Model-Informed Multistage Unsupervised Network for Hyperspectral Image Super-Resolution

Jiaxin Li, Ke Zheng, Lianru Gao, *Senior Member, IEEE*, Li Ni, Min Huang, and Jocelyn Chanussot, *Fellow, IEEE*

***Abstract*— By fusing a low-resolution hyperspectral image (LrMSI) with an auxiliary high-resolution multispectral image (HrMSI), hyperspectral image super-resolution (HISR) can generate a high-resolution hyperspectral image (HrHSI) economically. Despite the promising performance achieved by deep learning (DL), there are still two challenges remaining to be solved. First, most DL-based methods heavily rely on large-scale training triplets, which reduces them to limited generalization and poor practicability in real-world scenarios. Second, existing methods pursue higher performance by designing complex structures from off-the-shelf components while ignoring inherent information from the degradation model, hence leading to insufficient integration of domain knowledge and lower interpretability. To address those drawbacks, we propose a model-informed multistage unsupervised network, M2U-Net for short, by leveraging both deep image prior (DIP) and degradation model information. Generally, M2U-Net is built with a three-stage scheme, i.e., degradation information learning (DIL), initialized image establishment (IIE), and deep image generation (DIG) stages. The first stage is to exploit the deep information of the degradation model via a tiny network whose parameters and outputs will serve as guidance for the following two stages. Instead of feeding uninformed noise as input for stage three, the IIE stage aims to establish an initialized input with expressive HrHSI-relevant information by resorting to a spectral mapping learning network (SML-Net), thus facilitating the extraction of prior information and further magnifying the potential of DIP for high-quality reconstruction. Finally, we propose a dual U-shape network (Dual U-Net) as a powerful regularizer to capture image statistics, in which two U-Nets are coupled together by a cross-attention guidance (CAG) module to separately achieve spatial feature extraction and final image generation. The CAG module can incorporate abundant spatial information into the reconstruction process and, hence, guide the network toward a more plausible generation. Extensive experiments demonstrate the effectiveness of our proposed M2U-Net in terms of quantitative evaluation and visual quality. The code will be available at https://github.com/JiaxinLiCAS.**

***Index Terms*— Deep image prior (DIP), hyperspectral image (HSI), super-resolution, unsupervised learning.**

Nomenclature

Abbreviation	*Description*
CAG	Cross-attention guidance.
CIE	Cross-channel interaction and enhancement.
DIG	Deep image generation.
DIL	degradation information learning.
DIL-Net	Degradation information learning network.
DIP	Deep image prior.
DL	Deep learning.
Dual U-Net	Dual U-shape network.
GT	Ground truth.
HISR	Hyperspectral image super-resolution.
HrHSI	High-resolution hyperspectral image.
HrMSI	High-resolution multispectral image.
IIE	initialized image establishment.
LrHSI	Low-resolution hyperspectral image.
LrMSI	Low-resolution multispectral image.
PSF	Point spread function.
SML-Net	Spectral mapping learning network.
SRF	Spectral response function.

I. Introduction

HYPERSPECTRAL images (HSIs) are capable of providing subtle spectral information ranging from the visible to the infrared band, exhibiting promising potential in various fields, such as remote sensing, medical diagnosis, and food safety [1], [2], [3], [4], [5], [6], [7]. However, to maintain a sufficient signal-to-noise ratio (SNR), there is an inevitable decrease in the spatial resolution, which severely impedes its subsequent applications, especially fine-grained tasks [8], [9], [10], [11]. Conversely, multispectral images (MSIs) are captured with wider bandwidth, hence equipped with higher spatial resolution. Under this premise, it is a straightforward

Manuscript received 7 February 2024; revised 21 March 2024 and 15 April 2024; accepted 16 April 2024. Date of publication 18 April 2024; date of current version 1 May 2024. This work was supported by the National Key Research and Development Program of China under Grant 2021YFB3900502. *(Corresponding author: Ke Zheng.)*

Jiaxin Li is with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: lijiaxin203@mails.ucas.ac.cn).

Ke Zheng is with the College of Geography and Environment, Liaocheng University, Liaocheng 252059, China (e-mail: zhengke@lcu.edu.cn).

Lianru Gao and Li Ni are with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: gaolr@aircas.ac.cn; nili@aircas.ac.cn).

Min Huang is with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the School of Optoelectronics, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: huangmin@aircas.ac.cn).

Jocelyn Chanussot is with CNRS, Grenoble INP, GIPSA-Lab, University of Grenoble Alpes, 38000 Grenoble, France, and also with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: jocelyn@hi.is).

Digital Object Identifier 10.1109/TGRS.2024.3391014

and economical way to produce an HrHSI by fusing LrHSI and corresponding HrMSI, which is termed HISR [12], [13], [14], [15], [16].

The past decade has witnessed great development of HISR, in which pansharpening-, factorization-, and DL-based methods constitute the mainstream direction [17], [18], [19], [20]. In the early attempt, HISR is reformulated into multiple pansharpening problems that are solved in parallel by component substitution or multiresolution analysis [21]. However, this intuitive strategy inevitably leads to undesirable distortion, especially in complex scenarios. By designing different regularization terms to characterize the property of unknown HrHSI, factorization-based methods regard HISR as an ill-posed inverse problem [22], [23], [24], [25]. Therefore, its performance heavily depends on the accuracy of the prior hypothesis for the observed scenes.

Thanks to the powerful modeling ability of DL, the bottleneck of traditional methods is further broken through [26], [27], [28], [29], [30], [31], [32]. Despite the encouraging performance achieved by neural networks, there are still some drawbacks remaining to be solved. First, mainstream methods work in a supervised manner and, thus, need large numbers of training triplets, i.e., LrHSIs, HrMSIs, and HrHSIs, to achieve satisfactory results [33]. However, the unavailability of HrHSI GT makes the training of networks rely on spatially downsampled datasets, hence leading to insufficient utilization of original information [34]. Moreover, the supervised network only yields one satellite-specific model that performs well in particular sensors or similar scenes. In other words, it suffers from limited generalization ability when working with inconsistent testing data, e.g., different spectral bands or degradation models [35]. Second, existing DL models treat HISR as a black-box problem by designing complex structures to realize the nonlinear mapping learning and, hence, ignore the inherent domain knowledge of HISR, i.e., rich information from the degradation model. Considering those issues, endowing a DL network in an unsupervised manner while leveraging the degradation model is an urgent problem to be solved, but it still lacks study. Luckily, DIP, as an unsupervised technology, only requires the degraded images to realize the restoration of natural images. Hence, many researchers have begun to apply it to the task of HISR. However, the intrinsic differences between natural images and HSIs greatly hinder the potential of DIP. First, it is more difficult to generate desirable results from uninformed random noise due to the complicated spatial–spectral structure in latent HrHSI. Second, the primitive structure of generator networks is not sufficient to extract the image prior, hence leading to limited performance. Given this premise, we propose our M2U-Net by improving DIP in terms of the initialized input and the architecture of generator networks by virtue of the degradation model.

Specifically, M2U-Net is built with a three-stage pipeline, with DIL in the head, IIE in the body, and DIG in the tail. In the first stage, we design a tiny DIL-Net to exploit latent spatial–spectral correlations of to-be-fused images, whose parameters and outcomes will serve as guidance for the next two stages. The purpose of IIE stage is to establish an initialized image with sufficient HrHSI-relevant information for stage three, enabling DIP to better capture the image statistics and enhancing its potential for high-quality reconstruction. Specifically, a two-stream structure decorated with cascaded cross-channel interaction and enhancement (CIE) modules is chosen as the backbone of our SML-Net, capable of fully utilizing the model information from stage one and generating an initialized input for stage three. In the stage of DIG, we propose a Dual U-Net that treats the model structure as an implicit regularizer to capture image statistics. The proposed network consists of an auxiliary U-Net and a principal U-Net, one of which extracts multiscale spatial features of HrMSI and another receives corresponding features from the auxiliary U-Net via CAG module and gradually harnesses these features to achieve high-quality reconstruction. The overall process is performed on the observed HrMSI-LrHSI pair, successfully forming an unsupervised paradigm.

1) We propose a novel unsupervised network for the task of HISR. It purely depends on the observed HrMSI-LrHSI pair for network optimization without requiring extra training datasets. Moreover, the deep information from the degradation model is taken into account, empowering our model toward a more accurate reconstruction.
2) In the stage of IIE, we establish an initialized image as the input of stage three, hence providing sufficient HrHSI-relevant information for image recovery and further promoting the DIP potential for a better reconstruction.
3) By leveraging the power of DIP, we design a Dual U-Net that utilizes the model's architecture as an implicit regularizer to capture image statistics. Specifically, the proposed structure contains one auxiliary U-Net that extracts multiscale spatial features of HrMSI and one principal U-Net that receives corresponding features of the auxiliary U-Net via the CAG module and gradually achieves image reconstruction.

The rest of this article is organized as follows. In Section II, we give an introduction to the development of HISR. Sections III and IV reformulate this problem and describe the detailed structure of M2U-Net, respectively. Section V includes extensive experiments of the proposed network. Finally, Section VI summarizes this article.

II. Related Work

In this section, the mainstream algorithms of HISR are categorized into two groups, namely, traditional and DL-based methods. The former includes pansharpening- and factorization-based approaches, and the latter consists of two paradigms, i.e., supervised and unsupervised manners.

A. Traditional Methods

The early works attempt to extend pansharpening-based methods to HISR. For example, Gomez et al. [36] apply the wavelet technique to fuse HrMSI and corresponding LrHSI group-by-group according to their wavelength range. Later, Zhang and He [37] further introduce 3-D wavelet transform into HISR, which is more suitable for spatial enhancement. Selva et al. [38] design a hyperpansharpening framework by synthesizing a high-resolution image for each band of LrHSI. Generally, pansharpening-based methods are easy to implement and independent of the degradation model. However, the loss of spatial details and the distortion of spectral signatures frequently occur due to their limited modeling capacity.

By contrast, factorization-based approaches treat the HISR task as an ill-posed problem and harness the degradation model to establish the relationship between the observed HrMSI-LrHSI pair and latent HrHSI. Considering different strategies to model spatial–spectral structure, it can be further divided into matrix- and tensor-based paradigms. In the line of matrix decomposition, Han et al. [39] propose a twin-dictionary scheme to thoroughly characterize the spatial–spectral information, forming a theoretical optimization framework. Fu et al. [40] consider the local changes caused by the acquisition difference and impose an $\ell_{2,1}$ norm to compensate for the negative impacts. Li et al. [41] harness external and internal priors to regularize latent HrHSI, which considers the general image characteristics and the unique information of to-be-fused pairs. Facing the problem of spectral variability, Ren et al. [42] design a spectral unmixing-based model by introducing spatial, spectral, and index constraints. In contrast to matrix decomposition, tensor representation provides another perspective to model the structure of a three-order cube. For example, Li [43] reformulate the HISR task as the estimation of three dictionaries and a corresponding core tensor under the framework of Tucker decomposition. Following the same line, Dian et al. [44] further incorporate structure constraints into tensor factorization and perform decomposition for each clustered group, achieving satisfactory fusion performance. Similarly, Dian et al. [45] choose to harness low tensor-train rank to characterize the spatial–spectral-nonlocal correlation among similar patches. To better exploit correlations of the data cube, Xu et al. [46] employ a tensor ring to represent structure information and further introduce graph-Laplacian and Frobenius regularizations to stabilize fusion performance. Wang et al. [47] propose the first Bayesian tensor ring model, aiming to explore the sparsity property of latent tensors while achieving automatic decision-making of tensor ring rank. Similarly, Ye et al. [48] realize adaptive rank selection in the canonical polyadic factorization. Despite the satisfactory results achieved by factorization-based approaches, these regularizers need to be carefully constructed for different scenarios, which also becomes an obstacle to their further development.

B. DL-Based Methods

Unlike traditional methods, supervised approaches attempt to establish the relationship from HrMSI-LrHSI pairs to corresponding HrHSI GT and apply well-trained networks to super-resolve unseen images to desired resolution [49], [50], [51]. Under this premise, the pioneering works directly concatenate HrMSI and upsampled LrHSI as the network input and then employ a single-stream structure to realize mapping learning. For example, Han and Chen [52] simply apply a residual network to solve the HISR task and utilize the spatial information of HrMSI to guide the feature learning of middle layers. Similarly, Zhang et al. [53] propose three cascaded components with physical interpretability and impose spatial–spectral edge loss to constrain middle outputs. Instead of treating two inputs uniformly, existing methods prefer to build multistream architectures to diversely handle two different modalities. For example, Han et al. [54] design a two-way structure to separately extract spatial and spectral features and gradually merge these outputs of the same scale for image reconstruction. Considering the spectral coverage discrepancy between HrMSI and LrHSI, Sun et al. [55] divide the whole spectral range into overlapped and nonoverlapped sets and introduce component substitution to obtain high-fidelity outcomes. Focusing on the cross-modality feature utilization, Li et al. [56] construct their network backbone by cascading several feature extraction and fusion blocks, forming a coarse-to-fine learning scheme. From the perspective of loss constraints, Gao et al. [57] develop separable loss and self-supervised loss to separately enhance physical characteristics and balance spatial–spectral samples. Sun et al. [58] establish a multitask framework that can simultaneously realize noise removal and image fusion. Li et al. [59] propose a two-stage framework to realize a coarse-to-fine image reconstruction, enhancing the model representation learning. Beyond classic convolutional networks, some works successfully modify other advanced architectures for the task of HISR, aiming to eliminate the inherent limitations of local operators. For example, Jia et al. [60] develop spectral and spatial Transformers to separately handle the feature extraction of LrHSI and HrMSI, in which the pretraining strategy is introduced to enhance network performance. Cao et al. [61] harness the generation ability of diffusion models and design two task-oriented modules to decouple entangled information of input images.

