

# Enhanced Deep Image Prior for Unsupervised Hyperspectral Image Super-Resolution

Jiixin Li<sup>1</sup>, Ke Zheng<sup>1</sup>, Lianru Gao<sup>1</sup>, *Senior Member, IEEE*, Zhu Han<sup>2</sup>, *Student Member, IEEE*, Zhi Li<sup>1</sup>, and Jocelyn Chanussot<sup>3</sup>, *Fellow, IEEE*

**Abstract**—Depending on a large-scale paired dataset of low-resolution hyperspectral image (LrHSI), high-resolution multispectral image (HrMSI), and corresponding high-resolution hyperspectral image (HrHSI), the supervised paradigm has achieved impressive performance in the hyperspectral image super-resolution (HSR). However, the intrinsic data-intensive manner hinders its further application in real scenarios. Fortunately, deep image prior (DIP) allows us to achieve unsupervised super-resolution (SR) by solely utilizing degraded observations. However, its potential to accurately model complicated hyperspectral priors is still not fully exploited due to the following two factors: 1) existing methods tend to reconstruct the unknown HrHSI directly from a randomly generated noise, leaving it hard to leverage the scene-relevant information for prior learning and 2) the vanilla architecture is handcrafted for the generator network, which shows limitations in feature representation and thus fails to characterize the complicated image properties. To unleash the potential of DIP for the HSR task, we propose an enhanced DIP network, called EDIP-Net, by addressing the aforementioned impediments. Specifically, EDIP-Net is built with a two-stage four-component scheme, with a zero-shot learning (ZSL) stage for input image establishment and a deep image generation (DIG) stage for prior learning. First, we exploit the cross-scale spectral relationship inside the observations and thus design a degradation learning network to generate paired training samples from the observations themselves. As such, two image-coarse estimations are derived in a ZSL manner by learning an interactive spectral learning network. By replacing random noise with two estimations, we design a double U-shape architecture

for the generator network to capture their hyperspectral prior, each independently generating one HrHSI candidate. Under this premise, we further propose a degradation-aware decision fusion strategy to integrate the optimal results in a pixel-to-pixel manner. Extensive experiments demonstrate our superiority in achieving high-quality SR performance. The code will be available at <https://github.com/JiixinLiCAS>.

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

## I. INTRODUCTION

**H**YPERSPECTRAL images (HSIs) offer detailed spectral information spanning from visible to infrared wavelengths, enabling us to achieve accurate identification across various domains, including remote sensing, medical industry, and food safety [1], [2], [3], [4], [5], [6], [7]. However, the subtle spectral resolution inevitably leads to the sacrifice of spatial details, which greatly restricts their performance in the downstream tasks [8], [9], [10], [11], [12], [13], [14]. By contrast, multispectral images (MSIs) can provide finer spatial information with several spectral bands, showing complementary attributes to HSIs. Under this premise, hyperspectral image super-resolution (HSR) is a straightforward way to obtain high-resolution HSI (HrHSI) by fusing low-resolution HSI (LrHSI) and corresponding high-resolution MSI (HrMSI) [15], [16], [17], [18].

Numerous methods have been proposed for the task of HSR in the last decade, which can be roughly divided into model-based and deep learning (DL)-based categories [19], [20], [21], [22]. The former line of research leverages the degradation model to establish the relationship between the observed HrMSI–LrHSI pair and unknown HrHSI. By treating it as an ill-posed inverse problem, researchers focus on designing various regularizers to narrow the solution space for accurate reconstruction, such as total variation and nonlocal similarity [23], [24], [25], [26]. Despite the promising results achieved by hand-crafted priors, the intrinsic shallow structure restricts their further improvement when handling complicated cases. Moreover, different scenes exhibit diverse characteristics across the spatial–spectral dimension, which renders it difficult to discover a common prior for all scenes [27], [28]. Recently, supervised learning has dominated the research line of HSR by implicitly learning the common prior from large-scale paired training samples, i.e., LrHSI, HrMSI, and corresponding HrHSI [29]. By applying the well-trained model to unseen images, the captured data-driven prior can super-resolve the degraded observations into desirable outcomes.

Received 28 August 2024; revised 29 December 2024; accepted 16 January 2025. Date of publication 20 January 2025; date of current version 10 February 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 42325104 and Grant 62161160336. (*Corresponding author: Lianru Gao.*)

Jiixin Li and Zhi Li are 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; lizhi21@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 is 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).

Zhu Han is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, and the International Research Center of Big Data for Sustainable Development Goals, 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: hanzhu19@mails.ucas.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.2025.3531646

However, the performance gain heavily depends on the quality and quantity of training datasets, which leads to limited generalization and poor feasibility in real applications, especially the data-scarce HISR task. When the test images deviate from the training datasets, such as spectral band number, the degradation process, and the imaging conditions, the common prior from external datasets becomes inapplicable in such cases.

By contrast, Lempitsky et al. [30] state that the structure of a generator network can inherently capture the image statistics, thus adequate to serve as a powerful image prior. In other words, it is feasible to achieve unsupervised image restoration, including HISR, solely utilizing degraded images. Under this premise, many attempts have been made to apply deep image prior (DIP) to the task of HISR. The first work is proposed by Uezato et al. [31] that generate desirable HrHSI from a random noise input. Following that, Liu et al. [32], [33] incorporate the attention mechanism to achieve exhaustive prior learning. Cao et al. [34] further introduce Transformer to enhance the interaction between spatial and spectral domains. Despite their promising performance, the power of DIP is still not fully unleashed in terms of the following two factors, restricting its further potential to accurately model the complex hyperspectral prior. First, the current dominated work directly recovers the target HrHSI from randomly generated noise. Since the initialized input contains limited information about the task, it inevitably hinders DIP from leveraging the relevant information of the under-studying scene for further prior extraction. Second, the primitive structure of generator networks shows limitations in feature representation and thus fails to sufficiently model the complicated image characteristics. Considering the aforementioned problems, it is urgent for us to construct advanced networks to characterize the latent hyperspectral prior.

In this article, we propose an enhanced DIP network, termed EDIP-Net, to realize the full potential of DIP in the HISR. Specifically, it is built with a two-stage four-component scheme, with a zero-shot learning (ZSL) stage for input image establishment and a deep image generation (DIG) stage for prior learning. Specifically, we first design a degradation learning network to adaptively infer the unknown degradation parameters, i.e., point spread function (PSF) and spectral response function (SRF), and simultaneously generate paired samples for training an interactive spectral learning network. According to the cross-scale spectral relationship inside the observations, two image-coarse estimations with rich scene-relevant information can be obtained in a ZSL manner. By replacing random noise with two estimations, we design a double U-shape architecture for the generator network to capture their hyperspectral prior, each independently generating one HrHSI candidate. Under this premise, we apply the learned SRF to design a degradation-aware decision fusion strategy by comparing their spectral degraded outcomes with observed HrMSI. This straightforward strategy enables us to effectively integrate the optimal results in a pixel-to-pixel manner and provide a more plausible super-resolved result. The contributions of our proposed EDIP-Net are summarized as follows.

- 1) We propose a novel unsupervised network by leveraging the power of DIP. It purely depends on the observed

HrMSI–LrHSI to achieve HISR without any external datasets.

- 2) To promote the generator network to effectively leverage the relevant information about the under-studying scene, we design a degradation learning network to generate paired training samples from the observations themselves and thus generate two image-coarse estimates with sufficient scene-relevant information in a ZSL manner, which can facilitate the prior learning of the following DIG stage.
- 3) A double U-shape architecture is tailored for the generator network, allowing each stream to separately extract the hyperspectral prior of two input estimations and together produce two HrHSI candidates. Following that, we design a degradation-aware decision fusion strategy to effectively integrate the optimal result and generate the final super-resolved result.

