30, pp. 1423–1438, 2021.
- [47] T. Zhang, Y. Fu, L. Wang, and H. Huang, "Hyperspectral image reconstruction using deep external and internal learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8558–8567.
- [48] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised adaptation learning for hyperspectral imagery super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3070–3079.
- [49] W. Wei, J. Nie, L. Zhang, and Y. Zhang, "Unsupervised recurrent hyperspectral imagery super-resolution using pixel-aware refinement," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5500315, doi: [10.1109/TGRS.2020.3039534](https://doi.org/10.1109/TGRS.2020.3039534).
- [50] B. Pan, Q. Qu, X. Xu, and Z. Shi, "Structure-color preserving network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5520512, doi: [10.1109/TGRS.2021.3135028](https://doi.org/10.1109/TGRS.2021.3135028).
- [51] W. Dong, C. Zhou, F. Wu, J. Wu, G. Shi, and X. Li, "Model-guided deep hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 5754–5768, 2021.
- [52] X. Wang, J. Chen, Q. Wei, and C. Richard, "Hyperspectral image super-resolution via deep prior regularization with parameter estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 1708–1723, Apr. 2022.
- [53] Q. Xie, M. Zhou, Q. Zhao, Z. Xu, and D. Meng, "MHF-Net: An interpretable deep network for multispectral and hyperspectral image fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1457–1473, Mar. 2022.
- [54] Y. Qu, H. Qi, C. Kwan, N. Yokoya, and J. Chanussot, "Unsupervised and unregistered hyperspectral image super-resolution with mutual Dirichlet-net," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5507018, doi: [10.1109/TGRS.2021.3079518](https://doi.org/10.1109/TGRS.2021.3079518).
- [55] K. Zheng et al., "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2487–2502, Mar. 2021.
- [56] J. Yao, D. Hong, J. Chanussot, D. Meng, X. Zhu, and Z. Xu, "Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution," in *Proc. Comput. Vis. (ECCV)*, 2020, pp. 208–224.
- [57] B. J. Fubara, M. Sedky, and D. Dyke, "RGB to spectral reconstruction via learned basis functions and weights," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1984–1993.
- [58] Z. Zhu, H. Liu, J. Hou, H. Zeng, and Q. Zhang, "Semantic-embedded unsupervised spectral reconstruction from single RGB images in the wild," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2259–2268.
- [59] J. M. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
- [60] Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu, "HSCNN+: Advanced CNN-based hyperspectral recovery from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1052–10528.



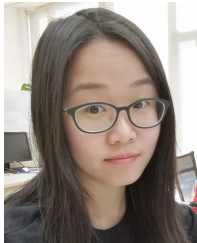
Qiaoying Qu received the B.S. degree from the College of Mathematics, Jilin University, Changchun, China, in 2021. She is currently pursuing the M.S. degree with the School of Statistics and Data Science, Nankai University. Her research interests include machine learning, hyperspectral super-resolution, and remote sensing image processing.



Bin Pan (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Astronautics, Beihang University, Beijing, China, in 2013 and 2019, respectively. Since 2019, he has been an Associate Professor with the School of Statistics and Data Science, Nankai University. His research interests include machine learning, remote sensing image processing, and multi-objective optimization.



Tao Li received the Ph.D. degree in computer science from Nankai University, Tianjin, China, in 2007. He is with the College of Computer Science, Nankai University, as a Professor. His main research interests include heterogeneous computing, machine learning, and the Internet of Things. He is a Distinguished Member of the China Computer Federation (CCF).



Xia Xu (Graduate Student Member, IEEE) received the B.S. and M.S. degrees from the School of Electrical Engineering, Yanshan University, Qinhuangdao, China, in 2012 and 2015, respectively, and the Ph.D. degree from the School of Astronautics, Beihang University, Beijing, China, in 2019. She is an Assistant Professor with the College of Computer Science, Nankai University, Tianjin, China. Her research interests include hyperspectral unmixing, multiobjective optimization, and remote sensing image processing.



Zhenwei Shi (Member, IEEE) received the Ph.D. degree in mathematics from the Dalian University of Technology, Dalian, China, in 2005. He was a Postdoctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA, from 2013 to 2014. He is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing. He has authored or coauthored 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 research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning. He serves as an Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the *Pattern Recognition*, the *ISPRS Journal of Photogrammetry and Remote Sensing*, and the *Infrared Physics and Technology*. His personal website is <http://levir.buaa.edu.cn/>.