Unfortunately, the aforementioned methods belong to the category of supervised paradigm and hence their competitive performance mainly relies on large-scale training triplets, which inherently reduces them to limited generalization and poor practicability in real-world scenarios. Considering those issues, researchers begin to develop unsupervised techniques for the task of HISR, aiming to get rid of the dependence on training datasets. Inspired by the unsupervised property of traditional methods, some works successfully apply the factorization spirit to the deep networks. For example, Gao et al. [50], Li et al. [62], Wu et al. [63], Zheng et al. [64], Li et al. [65], and Li et al. [66] decompose input HrMSI-LrHSI pair into the representation of abundances and endmembers and combine these latent features to reconstruct unknown HrHSI. Yang et al. [67] further embed tensor theory into the network structure and infer the desired tensors by tailor-designed layers. Besides, Liu et al. [35], [68] exploit the regularization ability of DIP and promote detail recovery under the attention mechanism. Along the same line, Nguyen et al. [69] replace vanilla ℓ_2 loss with SURE loss to avoid the overfitting problem and enhance the DIP results. Cao et al. [70] combine Transformer and convolutional neural networks to guarantee a sufficient utilization of the HrMSI-LrHSI pair. Zhang et al. [71] propose a two-stage framework to accurately model the image prior and exhibit promising performance in real scenarios. From the perspective of spectral SR, Li et al. [72], [73], [74] exploit the latent spatial–spectral correlation of input images and design a spectral learning network to directly achieve HrHSI generation from input HrMSI. Han et al. [75] consider the property of spectral similarity and, hence, perform spectral mapping for each independent cluster. Qin et al. [76] further introduce an augmentor network to enrich the input samples and promote the SR performance. Wu et al. [77] convert the HISR task into two subproblems, i.e., spectral SR and spatial SR, forming a novel fusion framework. Moreover, Guo et al. [78] and Dian et al. [79] utilize the learned degradation parameters to generate training datasets and conduct zero-shot learning for observed pairs.

III. Problem Formulation

A. Deep Image Prior

In the ill-posed inverse problem, such as denoising or super-resolution, the task can be formulated as an energy minimization problem as follows:

$$\min_{\mathbf{X}} E(\mathbf{X}, \mathbf{X_0}) + \mathcal{R}(\mathbf{X}) \tag{1}$$

where $E(\mathbf{X}, \mathbf{X_0})$ is the task-related degradation model that describes the relationship between desirable image $\mathbf{X}$ and degraded images $\mathbf{X_0}$. $\mathcal{R}(\mathbf{X})$ symbolizes the regularization term that needs careful handcrafting in the traditional methods. However, DIP [80] states that the network structure itself can serve as a powerful prior, which allows the removal of the regularization term and leads to the following equation:

$$\min_{\mathbf{X}} E(\mathbf{G}_\theta(\mathbf{E}), \mathbf{X_0}) \tag{2}$$

where $\mathbf{G}_\theta(\mathbf{E})$ is the designed network with parameters θ and random noise input $\mathbf{E}$. Under this premise, DIP is an unsupervised technology since the target image can be restored only with the degraded images.

B. Hyperspectral Image Super-Resolution

HISR is an ill-posed inverse problem, whose purpose is to estimate latent HrHSI $\mathcal{X} \in \mathbb{R}^{H\times W\times C}$ with high spatial–spectral resolution by fusing observed LrHSI $\mathcal{Y} \in \mathbb{R}^{h\times w\times C}$ and HrMSI $\mathcal{Z} \in \mathbb{R}^{H\times W\times c}$. Here, $H/h, W/w$, and C/c represent the height, width, and band number of corresponding images, where $w \ll W, h \ll H$, and $c \ll C$ with $r = H/h = W/w$ being the scale factor. Since LrHSI and HrMSI are the spatial and spectral degradation outcomes from HrHSI, their relationship $E(\mathbf{X}, \mathbf{X_0})$ can be mathematically established by the degradation model

$$\begin{aligned}\mathbf{Y} &= \mathbf{PX} + \mathbf{N}_y\\ \mathbf{Z} &= \mathbf{XS} + \mathbf{N}_z\end{aligned} \tag{3}$$

where $\mathbf{X} \in \mathbb{R}^{HW\times C}$, $\mathbf{Y} \in \mathbb{R}^{hw\times B}$, and $\mathbf{Z} \in \mathbb{R}^{HW\times b}$ are unfolded matrix version of $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively. $\mathbf{P} \in \mathbb{R}^{wh\times WH}$ represents the spatial blurring with point spread function (PSF) kernel and a downsampling operator. $\mathbf{S} \in \mathbb{R}^{B\times c}$ is the SRF responsible for the spectral degradation. $\mathbf{N}_y$ and $\mathbf{N}_z$ are independent Gaussian noises.

Similarly, the reconstruction of unknown HrHSI from observed HrMSI-LrHSI pair can be also formulated as the following optimization problem:

$$\min_{\mathbf{X}} \|\mathbf{Z} - \mathbf{XS}\|_F^2 + \|\mathbf{Y} - \mathbf{PX}\|_F^2 + \mathcal{R}(\mathbf{X}) \tag{4}$$

where $||\cdot||_F$ denotes the Frobenius norm. The first and second terms are called the data-fidelity term, and the last term is the prior term. The data terms can harness the complementary information from both LrHSI and HrMSI and ensure that the generated HrHSI is as similar as possible to the original LrHSI and HrMSI. However, designing a proper regularization term is still an open problem. HSIs often exhibit a wide range of structures, including edges, textures, and patterns, which are often complex and varied. Moreover, these handcrafted priors show limitations in capturing rich image statistics and, hence, leave considerable room for further improvement. Under the power of DIP, we substitute a well-designed network $\mathbf{G}_\theta(\cdot)$ for the explicit regularizer $\mathcal{R}(\cdot)$, which leads to the following formulation:

$$\begin{aligned}&\min_{\theta} \|\mathbf{Z} - \mathbf{XS}\|_F^2 + \|\mathbf{Y} - \mathbf{PX}\|_F^2\\ &\text{s.t. } \mathbf{X} = \mathbf{G}_\theta(\mathbf{E}).\end{aligned} \tag{5}$$

Therefore, the HISR task is converted to finding the optimal parameters θ instead of manually constructing explicit prior terms. Moreover, we only need an observed HrMSI-LrHSI pair to generate an unknown HrHSI, which is purely an unsupervised framework.

However, there are still two problems lying in front of DIP. One is that many works attempt to generate HrHSI directly from uninformed random noise $\mathbf{E}$. Despite the achievement made in natural images, the complicated spatial–spectral structure in latent HrHSI renders this task more challenging. Another is that the particularity of HISR is not fully considered when designing the structure of $\mathbf{G}_\theta(\cdot)$. In fact, different network architectures lead to diverse regularization results due to the preference of specific tasks. Considering those issues, we propose M2U-Net to enhance the DIP performance for the task of HISR. For the convenience of readers, the main abbreviations used in this article are listed in Nomenclature.

IV. Methodology

A. Three-Stage Pipeline

By leveraging both DIP and degradation model information, we propose M2U-Net to achieve unsupervised HISR. As shown in Fig. 1, our framework adopts a three-stage pipeline, including DIL, IIE, and DIG. In the first stage, DIL-Net is proposed to derive the required degradation information for the following two stages, including two LrMSIs, SRF, and PSF parameters. Then, we design SML-Net to establish the spectral relationship and simultaneously generate an initialized input with rich HrHSI-relevant information for stage three, aiming to facilitate the extraction of prior information and promote the potential of DIP for the HISR task. Finally, Dual U-Net with the CAG module is elaborately designed to recover latent HrHSI. It is worth noting that the overall pipeline is purely performed on observed HrMSI-LrHSI pair without any extra training datasets.

B. Degradation Information Learning

In this section, we design DIL-Net to exploit the deep information of the degradation model and then utilize it to guide the following two stages. Following (3), a latent spatial–spectral correlation of the to-be-fused HrMSI-LrHSI pair can be established as follows:

$$\mathbf{K}_1 = \mathbf{YS}, \quad \mathbf{K}_2 = \mathbf{PZ}, \;\; \mathbf{K}_1 = \mathbf{K}_2. \tag{6}$$

Here, two LrMSIs, i.e., $\mathbf{K}_1 \in \mathbb{R}^{hw\times b}$ and $\mathbf{K}_2 \in \mathbb{R}^{hw\times b}$, are obtained by spectrally and spatially downsampling LrHSI and

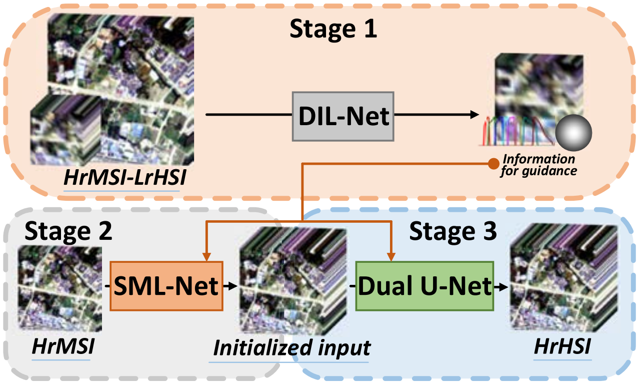

Fig. 1. Three-stage pipeline of M2U-Net.

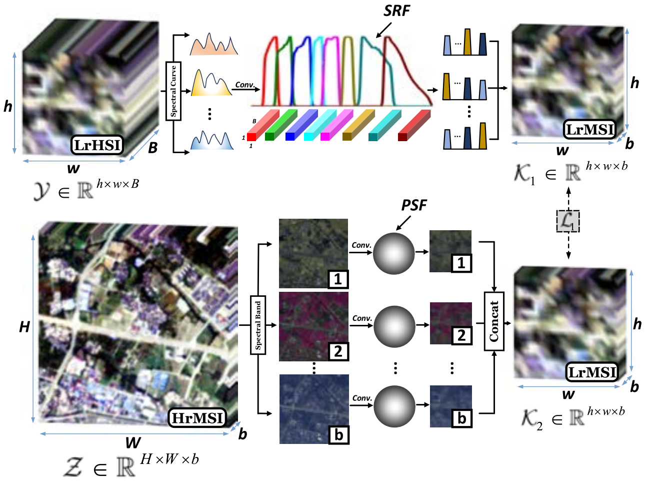
SRF
Spectral Curve
Conv.
LrHSI
h
w
B
$\mathcal{Y} \in \mathbb{R}^{h\times w\times B}$
LrMSI
b
$\mathcal{K}_1 \in \mathbb{R}^{h\times w\times b}$
$\mathcal{L}_1$
PSF
Spectral Band
Conv.
1
2
b
Concat
HrMSI
H
W
$\mathcal{Z} \in \mathbb{R}^{H\times W\times b}$
$\mathcal{K}_2 \in \mathbb{R}^{h\times w\times b}$

Fig. 2. Detailed structure of DIL-Net.

HrMSI, respectively. Under this premise, we directly construct DIL-Net to simulate the process in (6) and derive our needed model information.

As depicted in Fig. 2, the degradation process represented by $\mathbf{P}$ and $\mathbf{S}$ can be modeled by the convolutional operations. To be specific, the spatial degradation, as shown in the bottom of Fig. 2, is conducted for each spectral band independently with a shared PSF kernel followed by a downsampling operator. Hence, it can be perfectly modeled by a depthwise convolutional layer with trainable parameter $\theta^{(\mathrm{PSF})}$ and stride r, which is formulated as

$$\mathcal{K}_2 = \mathcal{F}_z\big(\mathcal{Z}; \theta^{(\mathrm{PSF})}\big). \tag{7}$$

Here, the spatial degradation is simplified by $\mathcal{F}_z(\cdot; \theta^{(\mathrm{PSF})})$ with $\theta^{(\mathrm{PSF})}$ to parameterize the PSF kernel. From the perspective of spectral degradation, SRF denotes the response of sensors at a specific wavelength, which symbolizes the integral operation along the channel dimension. In discrete settings, it can be further transformed into a weighted summation for each spectral curve. Therefore, a pointwise convolutional layer with learnable parameters $\theta^{(\mathrm{SRF})}$ is employed to model the transformation, as shown in the top of Fig. 2. This process can be expressed by

$$\mathcal{K}_1 = \mathcal{F}_y\big(\mathcal{Y}; \theta^{(\mathrm{SRF})}\big) \tag{8}$$

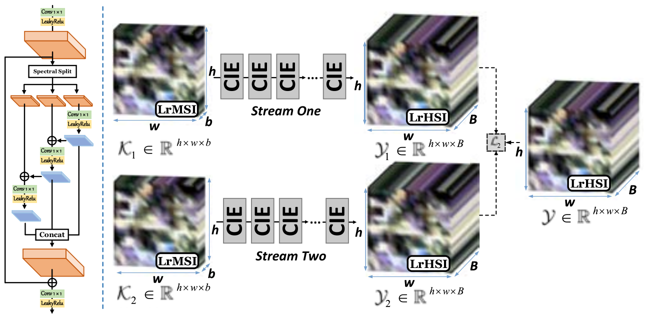
Conv 1×1
LeakyRelu
Spectral Split
Concat
CIE
Stream One
Stream Two
LrMSI
LrHSI
h
w
b
B
$\mathcal{K}_1 \in \mathbb{R}^{h\times w\times b}$
$\mathcal{K}_2 \in \mathbb{R}^{h\times w\times b}$
$\mathcal{Y}_1' \in \mathbb{R}^{h\times w\times B}$
$\mathcal{Y}_2' \in \mathbb{R}^{h\times w\times B}$
$\mathcal{L}_2$
$\mathcal{Y} \in \mathbb{R}^{h\times w\times B}$

Fig. 3. Detailed structure of SML-Net. (Left) Architecture of the CIE module.

where $\mathcal{F}_y(\cdot; \theta^{(\mathrm{SRF})})$ denotes the spectral degradation and $\theta^{(\mathrm{SRF})}$ is the SRF parameter.