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

## II. RELATED WORK

In this section, we divide the mainstream algorithms of HISR into two categories, namely, model-based and deep learning-based methods, and give a brief introduction to each research line. Then, some related technologies are described, including DIP and ZSL.

### A. Model-Based

The HISR task is naturally an ill-posed problem and researchers handcraft various priors to constrain the optimization process for a desirable reconstruction [35]. Among them, matrix decomposition and tensor representation are two popular approaches to modeling the latent HrHSI. Wang et al. [36] impose the low-rank structure prior via the subspace representations and form a self-supervised denoiser framework for the HISR. Apart from the low-rank property, Zhang et al. [37] consider the band correlations by designing a group spectral embedding. Han et al. [38] embed the global and local similarities into the sparse representation, guaranteeing a faithful reconstruction. Fu et al. [39] design a multitask framework to jointly realize denoising and super-resolution (SR). From the perspective of high-dimensional tensor, Ye et al. [40] characterize low-rank structures under the canonical polyadic factorization and adaptively achieve rank determination. Xu et al. [41] cluster the similar patches to account for the low-rank property. Differently, Zhang et al. [42] handcraft two graphs in the spatial–spectral domains to separately model the spatial–spectral structures under the Tucker decomposition. Dian et al. [43] combine the nonlocal self-similarity with the sparse Tucker factorization and achieve a reliable reconstruction. Dian et al. [44] further extend the existing tensor nuclear norm, allowing for a more

flexible low-rank definition. Similarly, Xu et al. [45] propose two variants for the tensor nuclear norm, alleviating the effect caused by mirroring boundary and singleton domain. Wu et al. [46] introduce a novel cross-domain decomposition to fully exploit the latent low-rank property.

Despite the satisfactory results achieved by model-based approaches, the prior term needs careful construction for different scenarios and shows limitations in modeling complicated scene structures, which becomes an obstacle to their further development.

### B. Deep Learning-Based

Deep learning-based methods mainly consist of two categories, i.e., supervised and unsupervised learning [47], [48], [49], [50], [51]. The former captures common priors from large-scale external datasets by fully considering their rich diversity across images. By applying these captured priors to unseen images, satisfactory results can be usually achieved. For example, Ran et al. [15] propose a general framework equipped with a guidance branch and a reconstruction branch, which forms a progressive learning manner. Sun et al. [52] combine wavelet decomposition with deep network to handle the noise-case HISR. Later, Sun et al. [53] apply the divide-and-conquer strategy to realize HISR with spectral coverage discrepancy. Zhu et al. [54] design a lightweight adversarial network by introducing the implicit neural representation, which can generate high-quality fusion results. Hu et al. [55] construct a two-path structure, with one path for spectral preservation and another for detail reconstruction. Dian et al. [56] propose a three-stage scheme by utilizing the network to learn the residual prior and incorporate the learned prior to guide the model optimization. Considering the intrinsic shortages of the convolutional operator, researchers begin to apply the Transformer architecture to the HISR task [57]. Deng et al. [58] consider the interaction among different paths while reducing computation complexity. Jia et al. [59] propose two Transformer modules, namely, spatial and spectral Transformer, to achieve feature extraction, and introduce a pretraining strategy to enhance their representation ability. Chen et al. [60] propose a cyclic Transformer architecture to promote the feature interactions across two input modalities. Ma et al. [61] further combine Transformer with convolutional layers to realize an unfolding network. Moreover, the diffusion model is also introduced to tackle HISR by designing cycle framework [62] or disentangled modulations [63]. Due to their powerful learning ability, supervised methods are capable of capturing rich priors from external datasets and therefore produce impressive reconstruction results [64], [65], [66], [67]. However, the consistency between training and testing datasets is the prerequisite of high performance, which means small variations will lead to significant performance degradation in such a supervised paradigm.

Considering the abovementioned issues, the unsupervised technique is brought into the HISR task [68]. This kind of category only needs observed HrMSI–low-resolution MSI (LrMSI) pair as the inputs and then train an image-specific network to generate the final output. In the early work,

researchers attempt to apply the decomposition strategy to learn the latent representation of the target image and utilize these features to reconstruct HrHSI. Specifically, Li et al. [69], Gao et al. [70], Qu et al. [71], Zheng et al. [72], Liu et al. [73], and Wu et al. [74] decompose the HrMSI–LrHSI pair into endmember and corresponding abundance and infer the unknown HrHSI by combining these middle features. Moreover, Yang et al. [75] and Wang et al. [76] represent the HrHSI as the combination of the three-order tensor and exploit the degradation relationship to realize unsupervised learning. Another strategy is to leverage the latent spectral relationship to guide the image reconstruction. Li et al. [77], [78] train a spectral learning network on paired samples generated from a degradation network, forming a simple but effective scheme. Similarly, Qin et al. [79] introduce the augmentation technique to enrich the data diversity and lead to a robust downsampling network for image generation. Moreover, Wu et al. [80] and Shi et al. [81] design a two-path architecture to separately achieve spatial and spectral SR, and leverage the degradation model to guide the network learning.

### C. Deep Image Prior

Instead of manually designing explicit regularizers for the inverse problem, Lempitsky et al. [30] state that the structure of a generator network exhibits powerful ability in modeling the complicated image prior. Under this premise, researchers attempt to unleash the power of DIP for the remote sensing field. For instance, Rasti et al. [82] and Zhou and Rodrigues [83] leverage DIP to achieve a high-quality hyperspectral unmixing where a simple architecture with a skip connection is constructed to infer unknown endmembers and corresponding abundances. Zhuang et al. [84] utilize a neural denoising network to learn representation coefficients and successfully represent the low rankness prior to remove the latent noise. Moreover, DIP has also been brought into the field of HISR. Instead of directly upsampling LrHSI by traditional bicubic interpolation, Zheng et al. [85] and Bandara et al. [86] apply DIP to provide a more informative input for the following hyperspectral pansharpening networks. Sidorov and Hardeberg [87] achieve promising performance for single HISR, and they also demonstrate their potential in other low-level task, including denoising and inpainting. Uezato et al. [31] make the first attempt to realize the HISR task by fusing one pair of HrMSI and LrHSI. Along the same line, Liu et al. [32], [88] embed the attention mechanism into the generator network, thus fully extracting the latent prior of the target image. Fang et al. [89] design two coupled DIP networks to separately generate the abundance and endmember, allowing for a sufficient spatial–spectral feature extraction. Li et al. [90] exploit the latent model information to facilitate the image prior extraction, forming a multistage learning framework. Moreover, Zhang et al. [91], [92] design a novel coarse-to-fine scheme to progressively model the complicated prior, making it easier to obtain satisfactory reconstruction results. Recently, Nie et al. [93] jointly achieve image registration and fusion tasks under the DIP framework. However, the potential of DIP is still not fully exploited for the



HISR task in terms of the initialized input and the structure of the generator network, thereby leaving large room for further improvement.