Given two LrMSIs, we can construct the loss function following (6):

$$\mathcal{L}_1 = \|\mathcal{K}_1 - \mathcal{K}_2\|_1. \tag{9}$$

$\mathcal{K}_1$ and $\mathcal{K}_2$ are tensor version of $\mathbf{K}_1$ and $\mathbf{K}_2$. Since estimating the unknown SRF and PSF is an ill-posed problem, we impose reasonable constraints on these parameters by considering their physical properties, i.e., nonnegativity and sum-to-one. Specifically, we apply clamp activation in each iteration to guarantee the nonnegativity. Besides, a normalization operator is also utilized to satisfy the sum-to-one constraint. Moreover, the target HrHSI typically comprises intricate objects varying in size and content, showcasing a wealth of spatial information including texture and edges. Nevertheless, the outcomes derived from the $L2$ loss method often suffer from a deficiency in high-frequency elements with overly smooth textures, yielding unsatisfactory images. In contrast, $L1$ loss tends to recover subtle and sharp details and, hence, is chosen as the loss criterion in our method.

C. Initialized Image Establishment

As mentioned earlier, we plan to harness the power of DIP to capture the image statistics of latent HrHSI. Despite the achievement made in natural images, it is still a challenging task to reconstruct high-quality HrHSI from uninformed random noise. Thus, we plan to generate an initialized input with expressive HrHSI-relevant information for DIP, aiming to facilitate the extraction of image statistics and promote its potential for better image recovery. Considering that HrHSI is unavailable, we generate a coarse estimation from the observed HrMSI-LrHSI pair with the assistance of stage one.

As can be seen in (3) and (6), two LrMSIs $\mathbf{K}_1$&$\mathbf{K}_2$ and HrMSI $\mathbf{Z}$ are spectrally degraded from LrHSI $\mathbf{Y}$ and unknown HrHSI $\mathbf{Z}$ via the same spectral degradation matrix $\mathbf{S}$, respectively. In other words, if we can establish the spectral mapping from $\mathbf{K}_1$&$\mathbf{K}_2$ to $\mathbf{Y}$ and apply this relationship to observed $\mathbf{Z}$, we can generate a coarse estimation with sufficient HrHSI-relevant information. With this observation, SML-Net is designed to achieve the spectral mapping from LrMSIs to LrHSI, as shown in Fig. 3. Considering that two LrMSIs are of equal importance to this task, a two-stream structure is chosen as the backbone, enabling our network to fully utilize their information and obtain a more expressive HrHSI-relevant image. Based on this, a series of CIE modules are cascaded to constitute the main part of SML-Net, aiming to realize a better spectral reconstruction. As depicted on the left of Fig. 3, the input first passes through a convolution layer with a LeakyReLu unit, transforming their channel into the desired number. Then, output feature maps are equally divided into three parts along the channel dimension. Before further operations, each part will receive the output of its preceding part via pixelwise summation. With the information passing through different parts, the interaction across different channels is gradually enhanced and their correlations are further established. After another convolution and LeakyRelu, the outputs from the three parts are concatenated together to recover the original size. Moreover, we further add a skip connection to bridge the output maps with the input, which can facilitate feature reuse and training convergence. Finally, we use another convolution with LeakyRelu to double the number of feature maps and gradually approximate the desired LrHSI. The reconstruction loss for optimizing SML-Net is given as follows:

$$\begin{cases} \mathcal{Y}_1, \mathcal{Y}_2 = \mathcal{F}_s(\mathcal{K}_1, \mathcal{K}_2; \theta^{(\mathrm{SML})}) \\ \mathcal{L}_2 = \|\mathcal{Y} - \mathcal{Y}_1\|_1 + \|\mathcal{Y} - \mathcal{Y}_2\|_1 \end{cases} \tag{10}$$

where $\mathcal{F}_s(\cdot; \theta^{(\mathrm{SML})})$ denotes SML-Net with trainable parameters $\theta^{(\mathrm{SML})}$.

With the well-trained $\mathcal{F}_s(\cdot; \theta^{(\mathrm{SML})})$, the initialized input can be obtained by

$$\begin{cases} \mathcal{X}_{s1}, \mathcal{X}_{s2} = \mathcal{F}_s(\mathcal{Z}, \mathcal{Z}; \theta^{(\mathrm{SML})}) \\ \mathcal{X}_{\mathrm{input}} = \frac{1}{2}\mathcal{X}_{s1} + \frac{1}{2}\mathcal{X}_{s2}. \end{cases} \tag{11}$$

$\mathcal{X}_{s1}$ and $\mathcal{X}_{s2}$ are two coarse estimations of HrHSI, and we obtain the final initialized input $\mathcal{X}_{\mathrm{input}}$ by an average operation. Despite the inevitable error during the learning process, the generated $\mathcal{X}_{\mathrm{input}}$ still provides sufficient HrHSI-relevant statistics compared with the random noisy input, which will be of benefit to DIP.

D. Deep Image Generation

Recent works [81], [82] have demonstrated that different network structures can lead to diverse priors, hence influencing the final reconstruction quality. On the other hand, U-Net, as a popular backbone, has shown great potential in the task of HISR [69]. Considering the aforementioned factors, we aim to construct our U-Net-based generator network to extract deep prior of the to-be-recovered HrHSI from the initialized $\mathcal{X}_{\mathrm{input}}$. Moreover, it is shown that rich spatial information from HrMSI can contribute to a powerful regularizer and, hence, further improve the fusion performance. Under this premise, we design Dual U-Net to fully capture the spatial–spectral prior of latent HrHSI for a high-quality reconstruction. As shown in Fig. 4, our proposed Dual U-Net mainly consists of one auxiliary U-Net and one principal U-Net, which are closely coupled through the CAG module. Overall, the auxiliary U-Net is responsible for extracting the multiscale spatial features of HrMSI for guidance, and the principal U-Net absorbs this information via the CAG module to gradually recover HrHSI.

To be specific, Dua U-Net retains the classic architecture proposed in [83]. The left-hand side of Fig. 4 is the down-sampling path, which progressively narrows down the spatial resolution of the initialized image $\mathcal{X}_{\mathrm{input}}$ and HrMSI $\mathcal{Z}$ with the repeated application of the CAG modules and pooling operators. The formulation of this path can be expressed as

$$\begin{cases} \mathcal{X}_1, \mathcal{Z}_1 = \mathbf{CAG}(\mathcal{X}_{\mathrm{input}}, \mathcal{Z}) \\ \mathcal{X}_2 = \mathcal{X}_1 \downarrow \\ \mathcal{Z}_2 = \mathcal{Z}_1 \downarrow \end{cases} \tag{12}$$

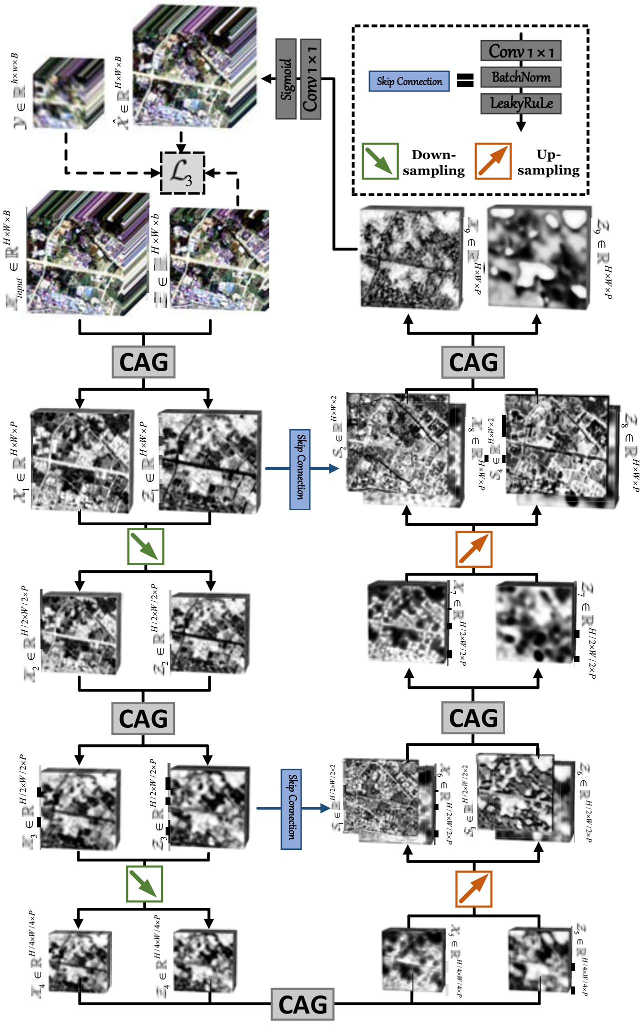
$\mathcal{Y} \in \mathbb{R}^{h\times w\times B}$
$\hat{\mathcal{X}} \in \mathbb{R}^{H\times W\times B}$
Sigmoid
Conv 1 × 1
Skip Connection
Conv 1 × 1
BatchNorm
LeakyRuLe
Down-sampling
Up-sampling
$\mathcal{L}_3$
$\mathcal{X}_{input} \in \mathbb{R}^{H\times W\times B}$
$\mathcal{Z} \in \mathbb{R}^{H\times W\times b}$
$\mathcal{X}_9 \in \mathbb{R}^{H\times W\times P}$
$\mathcal{Z}_9 \in \mathbb{R}^{H\times W\times P}$
CAG
$\mathcal{X}_1 \in \mathbb{R}^{H\times W\times P}$
$\mathcal{Z}_1 \in \mathbb{R}^{H\times W\times P}$
Skip Connection
$\mathcal{S}_2 \in \mathbb{R}^{H\times W\times 2}$
$\mathcal{X}_8 \in \mathbb{R}^{H\times W\times P}$
$\mathcal{S}_4 \in \mathbb{R}^{H\times W\times 2}$
$\mathcal{Z}_8 \in \mathbb{R}^{H\times W\times P}$
$\mathcal{X}_2 \in \mathbb{R}^{H/2\times W/2\times P}$
$\mathcal{Z}_2 \in \mathbb{R}^{H/2\times W/2\times P}$
$\mathcal{X}_7 \in \mathbb{R}^{H/2\times W/2\times P}$
$\mathcal{Z}_7 \in \mathbb{R}^{H/2\times W/2\times P}$
CAG
$\mathcal{X}_3 \in \mathbb{R}^{H/2\times W/2\times P}$
$\mathcal{Z}_3 \in \mathbb{R}^{H/2\times W/2\times P}$
Skip Connection
$\mathcal{S}_1 \in \mathbb{R}^{H/2\times W/2\times 2}$
$\mathcal{X}_6 \in \mathbb{R}^{H/2\times W/2\times P}$
$\mathcal{S}_3 \in \mathbb{R}^{H/2\times W/2\times 2}$
$\mathcal{Z}_6 \in \mathbb{R}^{H/2\times W/2\times P}$
$\mathcal{X}_4 \in \mathbb{R}^{H/4\times W/4\times P}$
$\mathcal{Z}_4 \in \mathbb{R}^{H/4\times W/4\times P}$
$\mathcal{X}_5 \in \mathbb{R}^{H/4\times W/4\times P}$
$\mathcal{Z}_5 \in \mathbb{R}^{H/4\times W/4\times P}$
CAG

Fig. 4. Detailed structure of Dual U-Net.

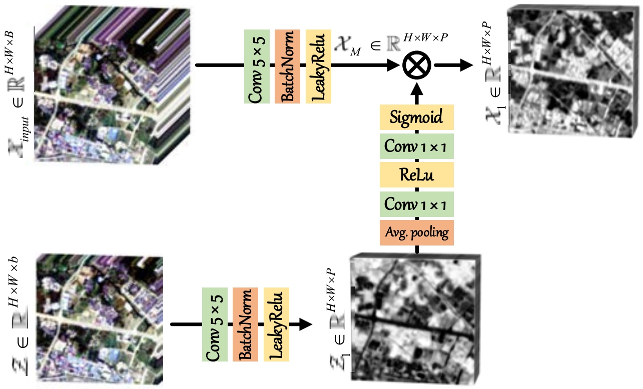
$\mathcal{X}_{input} \in \mathbb{R}^{H\times W\times B}$
Conv 5 × 5
BatchNorm
LeakyRelu
$\mathcal{X}_M \in \mathbb{R}^{H\times W\times P}$
$\mathcal{X}_1 \in \mathbb{R}^{H\times W\times P}$
Sigmoid
Conv 1 × 1
ReLu
Conv 1 × 1
Avg. pooling
$\mathcal{Z} \in \mathbb{R}^{H\times W\times b}$
Conv 5 × 5
BatchNorm
LeakyRelu
$\mathcal{Z}_1 \in \mathbb{R}^{H\times W\times P}$

Fig. 5. Detailed structure of the CAG module.

and

$$\begin{cases} \mathcal{X}_3, \mathcal{Z}_3 = \textbf{CAG}(\mathcal{X}_2, \mathcal{Z}_2) \\ \mathcal{X}_4 = \mathcal{X}_3 \downarrow \\ \mathcal{Z}_4 = \mathcal{Z}_3 \downarrow \end{cases} \tag{13}$$

where ↓ denotes the average pooling that realizes the spatial reduction with a scale factor of four. Note that the number of feature maps produced in Dual U-Net is set to P, and their size is marked next to themselves. The most important part of the Dual U-Net is the CAG module, which transforms the spatial information from the auxiliary U-Net to the corresponding stages in the principal U-Net via a cross-attention mechanism. Without loss of generality, we take $\mathcal{X}_{\text{input}}$ and $\mathcal{Z}$ as an example to demonstrate the working mechanism, as shown in Fig. 5. The top and bottom parts of the CAG module share the same architecture, i.e., a convolution layer followed by batch normalization and LeakyReLu, leading to the feature $\mathcal{Z}_1$ and $\mathcal{X}_M$. Considering the importance of spatial details for image recovery, we encapsulate abundant spatial information of $\mathcal{Z}_1$ into an attention map. Then, the derived weights are imposed on the corresponding outputs of the principal U-Nets, enabling these spatial features to guide the reconstruction process. Different from [82], Uezato et al. utilize two tailor-designed modules to separately deal with the features from the encoder and decoder parts. By contrast, the CAG module is applied to all features of the same scale. Moreover, we remove the upsample operator since we feed the initialized $\mathcal{X}_{\text{input}}$ with the same scale of HrHSI as the input of Dual U-Net, which may reduce unnecessary uncertainty in the optimization process.

The right-hand side of Fig. 4 is the upsampling path that gradually recovers the spatial size to approximate desired HrHSI, which is formulated as

$$\begin{cases} \mathcal{X}_5, \mathcal{Z}_5 = \textbf{CAG}(\mathcal{X}_4, \mathcal{Z}_4) \\ \mathcal{X}_6 = \mathcal{X}_5 \uparrow \\ \mathcal{Z}_6 = \mathcal{Z}_5 \uparrow \end{cases} \tag{14}$$

and

$$\begin{cases} \mathcal{X}_7, \mathcal{Z}_7 = \textbf{CAG}([\mathcal{S}_1, \mathcal{X}_6], [\mathcal{S}_3, \mathcal{Z}_6]) \\ \mathcal{X}_8 = \mathcal{X}_7 \uparrow \\ \mathcal{Z}_8 = \mathcal{Z}_7 \uparrow \\ \mathcal{X}_9, \mathcal{Z}_9 = \textbf{CAG}([\mathcal{S}_2, \mathcal{X}_8], [\mathcal{S}_2, \mathcal{Z}_8]) \end{cases} \tag{15}$$

where [,] denotes the concatenation operator along the spectral dimension and ↑ represents bilinear interpolation to achieve spatial upsampling. $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_4$ are generated from the downsampling path via an operation of 1×1 convolution, batch normalization, and LeakyReLu, enabling the upsampling path to reuse the detail information from the low-level features. Finally, we employ a 1×1 convolution along with the sigmoid unit to transform the feature map to the desired number and obtain the final output $\hat{\mathcal{X}}$.

Since there is no available GT for network optimization, we directly resort to (5) and incorporate the observed HrMSI-LrHSI to formulate the loss function

$$\mathcal{L}_3 = \left\| \mathcal{Z} - \mathcal{F}_z(\hat{\mathcal{X}}; \theta^{(\text{SRF})}) \right\|_1 + \left\| \mathcal{Y} - \mathcal{F}_y(\hat{\mathcal{X}}; \theta^{(\text{PSF})}) \right\|_1. \tag{16}$$

This model-constrained loss indicates that the spectral and spatial degraded versions of generated HrHSI $\hat{\mathcal{X}}$ should approximate to HrMSI $\mathcal{Z}$ and LrHSI $\mathcal{Y}$, respectively. Since the parameters of SRF and PSF are usually unavailable, we directly employ the degradation process learned in stage one to achieve the two transformations. In other words, our method can realize the HISR task in a blind manner.

TABLE I
SIMULATION RESULTS OF FOUR DATASETS

	Houston	Washington DC Mall	TianGong-1	Chikusei
Size of HrHSI	400 × 400 × 46	300 × 300 × 191	240 × 240 × 54	400 × 400 × 110
Spectral range of HrHSI (nm)	380 - 1050	401 – 2473	413 - 887	363 - 1018
Scale factor	8	10	12	16
Size of LrHSI	50 × 50 × 46	30 × 30 × 191	20 × 20 × 54	25 × 25 × 110
Size of HrMSI	400 × 400 × 8	300 × 300 × 8	240 × 240 × 8	400 × 400 × 8

V. Experiments and Results

In this section, extensive experiments, including parameter discussion, ablation study, and SR performance, are provided to quantitatively and visually demonstrate our effectiveness.

A. Datasets and Setups

1) Datasets: In our study, the performance of all methods is evaluated on four HSIs, including the University of Houston, Washington DC Mall, TianGong-1, and Chikusei datasets. Specifically, Houston datasets acquired by the ITRES CASI 1500 camera provide 48 spectral bands with a spectral range from 380 to 1050 nm. The Washington DC Mall dataset is captured by the HYDICE sensor with a total of 210 bands. The TianGong-1 dataset is produced for the task of scene classification and covers the range from visible to short-wave infrared. The Chikusei dataset is taken by Headwall Hyperspec-VNIR-C imaging sensor and has 128 bands from 363 to 1018 nm.

Considering the unavailability of HrHSI in real scenarios, a simulation experiment is conducted by treating the abovementioned datasets as in [84]. In this process, LrHSI is generated by degrading the reference with a Gaussian blurring kernel followed by a downsampling operator, and HrMSI is produced using multispectral SRF of WorldView 2. The specifications of simulated results are provided in Table I.

2) Benchmark: To assess the performance of M2U-Net, we choose nine competitors from traditional and DL-based categories, where the former one includes G-SOMP+ [85], CSU [86], CNMF [87], STEREO [88], CSTF [43], and SCOTT [89], and the latter one includes MIAE [90], SURE [69], and XINet [62]. We select six metrics to quantitatively evaluate the fusion performance in terms of spatial, spectral, and global qualities. Specifically, correlation coefficient (CC), root mean square error (RMSE), SNR [91], and universal image quality index (UIQI) [92], [93] are computed bandwise to evaluate the spatial reconstruction quality. For PSNR, UIQI, and CC, a higher value indicates better fusion performance, while the ideal value for RMSE is zero. SAM is computed along each spectrum vector to measure the spectral similarity with the best value at zero. Relative dimensionless global error in synthesis (ERGAS) [94] is a global indication of the fusion quality, and the ideal value is at zero. For visual evaluation, we use the SAM heatmap, the mean relative absolute error (MRAE) heatmap, and the residual heatmap to exhibit their differences.

TABLE II

QUANTITATIVE METRICS WITH DIFFERENT NUMBER OF P IN THE TIANGONG-1 DATASET

P	SAM	PSNR	ERGAS	CC	RMSE	UIQI
40	1.5355	40.5964	0.2598	0.9917	0.0080	0.9994
60	1.2161	42.4483	0.2059	0.9947	0.0063	0.9996
80	1.1168	43.3147	0.1935	0.9955	0.0058	0.9996
100	1.0416	43.8022	0.1754	0.9960	0.0054	0.9997
120	0.9931	44.3801	0.1647	0.9964	0.0051	0.9997
140	0.9843	44.4626	0.1638	0.9965	0.0050	0.9997
160	0.9693	44.7003	0.1596	0.9967	0.0049	0.9997
180	0.9172	45.1452	0.1502	0.9969	0.0047	0.9998
200	0.8674	45.5849	0.1414	0.9972	0.0044	0.9998
220	0.8863	45.4522	0.1437	0.9971	0.0045	0.9998
240	0.8814	45.5398	0.1440	0.9971	0.0045	0.9998
260	0.8431	45.9279	0.1387	0.9973	0.0043	0.9998
280	0.8593	45.6966	0.1421	0.9973	0.0044	0.9998
300	0.9180	45.3750	0.1544	0.9970	0.0047	0.9997
320	0.8484	45.8035	0.1391	0.9973	0.0043	0.9998
340	0.8439	45.9655	0.1387	0.9973	0.0043	0.9998

3) Implementation Details: Our proposed M2U-Net is optimized in a three-stage pipeline, as shown in Algorithm 1. To achieve a stable optimization, the linear decay strategy is adopted for the Adam optimizer. Specifically, the learning rate is initially fixed to 0.001, 0.004, and 0.004 for the three stages, respectively, and then linearly decreases to zero until the given epochs. The optimization is implemented under the PyTorch framework with four GeForce GTX 1080 Ti.

Algorithm 1 Proposed M2U-Net Algorithm

```
Input: Observed LrHSI $\mathbf{Y}$ and HrMSI $\mathbf{Z}$.
Procedure:
Stage1: DIL:
    a) Optimizing $\mathcal{L}_1$;
    b) Obtain two LrMSI $\mathbf{K}_1$ & $\mathbf{K}_2$, and degradation parameters $\theta^{(PSF)}$ and $\theta^{(SRF)}$
Stage2: IIE:
    a) Optimizing $\mathcal{L}_2$ under the guidance of $\mathbf{K}_1$ & $\mathbf{K}_2$;
    b) Obtain the initialized input $\mathcal{X}_{\text{input}}$
Stage3: Deep image generation:
    a) Optimizing $\mathcal{L}_3$ under the guidance of $\theta^{(PSF)}$ and $\theta^{(SRF)}$;
    b) Obtain the final output $\hat{\mathcal{X}}$
End Procedure
```

B. Parameters' Discussion

As mentioned earlier, the number of feature channels in each layer of Dual U-Net is set to P, which serves as an important parameter to influence the final reconstruction result. To analyze its effect on the SR performance, we experiment with the TianGong-1 dataset by setting P from 40 to 340 with a step of 20, and the corresponding results are shown in Table II. It can be observed that the performance is restricted under a small number of feature channels. With the increase in P, a noticeable improvement occurs in terms of all metrics because more spatial–spectral priors can be modeled. After the number of 260, the performance stays relatively stable with a little fluctuation. Considering the growing parameter burden caused by the rise of P, we balance the performance and the network parameter and set P to 260 for the TianGong-1 dataset. The same procedure is conducted for the other three datasets.

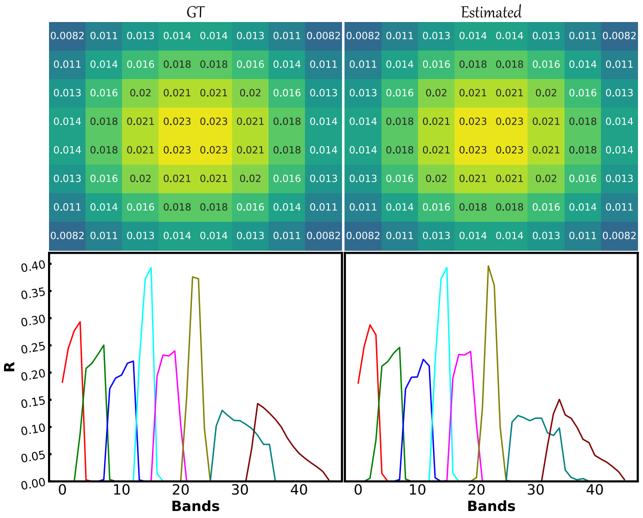

Fig. 6. GT and estimated PSF and SRF parameters in the Houston dataset.

C. Learned Degradation Information

We design DIL-Net in stage one to learn the deep degradation information, including two LrMSI $\mathcal{K}_1$ & $\mathcal{K}_2$, PSF, and SRF parameters, which serve as guidance for the other two stages. To verify their accuracy, we choose the Houston dataset as an example to visually exhibit their results. As shown in Fig. 6, the first column depicts the GT PSF and SRF, and the second column shows the corresponding estimated results. It can be observed that we achieve a precise estimation of PSF. Despite the error shown in the SRF, it still exhibits a similar curve with slight fluctuations. Besides, we illustrate the learned LrMSIs in Fig. 7 to further demonstrate the above conclusion. In the first row, $\mathcal{K}_1$ and $\mathcal{K}_2$ exhibit close spatial structure to GT with high PSNR values. The MRAE heatmap in the second row is magnified by a factor of 30 to display their error. Though two LrMSIs both achieve high accuracy, $\mathcal{K}_2$ generated from PSF is more accurate than $\mathcal{K}_1$ generated from SRF, showing consistent results with Fig. 6.

D. Ablation Study

Our proposed M2U-Net consists of three key nets to separately realize a three-stage pipeline. Hence, we implement extensive ablation experiments to verify their influence in the Chikusei dataset.

1) Influence of DIL-Net: PSF and SRF, serving as important parameters in the degradation model, are often treated as known priors to assist the HISR task. Considering their unavailability in real applications, it is not desirable to make such an assumption. Luckily, we construct DIL-Net to exploit

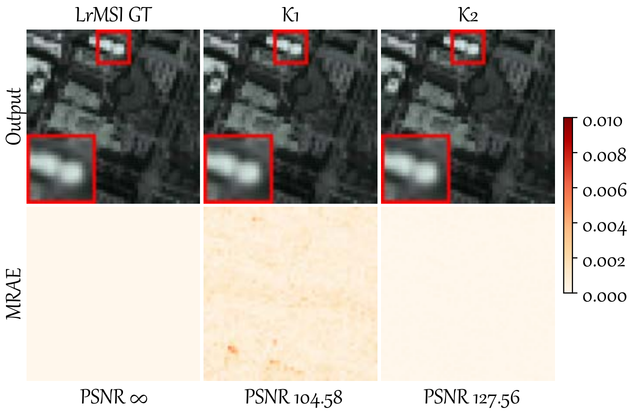

Fig. 7. GT LrMSI and two learned LrMSI $\mathcal{K}_1$ and $\mathcal{K}_2$, along with the quality measures, in the Houston dataset. For better visual quality, we magnify MRAE by a factor of 30.

TABLE III
ABLATION STUDY OF DIL-NET IN THE CHIKUSEI DATASET

Datasets	Configuration	SAM	PSNR	ERGAS	CC	RMSE	UIQI
Chikusei	blind	0.8687	49.7405	0.2783	0.9969	0.0040	0.9993
	non-blind	0.8685	49.8373	0.3067	0.9965	0.0039	0.9992

the deep information of the degradation model, in which PSF and SRF parameters are learned from the observed HrMSI-LrHSI pair. To verify the effectiveness of DIL-Net, we simply remove the first stage and provide M2U-Net with GT PSF and SRF. The blind and nonblind results in the Chikusei dataset are summarized in Table III. It can be seen that M2U-Net with accurate PSF and SRF can obtain better results in terms of SAM, PSNR, and RMSE. However, the improvement is not significant, and the blind version even achieves slightly better outcomes in other metrics, which demonstrates the effectiveness of DIL-Net.

2) *Influence of SML-Net:* In the stage of IIE, we design a two-stream SML to fully utilize the information of $\mathcal{K}_1$ and $\mathcal{K}_2$, aiming to generate an image with much more expressive HrHSI-relevant statistics by leveraging the established spectral mapping. To demonstrate the effect of this structure, we propose two variants by separately removing each stream, named stream one and stream two, respectively. As can be seen in Table IV, two variants with only one stream give relatively poor performance compared with the complete structure, in which stream one suffers more degradation in terms of all metrics. Therefore, the two-stream structure can bring about improvements in the SR result. Moreover, we show the quantitative evaluation of middle outcomes in the Chikusei dataset, as shown in Table V. It can be observed that $\mathcal{X}_{s1}$ and $\mathcal{X}_{s2}$ are the coarse estimation of the target HrHSI compared with other competitors, but they can still provide sufficient HrHSI-relevant information compared with the random noise, which will facilitate the prior extraction of DIP and promote the fusion performance.