#### D. Zero-Shot Learning

ZSL refers to the ability to perform a task only relying on the test image [94], where the training samples are solely extracted from the test image itself. Mansour and Heckel [95] generate training pairs by decomposing a noise image into two downsampled images. By learning the mapping function from one to the other and applying it to the original image, a denoised outcome can be yielded. Shocher et al. [96] further introduce ZSL into the SR field. Specifically, they extract the training pairs by downsampling the test image and learn a coarse-scale relationship to achieve the original-scale SR. Following the same line, Dian et al. [97] adaptively learn the degradation model to produce reduced-scale training samples. By adapting the downsampled relationship to full resolution, a high-quality result can be reconstructed. Similarly, Guo et al. [98] design a framework to simultaneously realize image registration, degradation learning, and fusion in a ZSL manner. Since the training process is purely performed on the test image, it focuses more on exploiting the inherent information of the observations instead of the external datasets. Hence, the reconstructed result usually contains sufficient image statistics of the under-studying scene. Considering its property, it is desirable to generate an initialized image by ZSL, enabling DIP with rich scene-relevant information for better prior learning. Moreover, ZSL only operates on the observations without relying on external datasets, which also satisfies the requirement of our unsupervised paradigm. Nevertheless, how to apply ZSL to the HISR task becomes another challenge that remains to be solved.

### III. PROBLEM FORMULATION

#### A. Hyperspectral Image Super-Resolution

Given the observed LrHSI  $\mathcal{Y} \in \mathbb{R}^{h \times w \times C}$  and HrMSI  $\mathcal{Z} \in \mathbb{R}^{H \times W \times c}$ , HISR aims to reconstruct the latent HrHSI  $\mathcal{X} \in \mathbb{R}^{H \times W \times C}$  with desirable spatial-spectral resolution, where  $H/h$ ,  $W/w$ , and  $C/c$  are the spatial and spectral dimensions, respectively. Note that, there usually exists a physical constraint due to the imaging mechanism, i.e.,  $w \ll W$ ,  $h \ll H$ , and  $c \ll C$ , where  $r = H/h = W/w$  denotes the scale factor. According to the linear degradation model, the observed HrMSI-LrHSI pair can be treated as the degraded version of unknown HrHSI, which is formulated as follows:

$$\begin{aligned} \mathbf{Y} &= \mathbf{P}\mathbf{X} + \mathbf{N}_y \\ \mathbf{Z} &= \mathbf{S}\mathbf{X} + \mathbf{N}_z \end{aligned} \quad (1)$$

For simplicity,  $\mathbf{X} \in \mathbb{R}^{HW \times C}$ ,  $\mathbf{Y} \in \mathbb{R}^{hw \times B}$ , and  $\mathbf{Z} \in \mathbb{R}^{HW \times b}$  are all unfolded into the matrix version from the three-order tensor.  $\mathbf{P} \in \mathbb{R}^{wh \times WH}$  denotes the spatial degradation matrix consisting of a PSF blurring kernel and a downsampling operator.  $\mathbf{S} \in \mathbb{R}^{B \times c}$  is the SRF of a multispectral sensor.  $\mathbf{N}_y$  and  $\mathbf{N}_z$  are the independent Gaussian noises.

Facing this ill-posed inverse problem, early work concentrates on handcrafting various regularizers to characterize the

potential property of the unknown HrHSI, thus forming the following optimization problem:

$$\min_{\mathbf{X}} \|\mathbf{Z} - \mathbf{S}\mathbf{X}\|_F^2 + \|\mathbf{Y} - \mathbf{P}\mathbf{X}\|_F^2 + \mathcal{R}(\mathbf{X}) \quad (2)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm, and  $\mathcal{R}(\cdot)$  indicates the designed regularizer for the image priors. However, handcrafting a proper and effective prior for a specific scene remains to be an art. Luckily, DIP allows us to model the latent priors with a well-designed generator network, thus getting rid of the manually designed regularizer. With this observation, we replace the explicit regularizer  $\mathcal{R}(\cdot)$  with the implicit prior captured by a network  $\mathbf{G}_\theta(\cdot)$ , and thereby construct the following formulation:

$$\begin{aligned} \min_{\theta} \|\mathbf{Z} - \mathbf{S}\mathbf{X}\|_F^2 + \|\mathbf{Y} - \mathbf{P}\mathbf{X}\|_F^2 \\ \text{s.t. } \mathbf{X} = \mathbf{G}_\theta(\mathbf{E}). \end{aligned} \quad (3)$$

Here,  $\mathbf{E}$  is the input of the generator network  $\mathbf{G}_\theta(\cdot)$  and  $\theta$  denotes its trainable parameters. It can be observed from (3) that reconstructing the unknown HrHSI is converted into finding the optimal parameters  $\theta$  of the generator network  $\mathbf{G}_\theta(\cdot)$ . Moreover, the optimization process purely depends on the observed HrMSI-LrHSI without external training datasets, satisfying an unsupervised paradigm. Though many works successfully apply DIP to solve the HISR task, its potential is still not fully exploited due to the following two factors. First, a randomly generated noise is usually fed into the generator network  $\mathbf{G}_\theta(\cdot)$  to recover complicated hyperspectral structures. However, it contains limited information about the under-studying scene, making it hard for the generator network to leverage the spatial-spectral attributes of observations. Second, different network structures lead to diverse priors, and thus tailor-designed architectures tend to provide more plausible reconstruction results. Considering these factors, we propose an EDIP-Net with a two-stage four-component scheme, and we will elaborate on its detailed structure in Section IV.

### IV. METHODOLOGY

#### A. Overall Scheme

In this section, we propose EDIP-Net to unleash the power of DIP for the task of HISR, and its overall pipeline is shown in Fig. 1. Specifically, our EDIP-Net adopts a two-stage scheme consisting of a ZSL stage that aims at producing two image-coarse estimations as the input of stage two and a DIG stage that focuses on modeling the image prior and generating the final HrHSI. In the first stage, we design a degradation learning network to adaptively infer the degradation parameters, i.e., PSF and SRF, and simultaneously generate paired training samples from the observed HrMSI-LrHSI correspondence. With the cross-scale spectral relationship, we further train an interactive spectral learning network with these synthetic datasets and thus generate two image-coarse estimations in a ZSL manner, providing sufficient scene-relevant information for the next stage. By feeding two estimations into the DIG stage, we construct a double U-shape network to capture the hyperspectral priors, with each subnet independently generating one HrHSI candidate. Under this premise,

a degradation-aware decision fusion strategy is proposed by resorting to the learned SRF, enabling a pixel-to-pixel fusion manner and generating a more plausible result.

### B. ZSL Stage

We adopt a three-step scheme for the ZSL stage, as shown in Fig. 2. Specifically, we find the cross-scale spectral relationship within the observations and thus design a degradation learning network to generate the training samples in the first step. With these synthetic datasets, we design an interactive spectral learning network to establish the relationship from generated LrMSIs to observed LrHSI. In the last step, we can directly generate two image-coarse estimations by applying the well-trained network to the original HrMSI. In the following section, the detailed process will be introduced.

1) *Cross-Scale Spectral Relationship*: The foundation of the ZSL stage is derived from the cross-scale spectral relationship, which is hidden within the degradation model. Specifically, a latent correlation between observed HrMSI and LrHSI can be discovered by degrading (1)

$$\mathbf{K}_1 = \mathbf{Y}\mathbf{S}, \quad \mathbf{K}_2 = \mathbf{P}\mathbf{Z}, \quad \mathbf{K}_1 = \mathbf{K}_2. \quad (4)$$

$\mathbf{K}_1 \in \mathbb{R}^{hw \times b}$  and  $\mathbf{K}_2 \in \mathbb{R}^{hw \times b}$  represent two LrMSIs with both low spatial-spectral resolutions, in which  $\mathbf{K}_1$  is the spectrally degraded LrHSI and  $\mathbf{K}_2$  is the spatially degraded HrMSI. Combining the latent correlation (4) with the degradation model (1), we can naturally derive our needed cross-scale spectral relationship

$$\begin{cases} \mathbf{Z} = \mathbf{X}\mathbf{S} \\ \mathbf{K}_1 = \mathbf{K}_2 = \mathbf{Y}\mathbf{S}. \end{cases} \quad (5)$$

This equation directly reflects the spectral relationship between the downsampled and original scale. Specifically, two LrMSIs  $\mathbf{K}_1$  and  $\mathbf{K}_2$  are spectrally degraded from LrHSI  $\mathbf{Y}$  via the degradation matrix  $\mathbf{S}$  in the spatially downsampled scale, and HrMSI  $\mathbf{Z}$  is also reduced from unknown HrHSI  $\mathbf{X}$  through the same matrix  $\mathbf{S}$  from the original scale. From the above analysis, if we successfully establish the spectral relationship from latent LrMSIs to observed LrHSI in the downsampled scale, we can recover an image-coarse estimation of the unknown HrHSI from the HrMSI. To realize our goal, we individually design the degradation learning network for training sampling generation and the interactive spectral learning network for the relationship establishment.