3) *Influence of Dual U-Net:* Dual U-Net is designed in the last stage to capture the image statistic with the initialized image $\mathcal{X}_{\text{input}}$ and HrMSI $\mathcal{Z}$ as the inputs. In this part, we will evaluate the effectiveness in terms of the structure and the input, respectively.

TABLE IV
ABLATION STUDY OF SML-NET IN THE CHIKUSEI DATASET

Datasets	Configuration	SAM	PSNR	ERGAS	CC	RMSE	UIQI
Chikusei	stream one	0.9248	49.0522	0.3079	0.9966	0.0043	0.9992
	stream two	0.9165	49.2854	0.3003	0.9966	0.0042	0.9992
	complete	0.8687	49.7405	0.2783	0.9969	0.0040	0.9993

TABLE V
QUANTITATIVE EVALUATION OF MIDDLE OUTCOMES IN THE CHIKUSEI DATASET

Datasets	Name	SAM	PSNR	ERGAS	CC	RMSE	UIQI
Chikusei	$\mathcal{X}_{s1}$	1.5185	41.6249	0.8841	0.9867	0.0078	0.9922
	$\mathcal{X}_{s2}$	1.1040	42.1037	0.7852	0.9877	0.0072	0.9962
	$\mathcal{X}_{input}$	1.1383	43.8104	0.6802	0.9906	0.0061	0.9962
	$\hat{\mathcal{X}}$	0.8687	49.7405	0.2783	0.9969	0.0040	0.9993

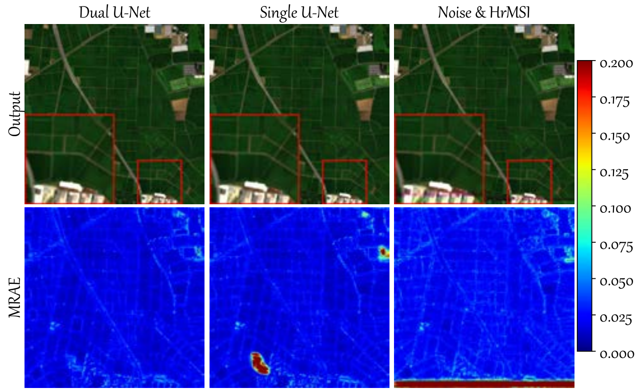

Fig. 8. Outputs and MARE heatmaps under different configurations.

TABLE VI
ABLATION STUDY OF DUAL U-NET IN THE CHIKUSEI DATASET

	Configuration	SAM	PSNR	ERGAS	CC	RMSE	UIQI
Structure	Single U-Net	1.0904	47.1172	0.3705	0.9928	0.0069	0.9982
	Dual U-Net	0.8687	49.7405	0.2783	0.9969	0.0040	0.9993
Input	noise & $\mathcal{Z}$	2.5476	38.8398	0.8123	0.9223	0.0345	0.9930
	$\mathcal{X}_{input}$ & $\mathcal{Z}$	0.8687	49.7405	0.2783	0.9969	0.0040	0.9993

a) Structure: As shown in Table VI, we remove the auxiliary U-Net and only retain the principal U-Net to extract the prior information of $\mathcal{X}_{\text{input}}$. Compared with Dual U-Net, the performance of the single U-Net drops sharply, which firmly demonstrates that the incorporation of spatial information from HrMSI can benefit the capture of image statistics and, hence, further promote the SR results. Moreover, Fig. 8 shows that the loss of spatial guidance from HrMSI leads to serious distortions in some local areas. Differently, Dual U-Net can better recover the spatial structure with fewer distortions under the guidance of HrMSI.

b) Input: To verify the effectiveness of the initialized $\mathcal{X}_{\text{input}}$, we replace it with a random noise. It can be seen in Table VI that there is a significant decline in all metrics compared with $\mathcal{X}_{\text{input}}$ since a random noise contains limited knowledge about the to-be-recovered image and, hence, hampers the capture of image priors. However, the generated $\mathcal{X}_{\text{input}}$ with sufficient HrHSI-relevant information can promote the

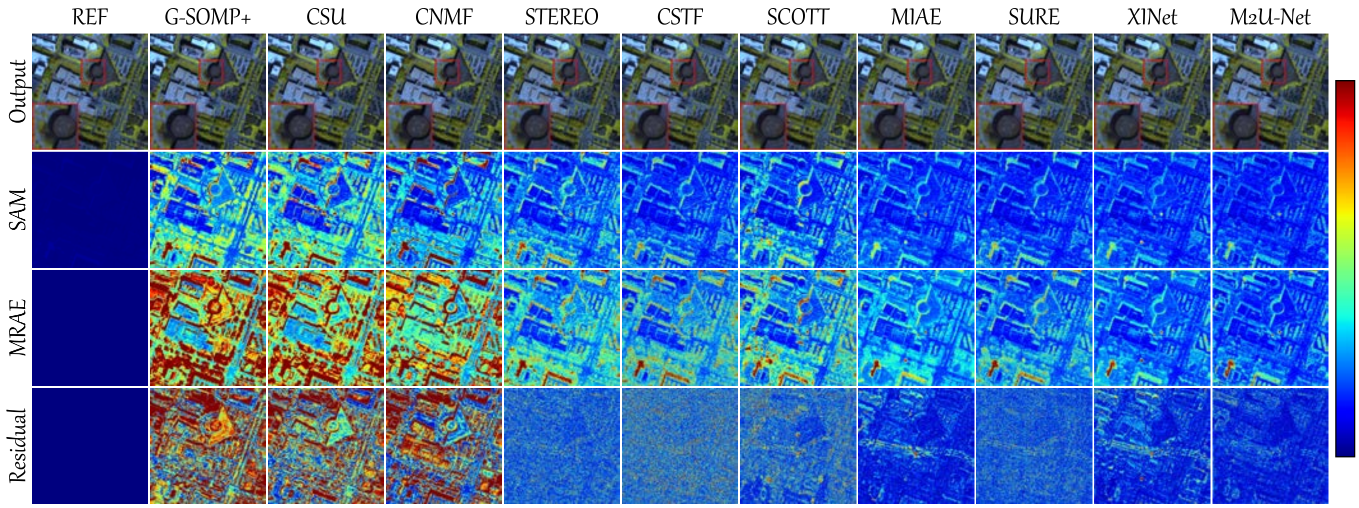

Fig. 9. Visual comparison of all methods in the Houston dataset. First row: pseudocolor image of SR outputs (R:46, G:30, B:14). Second row: the heatmap of SAM error. Third row: the heatmap of MRAE. Fourth row: the residual heatmap at band 31.

TABLE VII
QUANTITATIVE PERFORMANCE IN THE HOUSTON DATASET

Method	SAM	PSNR	ERGAS	CC	RMSE	UIQI
G-SOMP+	1.7447	37.4478	0.5945	0.9984	0.0092	0.9971
CSU	1.8510	38.2631	0.5479	0.9979	0.0084	0.9974
CNMF	1.5375	38.3817	0.5346	0.9987	0.0083	0.9960
STEREO	1.2055	46.1988	0.2833	0.9989	0.0039	0.9995
CSTF	0.9713	45.9991	0.2531	0.9991	0.0037	0.9996
SCOTT	1.1564	45.8172	0.2946	0.9988	0.0041	0.9995
MIAE	0.9002	48.5371	0.2516	0.9993	0.0033	0.9996
SURE	0.8681	48.6378	0.2164	0.9993	0.0030	0.9997
XINet	0.8679	49.2756	0.2524	0.9991	0.0031	0.9997
Ours	**0.7941**	**49.4177**	**0.2012**	**0.9994**	**0.0028**	**0.9998**

TABLE VIII
QUANTITATIVE PERFORMANCE IN THE WASHINGTON DC MALL DATASET

Method	SAM	PSNR	ERGAS	CC	RMSE	UIQI
G-SOMP+	2.6587	32.0508	1.7351	0.9686	0.0095	0.9809
CSU	2.8591	31.7495	1.7696	0.9647	0.0103	0.9830
CNMF	2.5965	33.0363	1.5768	0.9752	0.0084	0.9856
STEREO	2.5519	34.4927	2.3120	0.9470	0.0059	0.9653
CSTF	2.0705	35.1518	1.5513	0.9723	0.0055	0.9877
SCOTT	2.0401	35.1259	1.7666	0.9652	0.0054	0.9807
MIAE	1.6301	36.2152	2.1133	0.9635	0.0045	0.9752
SURE	1.6893	36.2620	3.0149	0.9638	0.0041	0.9728
XINet	1.6169	36.0969	1.8606	0.9635	0.0044	0.9833
Ours	**1.5625**	**38.0656**	**1.3038**	**0.9806**	**0.0038**	**0.9914**

DIP to fully model the latent statistics, which is of benefit to the SR task. Besides, Fig. 8 exhibits the gap between two different inputs. The output generated from noise and $\mathcal{Z}$ produces visible distortions at the bottom of the image and shows a limited reconstruction ability in the complex region, which needs more critical priors to recover. On the other hand, Dual U-Net with $\mathcal{X}_{\text{input}}$ and $\mathcal{Z}$ as the input still gives a stable result in the overall region.

E. Comparison With State-of-the-Arts

We select nice state-of-the-art methods as competitors to demonstrate the effectiveness of our proposed M2U-Net. The quantitative and visual results are illustrated in Tables VII–X and Figs. 9–13, respectively.

1) Houston: The SR results of the Houston dataset are illustrated in Fig. 9. G-SOMP+, CSU, and CNMF produce different levels of distortions in both spatial and spectral domains. STEREO and SCOTT give similar reconstruction results, while CSTF is more effective in spectral preservation, showing outstanding ability in the traditional methods. Among the DL-based ones, MIAE, SURE, and XINet all achieve satisfactory results in spatial enhancement and spectral fidelity. Most importantly, the error heatmap of the proposed M2U-Net has the bluest color among all methods, especially in the edge region, which demonstrates its best performance in the SR outcomes.

Moreover, Table VII provides a more straightforward assessment of all methods. It is noted that M2U-Net reaches the highest scores in terms of all metrics, showing consistent results with the visual inspection. Besides, XINet also achieves a competitive performance and shows an advantage over MIAE and SURE. CSTF outperforms other traditional methods in the SAM index, indicating its strong ability in spectral preservation, while three matrix-based methods are left far behind in all indicators, especially PSNR value. Besides, the PNSR as the function of the spectral band is illustrated in Fig. 13. As can be seen, XINet, MIAE, and our method take the leading position in the overall bands, which reflects their ability in spatial reconstruction.

2) Washington DC Mall: We show the SR results of the Washington DC Mall dataset in Fig. 10. It can be seen that the error map of three matrix-based methods is not desirable,

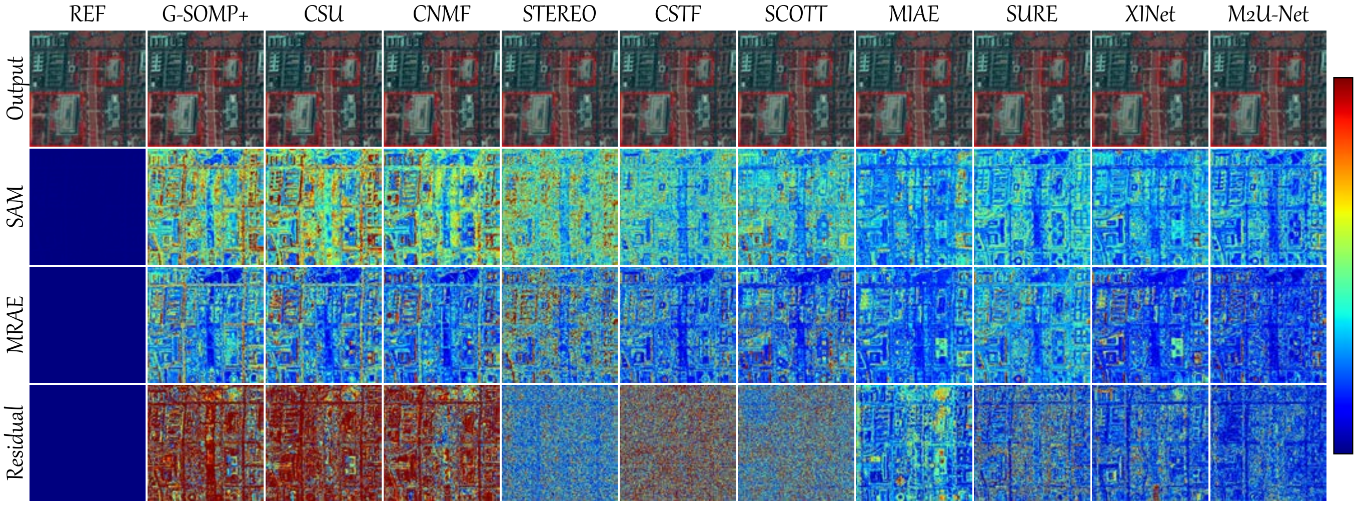

Fig. 10. Visual comparison of all methods in the Washington DC Mall dataset. First row: pseudocolor image of SR outputs (R:60, G:27, B:17). Second row: the heatmap of SAM error. Third row: the heatmap of MRAE. Fourth row: the residual heatmap at band 69.

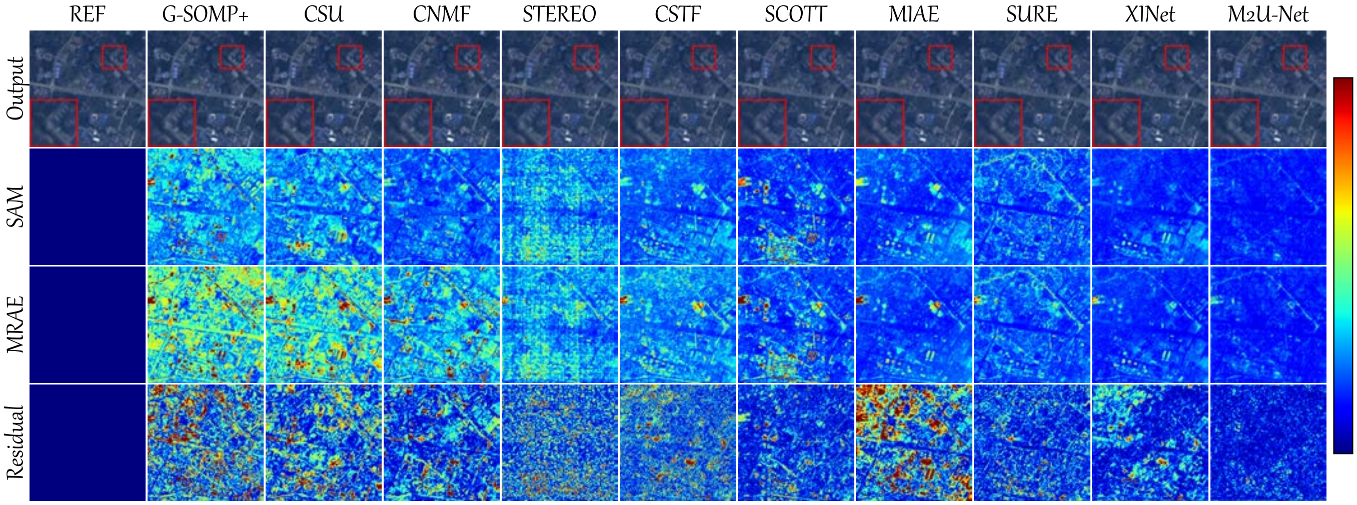

Fig. 11. Visual comparison of all methods in the TianGong-1 dataset. First row: pseudocolor image of SR outputs (R:30, G:20, B:7). Second row: the heatmap of SAM error. Third row: the heatmap of MRAE. Fourth row: the residual heatmap at band 44.

TABLE IX
QUANTITATIVE PERFORMANCE IN THE TIANGONG-1 DATASET

Method	SAM	PSNR	ERGAS	CC	RMSE	UIQI
G-SOMP+	1.7714	36.1735	0.4124	0.9892	0.0128	0.9980
CSU	1.5471	36.8678	0.3873	0.9903	0.0115	0.9983
CNMF	1.2508	37.0527	0.3904	0.9928	0.0113	0.9988
STEREO	1.6001	41.5475	0.2811	0.9915	0.0082	0.9991
CSTF	1.3019	41.3339	0.2488	0.9935	0.0072	0.9993
SCOTT	1.2381	41.1616	0.2461	0.9923	0.0077	0.9993
MIAE	1.1705	43.9053	0.2062	0.9958	0.0063	0.9994
SURE	1.2233	43.6018	0.2192	0.9947	0.0064	0.9995
XINet	1.0086	45.3571	0.1749	0.9962	0.0053	0.9996
Ours	**0.8431**	**45.9279**	**0.1387**	**0.9973**	**0.0043**	**0.9998**

especially in the spectral domain. Though three tensor-based methods produce better results, the spectral distortion cannot be ignored. As for DL-based ones, the difference between MIAR, SURE, and XINet is not significant, and all of them can generate high-quality HrHSI. It is worth mentioning that M2U-Net still provides the lowest error maps compared with all competitors, showing its superior ability in spatial and spectral recovery.

Quantitative results of all methods are summarized in Table VIII to objectively evaluate their performance. Our proposed M2U-Net achieves the best scores in all metrics, indicating its strong ability in detail recovery and spectral preservation. Though MIAE, SURE, and XINet have the closest results to ours in most indicators, there is still a gap lying in the PSNR value. Only the matrix-based methods give unsatisfactory results compared with the aforementioned competitors. Moreover, it can be seen in Fig. 13 that M2U-Net shows advantages over the spectral bands, which directly demonstrates the effectiveness of our method.

3) TianGong-1: The SR outcomes for TianGong-1 are depicted in Fig. 11. It can be observed that there are noticeable spatial and spectral distortions in matrix-based methods,

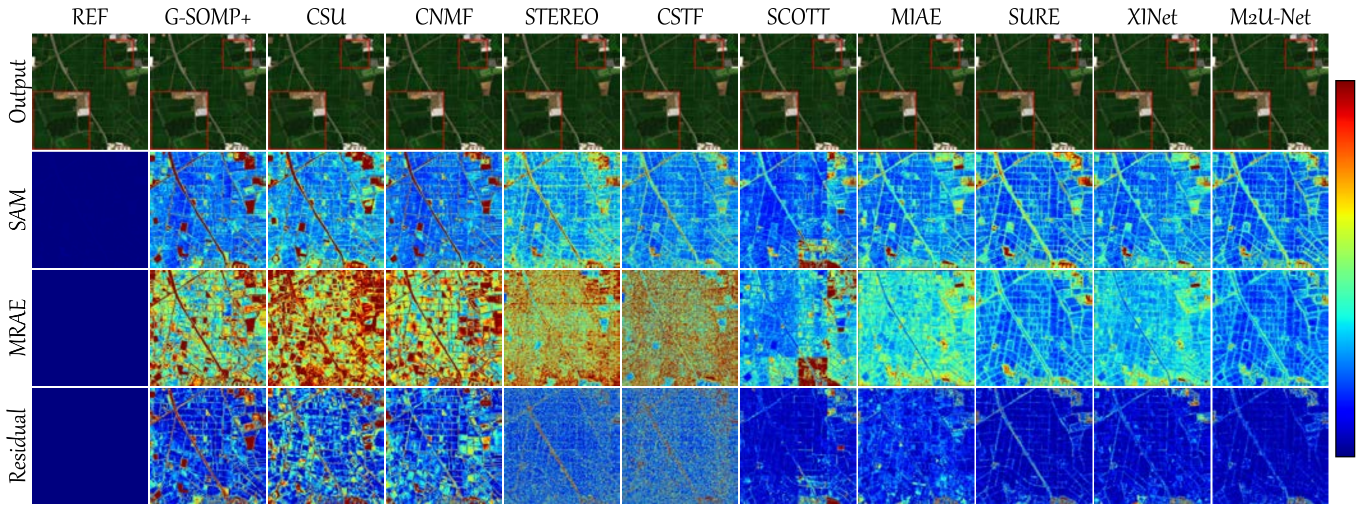

Fig. 12. Visual comparison of all methods in the Chikusei dataset. First row: pseudocolor image of SR outputs (R:57, G:37, B:17). Second row: the heatmap of SAM error. Third row: the heatmap of MRAE. Fourth row: the residual heatmap at band 26.

TABLE X
QUANTITATIVE PERFORMANCE IN THE CHIKUSEI DATASET

Method	SAM	PSNR	ERGAS	CC	RMSE	UIQI
G-SOMP+	1.3008	42.0384	0.6641	0.9927	0.0093	0.9963
CSU	1.4271	40.1465	0.8776	0.9874	0.0096	0.9930
CNMF	1.2194	41.3611	0.6565	0.9928	0.0091	0.9952
STEREO	1.1298	45.6701	0.5260	0.9930	0.0054	0.9955
CSTF	0.9232	44.4402	0.4857	0.9938	0.0058	0.9961
SCOTT	0.9544	45.0450	0.7615	0.9833	0.0103	0.9954
MIAE	0.9653	47.6483	0.4379	0.9944	0.0049	0.9981
SURE	0.9583	48.0581	0.3273	0.9965	0.0045	0.9992
XINet	0.9073	48.5201	0.4395	0.9945	0.0045	0.9986
Ours	**0.8687**	**49.7405**	**0.2783**	**0.9969**	**0.0040**	**0.9993**

TABLE XI
RUNNING TIME AND MODEL SIZE OF DL-BASED METHODS IN THE TIANGONG-1 DATASET

Method	M2U-Net			MIAE	SURE	XINet
	Stage 1	Stage 2	Stage 3			
Running Time (S)	20	90	7900	698	2256	5580
Parameters (K)	0.5	17	14135	203	195	701

i.e., G-SOMP+, CSU, and CNMF, partly owing to the limited expressiveness of designed priors. As for the tensor-based methods, STEREO obtains relatively poor results in terms of spectral reconstruction, and SCOTT exhibits limited spatial enhancement in the building area. Overall, DL-based methods yield more satisfactory results compared with traditional ones. However, it is still a challenge for them to reconstruct small objects with high fidelity, such as buildings. Differently, the proposed M2U-Net is robust to different objects and gives the best SR result with the lowest error maps.

The quantitative evaluation listed in Table IX also favors the conclusion made in visual comparison. Our proposed method achieves huge advantages in all metrics, especially for PSNR and SAM, which firmly exhibit its superior ability. XINet gives the second-best results, followed by MIAE. SCOTT is good at spectral recovery while performing unsatisfactorily in the spatial domain. Matrix-based methods are generally inferior to the aforementioned methods, in which CNMF shows impressive spectral fidelity among all competitors. Besides, the PSNR curve shown in Fig. 13 indicates that our method also acquires competitive results compared with others.

4) Chikusei: The visual results in the Chikusei dataset are displayed in Fig. 12. As we can see, nonnegligible distortion appears in the outcomes of three matrix-based methods, showing limited ability in the building areas. Tensor-based methods can produce more satisfactory results with fewer artifacts. Moreover, CSTF and SCOTT give precise spectral reconstruction, even closer to SURE. More importantly, M2U-Net consistently produces the best results among all competitors, with XINet and SURE following behind.

Table X displays the evaluation metrics in the Chikusei dataset, which further proves the judgments from the visual results. Specifically, our proposed M2U-Net obtains the highest scores in terms of all metrics, with a huge lead in the PSNR value, which convincingly indicates the excellent ability to recover spatial details and preserve spectral features. CSTF and SCOTT also exhibit favorable spectral reconstruction ability but are still inferior to DL-based methods in the spatial domain. Three matrix-based methods, G-SOMP+, CSU, and CNMF, suffer from serious distortion and lead to the loss of spatial details and spectral features. On the other hand, the PSNR curve in Fig. 13 shows that our method obtains overall good performance along all bands, which shows consistent conclusions with the quantitative results.

5) Running Time and Model Size of DL-Based Methods: Table XI displays the running time and model size of DL-based methods in the TianGong-1 dataset. Overall, MIAE and SURE have the smallest model sizes and also take the shortest running time due to their simple architecture. However, the main burden of our process is from the last stage. Specifically, the DIL-Net and SML-Net are lightweight

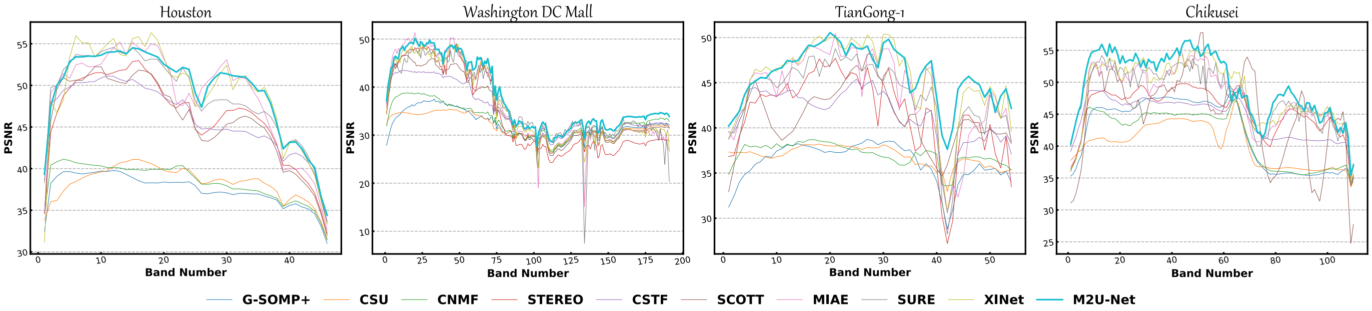

Fig. 13. PSNR value along the spectral band in four datasets.

since the DIL-Net only contains trainable degradation parameters and the SML-Net is responsible for the simple spectral learning. However, the last stage aims to capture the latent hyperspectral prior with two coupled U-Nets. As discussed in Table II, the size of Dual U-Net is influenced by the number of P that also determines the fusion performance. To achieve a better result, we set P as 260, which inevitably leads to the parameter burden.

F. Discussion on Regularization Parameters Estimation

Instead of manually designing explicit prior terms, DIP states that image prior can be found in the space of the network's parameters directly. However, it is still an open problem to design the optimal network for the HISR task. The most straightforward way to evaluate the estimated parameters lies in the super-resolved performance. In our experiments, we find that the estimation of the network parameters is closely related to two factors, i.e., the initialized input and the architecture of the designed network. First, most of the existing methods feed random noise as the input. However, the limited information contained in the input severely hinders the network from extracting the image prior, which can be demonstrated in Table VI. Second, the network architecture plays an important role in the parameter estimation. Since different network structures lead to different results, exploring the most relevant architecture for DIP to model a specific task is beneficial to the prior extraction. Specifically, in the task of HISR, we need to fully utilize the information of the observed HrMSI-LrHSI pair to guide the parameter estimation. Hence, we construct a Dual U-Net with one network to extract multiscale features of HrMSI for guidance and another network to realize image generation. The results in Table VI demonstrate the effectiveness by involving the information of HrMSI.

G. Robustness Analysis

In this section, we examine our method under more challenging conditions. First, we analyze the SR performance in four different scale factors in the Houston dataset, with CNMF, CSTF, and MIAE as competitors. Then, M2U-Net is performed on a real dataset to verify its potential.

1) Scale Factor: The SR results under four scale factors in the Houston data are shown in Fig. 14. As can be seen, the variation of the scale factor leads to different levels of influence on all methods, in which CNMF exhibits a more sensitive change. Though our proposed M2U-Net is also degraded with the increase in the scale factor, it still outperforms all competitors in all cases with less fluctuation compared with others, which firmly verifies the superiority of our method.