2) *Degradation Learning Network*: As shown in the top of Fig. 2, we design a degradation learning network to generate the training samples according to (4). Specifically, the spectral degradation represents the integral operation along the wavelength range using a multispectral SRF. In the discrete settings, each value in the degraded output is the weighted summation of the spectral vector in the corresponding location. Hence, a pointwise convolutional layer, i.e.,  $1 \times 1$  layer without any bias, is implemented to formulate the spectral transformation, which is expressed as

$$\mathbf{K}_1 = \mathcal{F}_y(\mathbf{Y}; \theta^{(\text{SRF})}) \quad (6)$$

where  $\mathcal{F}_y(\cdot; \theta^{(\text{SRF})})$  symbolizes the spectral degradation with  $\theta^{(\text{SRF})}$  acting as the SRF parameters.

Unlike spectral degradation, the spatial transform performs on each spectral band individually without interfering with the spectral dimension. Therefore, the depthwise convolution proposed by Chollet [99] is introduced to model such bandwise blurring with one shared PSF kernel, which is formulated as

$$\mathbf{K}_2 = \mathcal{F}_z(\mathbf{Z}; \theta^{(\text{PSF})}). \quad (7)$$

Here,  $\mathcal{F}_z(\cdot; \theta^{(\text{PSF})})$  denotes the spatial degradation and  $\theta^{(\text{PSF})}$  serves as the PSF parameters.

With two LrMSI  $\mathbf{K}_1$  and  $\mathbf{K}_2$ , we can generate our desirable training pairs by constructing the following loss function:

$$\mathcal{L}_1 = \|\mathbf{K}_1 - \mathbf{K}_2\|_1. \quad (8)$$

To further constrain the solution space of this ill-posed problem, we impose some physical constraints on the estimated parameters, i.e., nonnegativity and sum-to-one.

3) *Interactive Spectral Learning Network*: Once obtaining paired training samples, i.e., LrMSI-LrHSI pair, we design a simple but effective spectral learning network to establish their relationship, as shown in the middle of Fig. 2. Generally, we employ a two-stream architecture to fully leverage two generated LrMSIs. Inside the structure, multiple upsampling blocks are cascaded to progressively increase the spectral band of input LrMSIs and approach the desirable number of LrHSI. Specifically, we adopt the Res2Net module proposed by Gao et al. [100] for the upsampling part except the last one, in which the original  $3 \times 3$  kernel is changed into  $1 \times 1$  size to focus more on the spectral domain and enhance the interaction across different channels. Before feeding them into each upsampling module, we integrate the feature maps from the other stream through the addition operation, enabling them to further interact with each other and facilitate the information flow. The whole process can be simply formulated as

$$\mathbf{Y}_1, \mathbf{Y}_2 = \mathcal{F}_s(\mathbf{K}_1, \mathbf{K}_2). \quad (9)$$

Here,  $\mathcal{F}_s(\cdot)$  denotes the interactive spectral learning network, and its detailed structure can be unfolded as

$$\mathbf{Output}_{y,i} = \begin{cases} \text{Upsampling}_{y,i}(\mathbf{K}_1), & i = 1 \\ \text{Upsampling}_{y,i}(\mathbf{Output}_{y,i-1} + \mathbf{Output}_{z,i-1}), & i \neq 1 \end{cases} \quad (10)$$

$$\mathbf{Output}_{z,i} = \begin{cases} \text{Upsampling}_{z,i}(\mathbf{K}_2), & i = 1 \\ \text{Upsampling}_{z,i}(\mathbf{Output}_{z,i-1} + \mathbf{Output}_{y,i-1}), & i \neq 1 \end{cases} \quad (11)$$

where  $\text{Upsampling}_{y,i}$  and  $\text{Upsampling}_{z,i}$  represent the  $i$ th upsampling module of each stream, respectively. Besides,  $\mathbf{Output}_{y,i}$  and  $\mathbf{Output}_{z,i}$  are the output feature maps of the  $i$ th upsampling module.

With this observation, we establish our loss function in a zero-shot manner as follows:

$$\mathcal{L}_2 = \|\mathbf{Y} - \mathbf{Y}_1\|_1 + \|\mathbf{Y} - \mathbf{Y}_2\|_1. \quad (12)$$

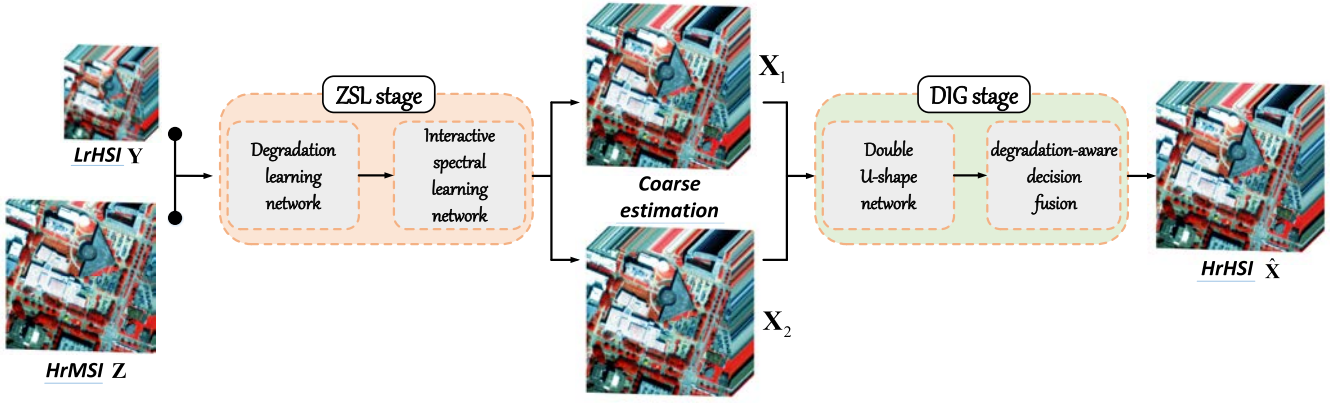


Fig. 1. Flowchart of the proposed EDIP-Net built with a two-stage four-component scheme.