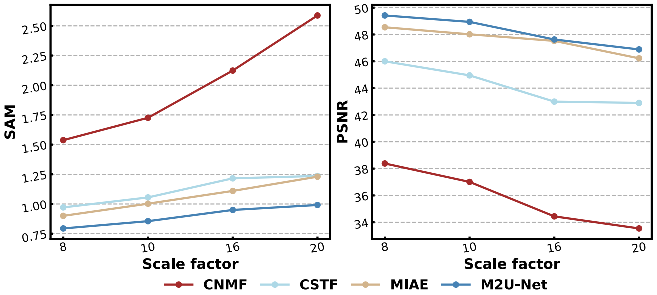

Fig. 14. PSNR and SAM metrics under the scale factor of 8, 10, 16, and 20 in the Houston dataset.

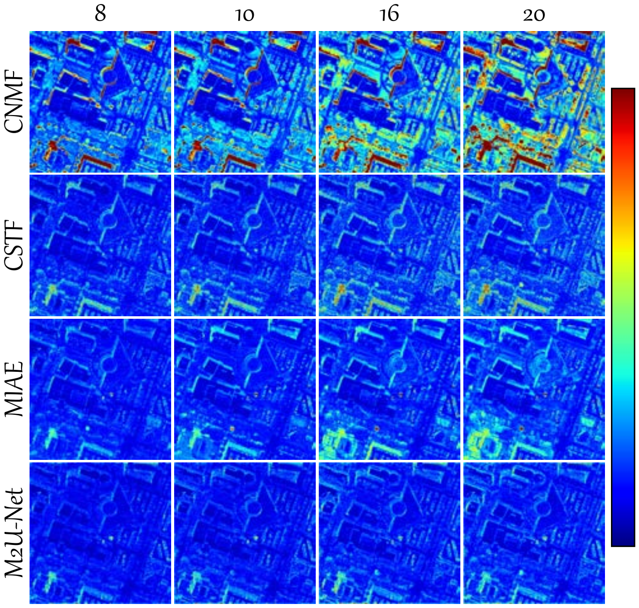

Fig. 15. SAM heatmap under four different scale factors in the Houston dataset.

Besides, we further illustrate their SAM and MRAE heatmap under four different scale factors in Figs. 15 and 16, respectively. As can be seen, the color of each error map gradually becomes brighter with the increase in the scale factor, which indicates that all methods suffer from different

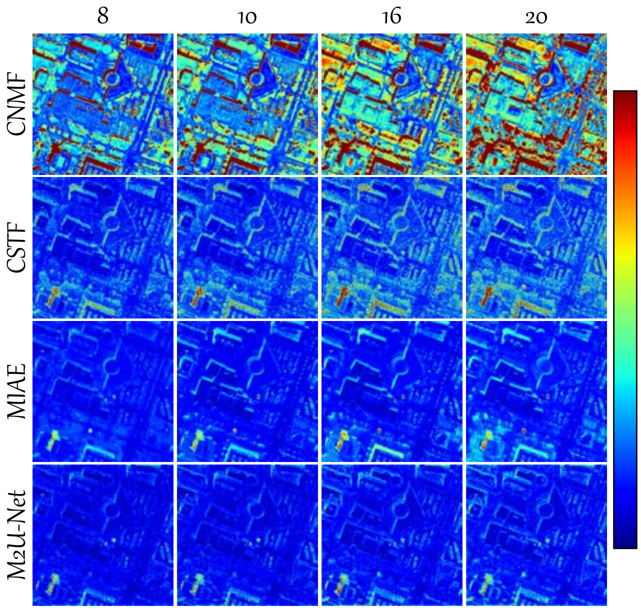

Fig. 16. MRAE heatmap under four different scale factors in the Houston dataset.

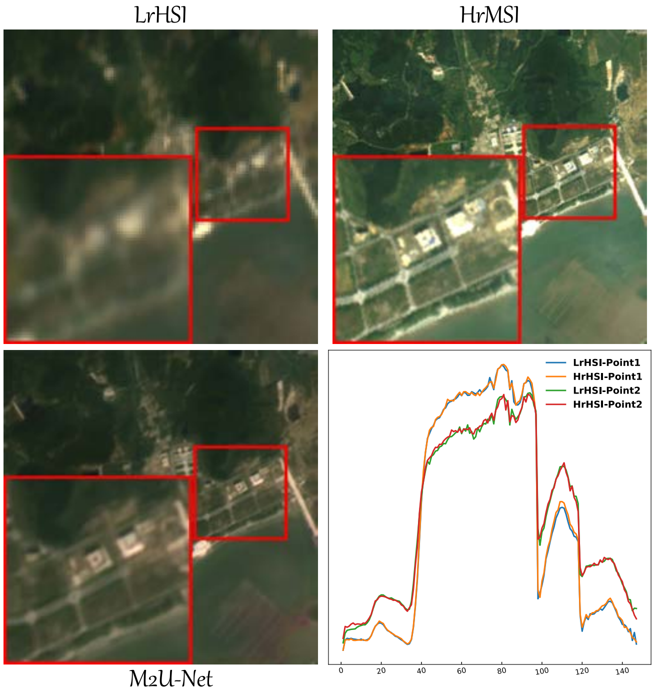

Fig. 17. SR result of M2U-Net performed in the Liao Ning-01 dataset.

levels of influence. However, our method still has the lowest error map in each case and undergoes minimal effect.

2) Real Dataset: Apart from the simulation experiments, we apply M2U-Net to a real dataset called Liao Ning-01.[1] This dataset is captured by the ZY-1 02D satellite over Dalian city, Liaoning province. It consists of one LrHSI with a size of 300 × 300 × 166 and one corresponding HrMSI with a size of 900 × 900 × 8. In our study, a local area is cropped for the examination, with LrHSI and HrMSI sizes of 100 × 100 × 147 and 300 × 300 × 8, respectively. As shown in Fig. 17, we display the input HrMSI-LrHSI pair and the reconstructed HrHSI from M2U-Net. Besides, we further enlarge a local area by a factor of four to intuitively observe their differences. Compared with the original LrHSI, the output of M2U-Net contains more spatial details, including textures and edges, which indicates our ability in spatial enhancement. Moreover, we select two points from LrHSI and its corresponding HrHSI to visualize their spectral curve. It can be observed that the reconstructed spectral feature is still close to that of LrHSI, verifying its satisfactory spectral fidelity.

VI. Conclusion

Considering the strong dependence on large training datasets and insufficient utilization of the degradation model, we design a model-informed multistage unsupervised network, M2U-Net for short, to deal with the HISR task. Specifically, our method is implemented in a three-stage manner, with DIL in the head, IIE in the body, and deep image generation in the tail. The first stage is to exploit the deep information of the degradation model and leverage it to guide the following two stages. The second stage utilizes the derived LrMSI to establish a spectral mapping function and obtains an initialized image with expressive HrHSI-relevant information as the input of stage three. Finally, we design Dual U-Net as a strong regularizer to capture the image prior and realize a high-quality reconstruction. The overall process purely depends on the observed HrMSI-LrHSI pair without any extra training triplets. We perform an extensive ablation study to demonstrate the effectiveness of each stage and compare M2U-Net with nine methods to verify our superior SR performance in both simulated and real datasets.

In future works, the proposed method will be further improved in two directions: one is to enhance the robustness against the noise, and the other is to reduce parameter burden while keeping high-quality reconstruction.

References

[1] M. Wang et al., "Tensor decompositions for hyperspectral data processing in remote sensing: A comprehensive review," *IEEE Geosci. Remote Sens. Mag.*, vol. 11, no. 1, pp. 26–72, Mar. 2023.

[2] X. Cheng, M. Zhang, S. Lin, Y. Li, and H. Wang, "Deep self-representation learning framework for hyperspectral anomaly detection," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–16, 2024.

[3] M. Wang, D. Hong, B. Zhang, L. Ren, J. Yao, and J. Chanussot, "Learning double subspace representation for joint hyperspectral anomaly detection and noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 3261964.

[4] Y. Ding et al., "Multi-scale receptive fields: Graph attention neural network for hyperspectral image classification," *Expert Syst. Appl.*, vol. 223, Aug. 2023, Art. no. 119858.

[5] J. He, Q. Yuan, J. Li, Y. Xiao, and L. Zhang, "A self-supervised remote sensing image fusion framework with dual-stage self-learning and spectral super-resolution injection," *ISPRS J. Photogramm. Remote Sens.*, vol. 204, pp. 131–144, Oct. 2023.

[6] L. Ren, D. Hong, L. Gao, X. Sun, M. Huang, and J. Chanussot, "Orthogonal subspace unmixing to address spectral variability for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 3236471.

[7] L. Ren, D. Hong, L. Gao, X. Sun, M. Huang, and J. Chanussot, "Hyperspectral sparse unmixing via nonconvex shrinkage penalties," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5500415.

[8] G. Vivone, "Multispectral and hyperspectral image fusion in remote sensing: A survey," *Inf. Fusion*, vol. 89, pp. 405–417, Jan. 2023.

[1] https://drive.google.com/drive/folders/1JLCCB6ld5R49HDLN5SsMISx1d0fuqRjO

[9] Q. Li, Y. Yuan, X. Jia, and Q. Wang, "Dual-stage approach toward hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 31, pp. 7252–7263, 2022.

[10] J. He et al., "Spectral super-resolution meets deep learning: Achievements and challenges," *Inf. Fusion*, vol. 97, Sep. 2023, Art. no. 101812.

[11] 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.

[12] R. Ran, L.-J. Deng, T.-X. Jiang, J.-F. Hu, J. Chanussot, and G. Vivone, "GuidedNet: A general CNN fusion framework via high-resolution guidance for hyperspectral image super-resolution," *IEEE Trans. Cybern.*, vol. 53, no. 7, pp. 4148–4161, Jul. 2023, doi: 10.1109/TCYB.2023.3238200.

[13] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[14] J. R. Patel, M. V. Joshi, and J. S. Bhatt, "A novel approach for hyperspectral image superresolution using spectral unmixing and transfer learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 1512–1515.

[15] Y. Li, W. Guo, W. Xie, T. Jiang, and Q. Du, "MMIF: Interpretable hyperspectral and multispectral image fusion via maximum mutual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 3343711.

[16] S.-Q. Deng, L.-J. Deng, X. Wu, R. Ran, D. Hong, and G. Vivone, "PSRT: Pyramid shuffle-and-reshuffle transformer for multispectral and hyperspectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 3244750.

[17] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. Plaza, "BockNet: Blind-block reconstruction network with a guard window for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 3335484.

[18] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. J. Plaza, "PDBSNet: Pixel-shuffle downsampling blind-spot reconstruction network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 5511914.

[19] D. Wang, L. Zhuang, L. Gao, X. Sun, X. Zhao, and A. Plaza, "Sliding dual-window-inspired reconstruction network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 3351179.

[20] C. Zhu et al., "Hyperspectral and multispectral remote sensing image fusion using SwinGAN with joint adaptive spatial–spectral gradient loss function," *Int. J. Digit. Earth*, vol. 16, no. 1, pp. 3580–3600, Oct. 2023.

[21] W.-J. Guo, W. Xie, K. Jiang, Y. Li, J. Lei, and L. Fang, "Toward stable, interpretable, and lightweight hyperspectral super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 22272–22281.

[22] S. Deng, L.-J. Deng, X. Wu, R. Ran, and R. Wen, "Bidirectional dilation transformer for multispectral and hyperspectral image fusion," in *Proc. 32nd Int. Joint Conf. Artif. Intell.*, Aug. 2023, pp. 3633–3641.

[23] C. Zhu, S. Deng, Y. Zhou, L.-J. Deng, and Q. Wu, "QIS-GAN: A lightweight adversarial network with quadtree implicit sampling for multispectral and hyperspectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 3332176.

[24] Q. Ma, J. Jiang, X. Liu, and J. Ma, "Learning a 3D-CNN and transformer prior for hyperspectral image super-resolution," *Inf. Fusion*, vol. 100, Dec. 2023, Art. no. 101907.

[25] R. Ran, L.-J. Deng, T.-J. Zhang, J. Chang, X. Wu, and Q. Tian, "KNLConv: Kernel-space non-local convolution for hyperspectral image super-resolution," *IEEE Trans. Multimedia*, early access, pp. 1–13, Mar. 28, 2024.

[26] Z. Han, D. Hong, L. Gao, B. Zhang, M. Huang, and J. Chanussot, "AutoNAS: Automatic neural architecture search for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 3186480.

[27] Z. Han, D. Hong, L. Gao, J. Yao, B. Zhang, and J. Chanussot, "Multimodal hyperspectral unmixing: Insights from attention networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5524913.

[28] Z. Han, C. Zhang, L. Gao, Z. Zeng, B. Zhang, and P. M. Atkinson, "Spatio-temporal multi-level attention crop mapping method using time-series SAR imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 206, pp. 293–310, Dec. 2023.

[29] Z. Chen, G. Wu, H. Gao, Y. Ding, D. Hong, and B. Zhang, "Local aggregation and global attention network for hyperspectral image classification with spectral-induced aligned superpixel segmentation," *Expert Syst. Appl.*, vol. 232, Dec. 2023, Art. no. 120828.

[30] Z. Chen, D. Hong, and H. Gao, "Grid network: Feature extraction in anisotropic perspective for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[31] Y. Chen, Q. Yuan, Y. Tang, Y. Xiao, J. He, and Z. Liu, "SENSE: Hyperspectral video object tracker via fusing material and motion cues," *Inf. Fusion*, vol. 109, Apr. 2024, Art. no. 102395.

[32] X. Cheng, M. Zhang, S. Lin, K. Zhou, S. Zhao, and H. Wang, "Two-stream isolation forest based on deep features for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[33] Q. Ma, J. Jiang, X. Liu, and J. Ma, "Deep unfolding network for spatiospectral image super-resolution," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 28–40, 2022.

[34] Q. Ma, J. Jiang, X. Liu, and J. Ma, "Multi-task interaction learning for spatiospectral image super-resolution," *IEEE Trans. Image Process.*, vol. 31, pp. 2950–2961, 2022.