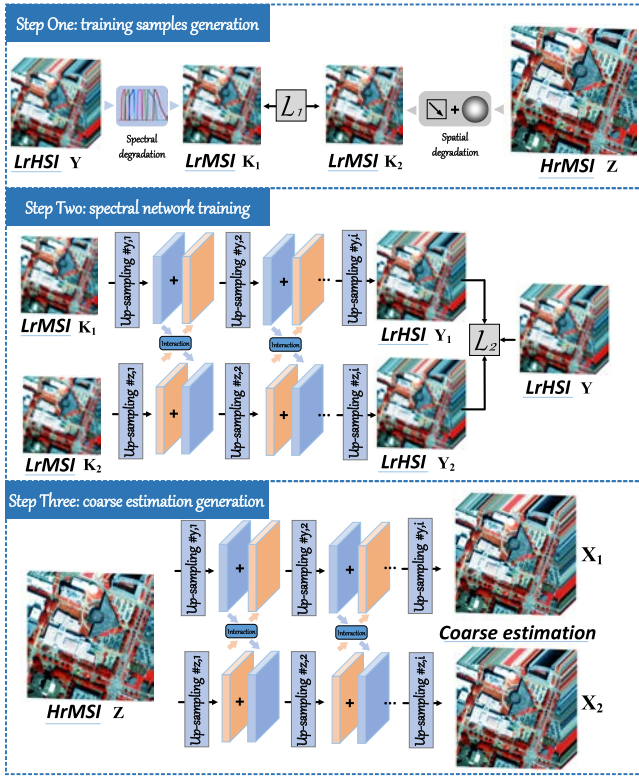


Fig. 2. Detailed structure of ZSL stage.

Once training is down, we can obtain two image-coarse estimations by applying  $\mathcal{F}_s(\cdot)$  to observed HrMSI according to the cross-scale spectral relationship (5)

$$\mathbf{X}_1, \mathbf{X}_2 = \mathcal{F}_s(\mathbf{Z}, \mathbf{Z}). \quad (13)$$

Since the training pairs are extracted from the observations themselves, the model mainly focuses on the internal characteristics of the under-studying scene and thus can produce two outcomes with sufficient scene-relevant information, which will benefit the following DIG stage.

### C. DIG Stage

1) *Double U-Shape Network*: The DIP proposed by Lempitsky et al. [30] points out that a generator network can be a

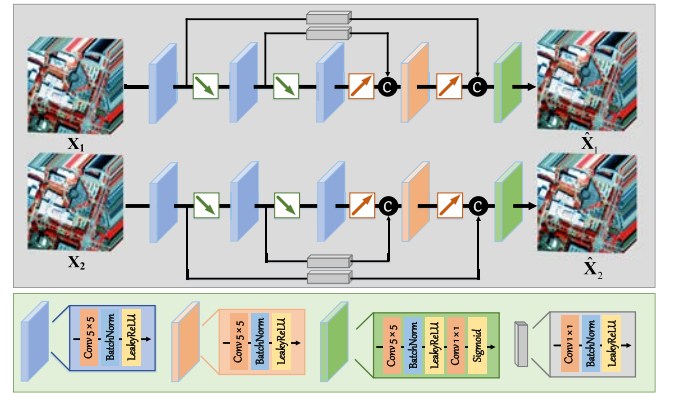


Fig. 3. Detailed structure of the double U-shape network.

powerful prior for the inverse problem. Under this premise, we propose a double U-shape network with the generated image-coarse estimation  $\mathbf{X}_1$  and  $\mathbf{X}_2$  as the inputs. As shown in Fig. 3, the whole network is a two-stream symmetric encoder-decoder structure with skip connections, which is built on the U-Net. Without loss of generality, we take  $\mathbf{X}_1$  as an example to demonstrate the working mechanism. Specifically, each stream consists of five blocks, with the first three blocks for encoding and the last two blocks for decoding. In the encoder part, three blocks with two downsampling operators are cascaded to gradually decrease the spatial size of input  $\mathbf{X}_1$ , aiming to obtain multiscale middle features. This process can be formulated as

$$\begin{cases} \mathbf{X}_{1,1} = F_{\text{en}}^1(\mathbf{X}_1) \\ \mathbf{X}_{1,2} = F_{\text{en}}^2(\mathbf{X}_{1,1} \downarrow) \\ \mathbf{X}_{1,3} = F_{\text{en}}^3(\mathbf{X}_{1,2} \downarrow) \end{cases} \quad (14)$$

where  $\downarrow$  denotes the spatial downsampling operator using the average pooling and  $F_{\text{en}}^i(\cdot)$  means the  $i$ th encoding block that contains a  $5 \times 5$  convolution layer, followed by a batch normalization and a LeakyReLU activation.

By contrast, the decoder part restores the spatial size of those downsampled features step by step under the guidance of skip connections. Specifically, unlike the traditional U-Net that directly concatenates the encoded and decoded features, we add a skip block before their fusion, which not only



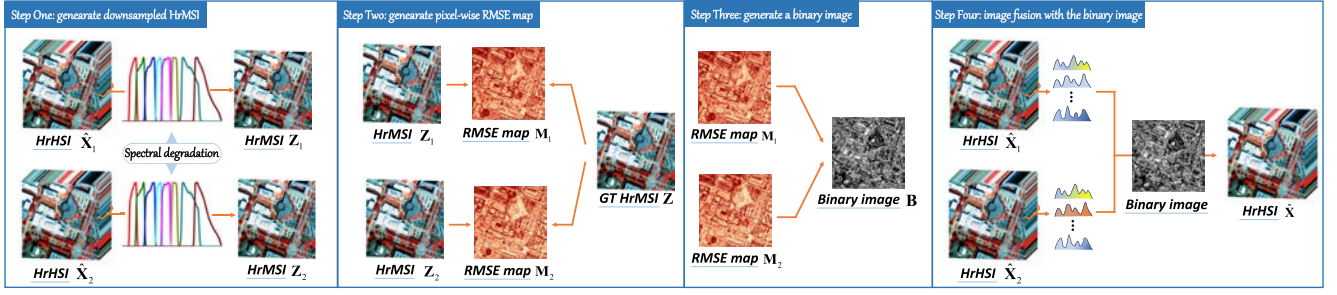


Fig. 4. Flowchart of the proposed degradation-aware decision fusion strategy.

alleviates the parameter burden by condensing its spectral band but also enhances their feature representation ability. Therefore, the decoder part is expressed as follows:

$$\begin{cases} \mathbf{X}_{1,4} = F_{\text{dn}}^1 \left( \left[ F_{\text{skip}}^1(\mathbf{X}_{1,2}), \mathbf{X}_{1,3} \uparrow \right] \right) \\ \hat{\mathbf{X}}_1 = F_{\text{dn}}^2 \left( \left[ F_{\text{skip}}^2(\mathbf{X}_{1,1}), \mathbf{X}_{1,4} \uparrow \right] \right) \end{cases} \quad (15)$$

where  $[\cdot]$  represents the concatenation operator along the spectral dimension and  $\uparrow$  is the bilinear interpolation. Besides,  $F_{\text{dn}}^i(\cdot)$  means the  $i$ th decoding block and  $F_{\text{skip}}^i(\cdot)$  is the  $i$ th skip block. Similarly, we can obtain  $\hat{\mathbf{X}}_2$  under the same architecture with  $\mathbf{X}_2$  as the input.

Since our proposed EDIP-Net is an unsupervised model without relying on external datasets, we formulate the loss function by resorting to (3) and thereby lead to the following optimization problem:

$$\begin{aligned} \mathcal{L}_3 = & \underbrace{\|\mathbf{Y} - \mathcal{F}_z(\hat{\mathbf{X}}_1; \theta^{(\text{PSF})})\|_1 + \|\mathbf{Z} - \mathcal{F}_y(\hat{\mathbf{X}}_1; \theta^{(\text{SRF})})\|_1}_{\hat{\mathbf{X}}_1} \\ & + \underbrace{\|\mathbf{Y} - \mathcal{F}_z(\hat{\mathbf{X}}_2; \theta^{(\text{PSF})})\|_1 + \|\mathbf{Z} - \mathcal{F}_y(\hat{\mathbf{X}}_2; \theta^{(\text{SRF})})\|_1}_{\hat{\mathbf{X}}_2}. \end{aligned} \quad (16)$$

The above loss term compels the generated  $\hat{\mathbf{X}}_1$  and  $\hat{\mathbf{X}}_2$  to follow the degradation model (1), i.e., the spectrally and spatially degraded outcomes should close to HrMSI  $\mathbf{Z}$  and LrHsi  $\mathbf{Y}$ , respectively. Though the parameters of SRF and PSF are hard to acquire, we directly utilize the estimated ones in the ZSL stage to realize the two transformations.

2) *Degradation-Aware Decision Fusion*: Instead of simply calculating the average of  $\hat{\mathbf{X}}_1$  and  $\hat{\mathbf{X}}_2$ , we propose a more advanced strategy to realize a pixel-to-pixel fusion. Fundamental to our strategy is the fact that the better the reconstruction result, the closer its spectral degradation result is to HrMSI. With this observation, we design a four-step fusion scheme, as shown in Fig. 4. First, two HrMSI  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  are generated by spectrally downsampling the HrHsi candidates, i.e.,  $\hat{\mathbf{X}}_1$  and  $\hat{\mathbf{X}}_2$ , which can be formulated as

$$\begin{cases} \mathbf{Z}_1 = \mathcal{F}_y(\hat{\mathbf{X}}_1; \theta^{(\text{SRF})}) \\ \mathbf{Z}_2 = \mathcal{F}_y(\hat{\mathbf{X}}_2; \theta^{(\text{SRF})}). \end{cases} \quad (17)$$

With two synthetic HrMSI  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$ , we continue to calculate their root-mean-square error (RMSE) map with ground truth

(GT) HrMSI  $\mathbf{Z}$  in a pixelwise manner, leading to the following process:

$$\begin{cases} \mathbf{M}_1 = \mathcal{F}_M(\mathbf{Z}_1, \mathbf{Z}) \\ \mathbf{M}_2 = \mathcal{F}_M(\mathbf{Z}_2, \mathbf{Z}) \end{cases} \quad (18)$$

where  $\mathbf{M}_1$  and  $\mathbf{M}_2$  are the corresponding RMSE map of  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$ , respectively, and the calculation process is simplified by  $\mathcal{F}_M(\cdot)$ . Once obtaining two error maps, we further generate a binary image by comparing their error values in each pixel location, which is expressed as follows:

$$\mathbf{B}(i, j) = \begin{cases} 0, & \mathbf{M}_1(i, j) \geq \mathbf{M}_2(i, j) \\ 1, & \mathbf{M}_1(i, j) < \mathbf{M}_2(i, j) \end{cases} \quad (19)$$

where  $\mathbf{B}(i, j)$ ,  $\mathbf{M}_1(i, j)$ , and  $\mathbf{M}_2(i, j)$  represent the value on the corresponding pixel coordinate of  $(i, j)$ . Under the guidance of  $\mathbf{B}$ , we can effectively integrate the optimal results of two HrHsi candidates  $\hat{\mathbf{X}}_1$  and  $\hat{\mathbf{X}}_2$  in a pixel-to-pixel manner, and the specific operation can be expressed as

$$\hat{\mathbf{X}}(i, j) = \begin{cases} \hat{\mathbf{X}}_1(i, j), & \mathbf{B}(i, j) = 1 \\ \hat{\mathbf{X}}_2(i, j), & \mathbf{B}(i, j) = 0 \end{cases} \quad (20)$$

where  $\hat{\mathbf{X}}(i, j)$ ,  $\hat{\mathbf{X}}_1(i, j)$ , and  $\hat{\mathbf{X}}_2(i, j)$  denote the spectral vector in the location of  $(i, j)$ . In such a straightforward fusion strategy, we can finally obtain the fused result  $\hat{\mathbf{X}}$ .

## V. EXPERIMENTS

In this section, the benchmark used for the evaluation is first introduced, and then, an extensive ablation study is conducted to evaluate the effectiveness of our proposed components. Besides, the fused results from different methods are compared using visual and quantitative measures, demonstrating our competitive performance. Last, robustness analysis is performed to verify our stability in challenging cases.

### A. Benchmark

1) *Datasets*: Four publicly available datasets, including the Houston dataset,<sup>1</sup> Washington DC Mall dataset,<sup>2</sup> TianGong-1 (TG) dataset,<sup>3</sup> and Chikusei dataset<sup>4</sup>, are adopted for the simulation experiments. Each of them goes through a preprocessing

<sup>1</sup><https://hyperspectral.ee.uh.edu/>

<sup>2</sup><https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>

<sup>3</sup><https://www.msadc.cn/dataHome>

<sup>4</sup><http://naotoyokoya.com/Download.html>

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

stage for fair quality evaluation, including low-quality band elimination and region cropping. The processed datasets serve as the GT image for reference and the input HrMSI-LrHSI pair is generated through the predefined blurring kernel and SRF from WorldView 2 multispectral sensor. The specifications of simulated results are provided in Table I.

2) *Comparison Methods and Evaluation Metrics*: We select nine typical fusion methods for experiments, covering existing mainstream categories. Specifically, G-SOMP+ [101], CSU [102], CNMF [103], STEREO [104], CSTF [105], and SCOTT [106] are from the model-based category, while ADASR [79], MIAE [73], and SURE [107] fall within the scope of deep learning.

Two kinds of evaluation manners are adopted to comprehensively assess the fused results. On the one hand, six metrics are selected to quantitatively evaluate different methods in terms of the spatial, spectral, and global qualities, including correlation coefficient (CC), RMSE, peak signal-to-noise ratio (PSNR) [108], universal image quality index (UIQI) [109], [110], relative dimensionless global error in synthesis (ERGAS), and spectral angle mapper (SAM). For PSNR, UIQI, and CC, a higher value means better fusion performance, while the ideal value for RMSE, ERGAS, and SAM is zero. On the other hand, three kinds of error maps are utilized to visually assist the quantitative evaluation, i.e., SAM heatmap, mean relative absolute error (MRAE) heatmap, and residual heatmap.

3) *Implementation Details*: Our proposed EDIP-Net is built with a two-stage framework, as shown in Algorithm 1. Specifically, it is optimized via Adam optimizer with a linear function decay scheme. The initial learning rate is set as  $1 \times 10^{-3}$ ,  $4 \times 10^{-3}$ , and  $4 \times 10^{-3}$  for our three subnets with a training epoch of 2000, 2000, and 7000, respectively, and then, each of them linearly decreases to zero until the given epochs.

### B. Degradation Learning Network

In the stage of ZSL, we design a simple degradation learning network to infer the unknown PSF and SRF parameters, thus generating paired training samples for the spectral learning network. In this section, we choose the Houston dataset as an example to visually demonstrate their accuracy. As mentioned earlier, spatial degradation is performed to simulate the LrHSI by using a  $8 \times 8$  Gaussian PSF with a standard deviation equating to 3.4, and spectral degradation is conducted via the eight-band WorldView 2 multispectral sensor. As shown in Fig. 5, the GT and estimated degradation parameters from our method and ADASR are depicted in the top, middle, and bottom parts, respectively. It can be observed that our

### Algorithm 1 Proposed EDIP-Net Algorithm

**Input:** Observed LrHSI  $\mathbf{Y}$  and HrMSI  $\mathbf{Z}$ .

**Procedure:**

**Zero-shot learning stage:**

- Obtain  $\mathbf{K}_1$  and  $\mathbf{K}_2$  by optimizing  $\mathcal{L}_1$
- Obtain two image-coarse estimates  $\mathbf{X}_1$  and  $\mathbf{X}_2$  by optimizing  $\mathcal{L}_2$

**Deep image generation stage:**

- Obtain two HrHSI candidates  $\hat{\mathbf{X}}_1$  and  $\hat{\mathbf{X}}_2$  by optimizing  $\mathcal{L}_3$ ;
- Obtain the final HrHSI  $\hat{\mathbf{X}}$ .