[35] S. Liu, S. Miao, S. Liu, B. Li, W. Hu, and Y. Zhang, "Circle-Net: An unsupervised lightweight-attention cyclic network for hyperspectral and multispectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 4499–4515, 2023.

[36] R. B. Gomez, A. Jazaeri, and M. Kafatos, "Wavelet-based hyperspectral and multispectral image fusion," in *Geo-Spatial Image and Data Exploitation II*, vol. 4383. Bellingham, WA, USA: SPIE, 2001, pp. 36–42.

[37] Y. Zhang and M. He, "Multi-spectral and hyperspectral image fusion using 3-D wavelet transform," *J. Electron. China*, vol. 24, no. 2, pp. 218–224, Mar. 2007.

[38] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hyper-sharpening: A first approach on SIM-GA data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3008–3024, Jun. 2015.

[39] X. Han, J. Yu, J.-H. Xue, and W. Sun, "Hyperspectral and multispectral image fusion using optimized twin dictionaries," *IEEE Trans. Image Process.*, vol. 29, pp. 4709–4720, 2020.

[40] X. Fu, S. Jia, M. Xu, J. Zhou, and Q. Li, "Fusion of hyperspectral and multispectral images accounting for localized inter-image changes," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517218.

[41] S. Li, R. Dian, and H. Liu, "Learning the external and internal priors for multispectral and hyperspectral image fusion," *Sci. China Inf. Sci.*, vol. 66, no. 4, Apr. 2023, Art. no. 140303.

[42] K. Ren, W. Sun, X. Meng, G. Yang, J. Peng, and J. Huang, "A locally optimized model for hyperspectral and multispectral images fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.
K. Ren, W. Sun, X. Meng, G. Yang, J. Peng, and J. Huang, "A locally optimized model for hyperspectral and multispectral images fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022, Art. no. 5519015.

[43] 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.

[44] 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.

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

[46] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral images super-resolution via learning high-order coupled tensor ring representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4747–4760, Nov. 2020.

[47] Y. Wang, W. Li, N. Liu, Y. Gui, and R. Tao, "FuBay: An integrated fusion framework for hyperspectral super-resolution based on Bayesian tensor ring," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, doi: 10.1109/TNNLS.2023.3281355.

[48] F. Ye, Z. Wu, X. Jia, J. Chanussot, Y. Xu, and Z. Wei, "Bayesian nonlocal patch tensor factorization for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 32, pp. 5877–5892, 2023.

[49] J. Li et al., "Deep learning in multimodal remote sensing data fusion: A comprehensive review," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102926.

[50] L. Gao, J. Li, K. Zheng, and X. Jia, "Enhanced autoencoders with attention-embedded degradation learning for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5509417.

[51] Q. Li, Y. Yuan, and Q. Wang, "Multiscale factor joint learning for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 3312436.

[52] X.-H. Han and Y.-W. Chen, "Deep residual network of spectral and spatial fusion for hyperspectral image super-resolution," in *Proc. IEEE 5th Int. Conf. Multimedia Big Data*, Sep. 2019, pp. 266–270.

[53] X. Zhang, W. Huang, Q. Wang, and X. Li, "SSR-NET: Spatial–spectral reconstruction network for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5953–5965, Jul. 2021.

[54] X.-H. Han, Y. Zheng, and Y.-W. Chen, "Multi-level and multi-scale spatial and spectral fusion CNN for hyperspectral image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 4330–4339.

[55] W. Sun, K. Ren, X. Meng, C. Xiao, G. Yang, and J. Peng, "A band divide-and-conquer multispectral and hyperspectral image fusion method," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5502113, doi: 10.1109/TGRS.2020.3046321.

[56] J. Li, S. Du, R. Song, Y. Li, and Q. Du, "Progressive spatial information-guided deep aggregation convolutional network for hyperspectral spectral super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, doi: 10.1109/TNNLS.2023.3325682.

[57] H. Gao, S. Li, and R. Dian, "Hyperspectral and multispectral image fusion via self-supervised loss and separable loss," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 3204769.

[58] W. Sun et al., "MLR-DBPFN: A multi-scale low rank deep back projection fusion network for anti-noise hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 3146296.

[59] Q. Li, M. Gong, Y. Yuan, and Q. Wang, "Symmetrical feature propagation network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5536912.

[60] S. Jia, Z. Min, and X. Fu, "Multiscale spatial–spectral transformer network for hyperspectral and multispectral image fusion," *Inf. Fusion*, vol. 96, pp. 117–129, Aug. 2023.

[61] Z. Cao, S. Cao, L.-J. Deng, X. Wu, J. Hou, and G. Vivone, "Diffusion model with disentangled modulations for sharpening multispectral and hyperspectral images," *Inf. Fusion*, vol. 104, Apr. 2024, Art. no. 102158.

[62] J. Li, K. Zheng, Z. Li, L. Gao, and X. Jia, "X-shaped interactive autoencoders with cross-modality mutual learning for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5518317, doi: 10.1109/TGRS.2023.3300043.

[63] H. Wu, K. Zhang, S. Wu, S. Shi, C. Bian, and M. Zhang, "Unsupervised encoder–decoder network under spatial and spectral guidance for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 3320404.

[64] 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. 2020.

[65] J. Li, K. Zheng, L. Ni, and L. Gao, "Dual U-Nets autoencoders for unsupervised hyperspectral image super-resolution," in *Proc. Int. Conf. Remote Sens., Mapping, Geographic Syst. (RSMG)*, Nov. 2023, pp. 132–137.

[66] J. Li, K. Zheng, L. Gao, and L. Ni, "Interactive autoencoders with degradation constraint for hyperspectral super-resolution," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2023, pp. 7447–7450.

[67] J. Yang, L. Xiao, Y.-Q. Zhao, and J. C. Chan, "Unsupervised deep tensor network for hyperspectral–multispectral image fusion," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, pp. 1–15, May 3, 2023.

[68] S. Liu, S. Miao, J. Su, B. Li, W. Hu, and Y.-D. Zhang, "UMAG-Net: A new unsupervised multiattention-guided network for hyperspectral and multispectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7373–7385, 2021, doi: 10.1109/JSTARS.2021.3097178.

[69] H. V. Nguyen, M. O. Ulfarsson, J. R. Sveinsson, and M. D. Mura, "Deep SURE for unsupervised remote sensing image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 3215902.

[70] X. Cao, Y. Lian, K. Wang, C. Ma, and X. Xu, "Unsupervised hybrid network of transformer and CNN for blind hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5507615.

[71] L. Zhang, J. Nie, W. Wei, and Y. Zhang, "Unsupervised test-time adaptation learning for effective hyperspectral image super-resolution with unknown degeneration," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, pp. 1–17, Feb. 5, 2024.

[72] J. Li, K. Zheng, J. Yao, L. Gao, and D. Hong, "Deep unsupervised blind hyperspectral and multispectral data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[73] J. Li, K. Zheng, W. Liu, Z. Li, H. Yu, and L. Ni, "Model-guided coarse-to-fine fusion network for unsupervised hyperspectral image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[74] J. Li, K. Zheng, L. Ni, and L. Gao, "A new unsupervised network for hyperspectral and multispectral image fusion," in *Proc. 13th Workshop Hyperspectral Imag. Signal Processing: Evol. Remote Sens. (WHISPERS)*, Oct. 2023, pp. 1–5.

[75] X. Han, J. Yu, J. Luo, and W. Sun, "Hyperspectral and multispectral image fusion using cluster-based multi-branch BP neural networks," *Remote Sens.*, vol. 11, no. 10, p. 1173, 2019.

[76] J. Qin, L. Fang, R. Lu, L. Lin, and Y. Shi, "ADASR: An adversarial auto-augmentation framework for hyperspectral and multispectral data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[77] H. Wu, S. Wu, K. Zhang, X. Liu, S. Shi, and C. Bian, "Unsupervised blind spectral–spatial cross-super-resolution network for HSI and MSI fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5511914.

[78] A. Guo, R. Dian, and S. Li, "A deep framework for hyperspectral image fusion between different satellites," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 7939–7954, Jul. 2023.

[79] R. Dian, A. Guo, and S. Li, "Zero-shot hyperspectral sharpening," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 12650–12666, Oct. 2023.

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

[81] A. Qayyum, I. Ilahi, F. Shamshad, F. Boussaid, M. Bennamoun, and J. Qadir, "Untrained neural network priors for inverse imaging problems: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 6511–6536, May 2023.

[82] T. Uezato, D. Hong, N. Yokoya, and W. He, "Guided deep decoder: Unsupervised image pair fusion," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 87–102.

[83] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[84] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.

[85] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2014, pp. 63–78.

[86] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3586–3594.

[87] 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. 2011.

[88] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, Dec. 2018.

[89] C. Prévost, K. Usevich, P. Comon, and D. Brie, "Hyperspectral super-resolution with coupled Tucker approximation: Recoverability and SVD-based algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 931–946, 2020.

[90] J. Liu, Z. Wu, L. Xiao, and X.-J. Wu, "Model inspired autoencoder for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5522412.

[91] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Quantitative quality evaluation of pansharpened imagery: Consistency versus synthesis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1247–1259, Mar. 2016.

[92] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Aug. 2002.

[93] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Jul. 2004.

[94] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. Int. Conf. Fusion Earth Data*, 2000, pp. 99–103.

Jiaxin Li received the B.E. degree from Chongqing University, Chongqing, China, in 2020. He is currently pursuing the Ph.D. degree in cartography and geography information systems with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

He has published over ten peer-reviewed papers with more than 300 citations in Google Scholar, including four ESI highly cited papers. His research interests include multimodal remote sensing data fusion, hyperspectral image processing, and deep learning.

Mr. Li received the National Scholarship for Doctoral Students in 2023 and the Excellent Papers Award at the 2023 International Conference on Remote Sensing, Mapping and Geographic Information Systems. He served as a reviewer for more than ten journals. More information can be found at GitHub: https://github.com/JiaxinLiCAS.

Ke Zheng received the B.S. degree in geographic information systems from Shandong Agricultural University, Taian, China, in 2012, and the M.S. and Ph.D. degrees in remote sensing from the College of Geosciences and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing, China, in 2016 and 2020, respectively.

He spent two years as a Post-Doctoral Associate at the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Science, Beijing. He is currently an Instructor with the College of Geography and Environment, Liaocheng University, Liaocheng, Shandong, China. His research interests include image processing, machine learning, and deep learning and their application in Earth Vision.

Lianru Gao (Senior Member, IEEE) received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2002, and the Ph.D. degree in cartography and geographic information systems from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, in 2007.

He was a Visiting Scholar at the University of Extremadura, Cáceres, Spain, in 2014, and Mississippi State University (MSU), Starkville, MS, USA, in 2016. He is currently a Professor with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, CAS. In the last ten years, he was the PI of ten scientific research projects at national and ministerial levels, including projects by the National Natural Science Foundation of China (2018–2020, 2022–2025, and 2024–2028) and the National Key Research and Development Program of China (2021–2025). He has published more than 240 peer-reviewed articles, and there are more than 150 journal articles included in the Science Citation Index (SCI). He is a coauthor of three academic books, including *Hyperspectral Image Information Extraction*. He obtained 30 national invention patents in China. He was supported by the National Science Foundation for Distinguished Young Scholars of China in 2023. His research interests include hyperspectral image processing and information extraction.

Dr. Gao is a fellow of the Institution of Engineering and Technology. He was awarded the Outstanding Science and Technology Achievement Prize of the CAS in 2016. He won the Second Prize of the State Scientific and Technological Progress Award in 2018. He received the 2021 Outstanding Paper Award from the IEEE Workshop on Hyperspectral Image Processing: Evolution in Remote Sensing (WHISPERS). He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and *IET Image Processing*.

Li Ni received the Ph.D. degree in cartography and geographical information systems from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2015.

She is currently an Associate Researcher with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. Her research interests include land surface temperature retrieval and application of hyperspectral remote sensing.

Min Huang received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 1999, and the Ph.D. degree in optical engineering from Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences (CAS), Xi'an, China, in 2009.

He is currently a Professor with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, CAS, Beijing, China, and the School of Optoelectronics, University of Chinese Academy of Science, Beijing. In the last ten years, he was the Principal Investigator of ten scientific research projects at national and ministerial levels, including projects by the Key Research Program of CAS Support Technology and the National High-Tech Research and Development Program. He has published more than 30 peer-reviewed articles and obtained more than 20 national invention patents in China. His research interests include imaging spectrometry and computational optical imaging.

Dr. Huang received the Second Prize of the State Scientific and Technological Progress Award in 2010.

Jocelyn Chanussot (Fellow, IEEE) received the M.Sc. degree in electrical engineering from Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998.

Since 1999, he has been with Grenoble INP, where he is currently a Professor of signal and image processing. He has been a Visiting Scholar with Stanford University, Stanford, CA, USA; the KTH Royal Institute of Technology, Stockholm, Sweden; and the National University of Singapore (NUS), Singapore. Since 2013, he has been an Adjunct Professor with the University of Iceland, Reykjavík, Iceland. He holds the AXA Chair in remote sensing and an Adjunct Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning, and artificial intelligence.

Dr. Chanussot was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society from 2006 to 2008 and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing in 2009. He is a member of the Institut Universitaire de France from 2012 to 2017. He was the Founding President of the IEEE Geoscience and Remote Sensing French Chapter from 2007 to 2010, which received the 2010 IEEE GRS-S Chapter Excellence Award. He has received multiple outstanding paper awards. He was the Vice-President of the IEEE Geoscience and Remote Sensing Society, in charge of meetings and symposia, from 2017 to 2019. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing (WHISPERS). From 2005 to 2008, he was the Co-Chair of the GRS Data Fusion Technical Committee, where he was the Chair from 2009 to 2011. He is an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING and *Proceeding of the IEEE*. He was the Editor-in-Chief of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2011 to 2015. In 2014, he served as a Guest Editor for *IEEE Signal Processing Magazine*. He has been a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters) since 2018.