**End Procedure**

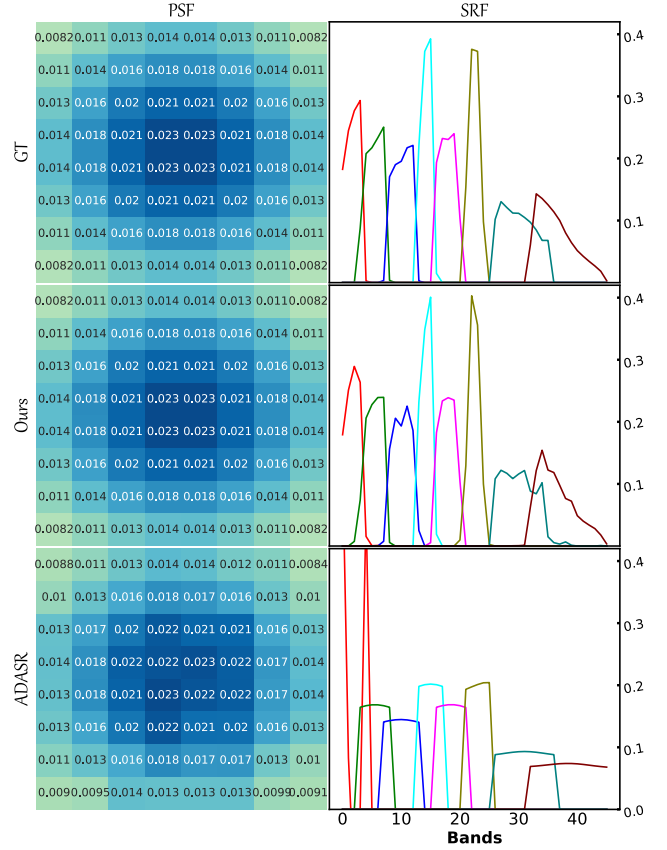


Fig. 5. GT and estimated PSF and SRF parameters in the Houston dataset.

estimated PSF satisfies the Gaussian distribution, showing higher precision compared with ADASR. By contrast, a large deviation can be observed in the estimated SRF from ADASR, especially at the peak of each curve. However, our prediction still exhibits a similar shape to the true ones.

### C. Ablation Study

In this section, some necessary ablation study is performed to demonstrate the effectiveness of our proposed key components, in which the Houston dataset is taken as an example.

1) *With and Without PSF and SRF Provided*: PSF and SRF parameters, acting as a bridge between observations and unknowns, play an important role in the HISR task. When these degradation parameters are provided as known priors,



TABLE II  
QUANTITATIVE EVALUATION IN THE HOUSTON DATASET  
WITH AND WITHOUT PSF AND SRF PROVIDED

	Outcomes	SAM	PSNR	ERGAS	CC	RMSE	UIQI
With PSF and SRF Provided	$\hat{\mathbf{X}}_1$	0.8084	48.7601	0.2065	0.9994	0.0028	0.9998
	$\hat{\mathbf{X}}_2$	0.8334	48.2191	0.2183	0.9993	0.0030	0.9998
	$\hat{\mathbf{X}}$	0.7944	49.1987	0.2027	0.9994	0.0028	0.9998
Without PSF and SRF Provided	$\hat{\mathbf{X}}_1$	0.8036	48.8817	0.2051	0.9994	0.0028	0.9998
	$\hat{\mathbf{X}}_2$	0.8356	48.3792	0.2163	0.9993	0.0030	0.9998
	$\hat{\mathbf{X}}$	0.7928	49.3325	0.2010	0.9994	0.0028	0.9998

TABLE III  
ABLATION STUDY OF ZSL STAGE IN THE HOUSTON DATASET

	Outcomes	SAM	PSNR	ERGAS	CC	RMSE	UIQI
Noise	$\hat{\mathbf{X}}_1$	1.8899	30.6319	1.3404	0.9802	0.0212	0.9976
	$\hat{\mathbf{X}}_2$	2.3320	28.8541	1.6501	0.9702	0.0255	0.9953
	$\hat{\mathbf{X}}$	1.7026	31.6921	1.1908	0.9842	0.0188	0.9981
ZSL	$\hat{\mathbf{X}}_1$	0.8036	48.8817	0.2051	0.9994	0.0028	0.9998
	$\hat{\mathbf{X}}_2$	0.8356	48.3792	0.2163	0.9993	0.0030	0.9998
	$\hat{\mathbf{X}}$	0.7928	49.3325	0.2010	0.9994	0.0028	0.9998

the task is converted into an ideal one. By contrast, it is more challenging to tackle a situation without knowing their information, which is usually encountered in real applications. Considering this reality, we design a simple degradation learning network to adaptively infer these parameters. To verify its performance, we simply remove this subnet and provide EDIP-Net with GT PSF and SRF for an experiment. The results of three HrHSI outcomes are shown in Table II. By comparing their quantitative performance, we can find that the fluctuation caused by the parameter estimation is not significant and our version without knowing the degradation even achieves slightly better results in some of the metrics, which firmly demonstrates the stability of the proposed EDIP-Net in the challenging situation.

2) *ZSL Stage*: Different from existing DIP-based methods that treat a randomly generated noise as the input, we propose a ZSL stage to initially generate two image-coarse estimations, which can facilitate the prior modeling of DIP by leveraging sufficient scene-relevant information. In this section, we conduct another experiment by directly feeding the random noise into the DIG stage, aiming to verify the effectiveness of the proposed ZSL stage. As can be seen in Table III, there is a huge decline in all metrics without the assistance of the ZSL stage, which reflects its improvement in fusion performance. Moreover, we also illustrate their visual results in Fig. 6. The outputs generated from the noise input fail to reconstruct the high-quality HrHSI and lead to noticeable distortions in the edge of each image. By contrast, the ZSL stage is capable of providing rich scene prior and promoting the DIP to model the relevant statistics, which successfully improves the reconstruction quality.

3) *Degradation-Aware Decision Fusion*: Instead of simply averaging two HrHSI candidates  $\hat{\mathbf{X}}_1$  and  $\hat{\mathbf{X}}_2$ , we propose a straightforward but effective strategy to integrate their information. In this section, we exhibit the results of some key outcomes in the Houston dataset to intuitively illustrate its

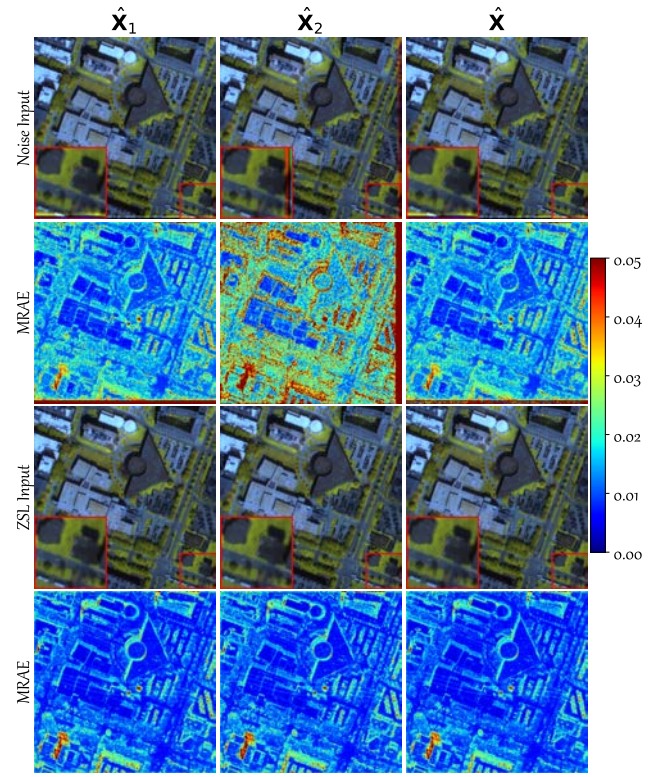


Fig. 6. Outputs and MRAE heatmaps of some key outcomes with and without the ZSL stage.

effectiveness. As shown in Table IV, we first calculate the quantitative metrics of five HrHSIs generated in the training process, including two image-coarse estimations  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , two HrHSI candidates  $\hat{\mathbf{X}}_1$  and  $\hat{\mathbf{X}}_2$ , and the final result  $\hat{\mathbf{X}}$ . Observing  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , we can find that despite their relatively low accuracy compared with other state-of-the-art methods, they still provide sufficient scene-relevant information. Moreover, their visual results shown in Fig. 7 further represent their abundant spatial details and spectral features. With the two estimations as the input, the double U-shape network can fully leverage the information of the under-studying scene and better model the complicated hyperspectral priors, thus producing two more satisfactory HrHSI candidates  $\hat{\mathbf{X}}_1$  and  $\hat{\mathbf{X}}_2$  with less spatial-spectral distortion. Since two candidates exhibit different reconstruction performances in the same area, we select the optimal result according to their degraded outcomes and therefore generate the final  $\hat{\mathbf{X}}$ . It can be observed that  $\hat{\mathbf{X}}$  yields the best scores among all metrics, especially PSNR value, which convincingly demonstrates the effectiveness of our proposed fusion strategy.

#### D. Comparison With the State-of-the-Art Methods

Nine state-of-the-art methods, ranging from model-based to deep learning categories, are selected as competitors to verify the superiority of our proposed EDIP-Net. The quantitative and visual results are illustrated in Tables V–VIII and Figs. 8–12, respectively.

1) *Houston*: Fig. 8 shows the visual results and corresponding error maps in the Houston dataset. It can be

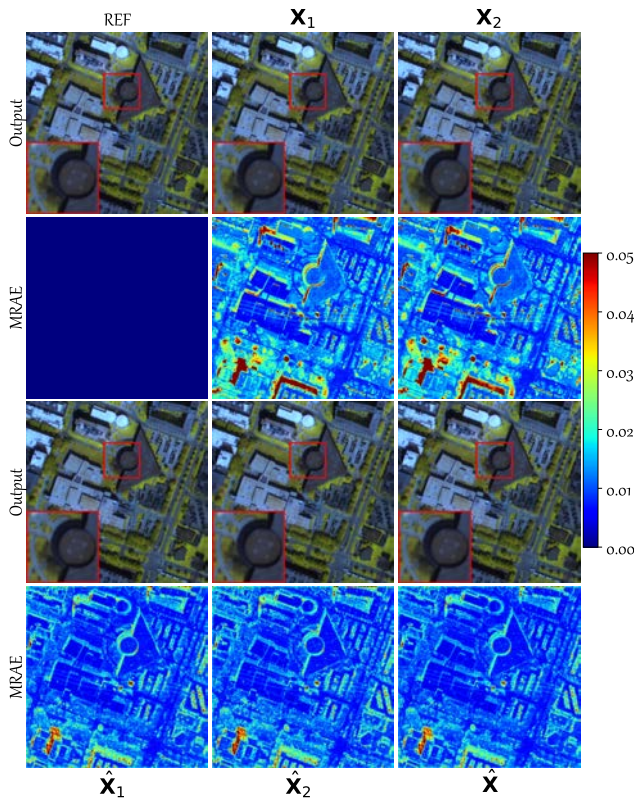


Fig. 7. Outputs and MRAE heatmaps of some key outcomes.

TABLE IV

ABLATION STUDY OF THE FUSION STRATEGY IN THE HOUSTON DATASET

Outcomes	SAM	PSNR	ERGAS	CC	RMSE	UIQI
$X_1$	1.0859	46.0369	0.3352	0.9987	0.0040	0.9993
$X_2$	1.0915	45.4356	0.3257	0.9987	0.0041	0.9993
$\hat{X}_1$	0.8036	48.8817	0.2051	<b>0.9994</b>	<b>0.0028</b>	<b>0.9998</b>
$\hat{X}_2$	0.8356	48.3792	0.2163	0.9993	0.0030	<b>0.9998</b>
$\hat{X}$	<b>0.7928</b>	<b>49.3325</b>	<b>0.2010</b>	<b>0.9994</b>	<b>0.0028</b>	<b>0.9998</b>

TABLE V

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
ADASR	1.1706	46.3849	0.3005	0.9987	0.0041	0.9994
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
Ours	<b>0.7928</b>	<b>49.3325</b>	<b>0.2010</b>	<b>0.9994</b>	<b>0.0028</b>	<b>0.9998</b>

observed that all fused results can provide more spatial details compared with the original LrHSI. Despite their close visual appearance, more subtle differences can be found in the error map. For model-based methods, their results generally

suffer from distortion in spatial and spectral dimensions, especially in the shaded areas. However, CSTF exhibits a relatively strong ability in the spectral reconstruction. For DL-based methods, some noticeable errors can be observed in the results of ADASR, especially in the lower left corner. SURE achieves the closest results to our EDIP-Net, but their reconstruction results in shadows are not as ideal as ours. Compared with all competitors, the proposed EDIP-Net produces superior performance in terms of spatial and spectral reconstruction, which is validated in the error maps.

The quantitative results presented in Table V further verify our superiority. Specifically, the proposed EDIP-Net reaches the best performance in all metrics, showing powerful reconstruction ability in spatial enhancement and spectral fidelity. At the same time, SURE achieves the second-best result and outperforms MIAE and ADASR. Moreover, CSTF and STEREO acquire the highest SAM and PSNR values among the traditional methods, respectively. Moreover, we also plot the PSNR value along different spectral bands in Fig. 12, which directly reflects our leading position in most reconstructed bands.

2) *Washington DC Mall*: The visual results and corresponding error maps of each method are shown in Fig. 9. Unsurprisingly, our method gives the smallest reconstruction error when compared with other competitors, especially in the building areas. At the same time, SURE, MIAE, and ADASR generally produce better results than traditional methods even though noticeable distortion can be observed in the vegetation area from ADASR. As for these traditional ones, STEREO, CSTF, and SCOTT perform slightly better than the other three matrix decomposition methods.

The quantitative evaluation in Table VI also supports the previous conclusions. Our proposed EDIP-Net achieves the best results in terms of all metrics, showing significant leading advantages in both spatial and spectral domains. It is worth mentioning that SURE, MIAE, and ADASR also achieve competitive results due to the capability of deep networks. However, other traditional methods generally fail to produce satisfactory outcomes partly because of their limited representation ability, and only SCOTT obtains relatively high scores. Besides, the PSNR value shown in Fig. 12 also demonstrates our spatial reconstruction ability, with the highest scores in the overall bands.

3) *TianGong-1*: The reconstruction results are illustrated in Fig. 10 to intuitively exhibit their differences. There exist noticeable distortions in the traditional methods, which fail to give a satisfactory result, especially in the building area. By contrast, deep learning-based methods can better restore the spatial details while preserving original spectral information. On the whole, our proposed EDIP-Net achieves stable and reliable results across different regions, which directly verifies our superiority in image fusion.

We list the quantitative results in Table VII to exhibit their reconstruction performance. Our method outperforms other competitors in all metrics, which firmly demonstrates our strong ability in spatial enhancement and spectral preservation. Moreover, MIAE obtains the suboptimal result, followed



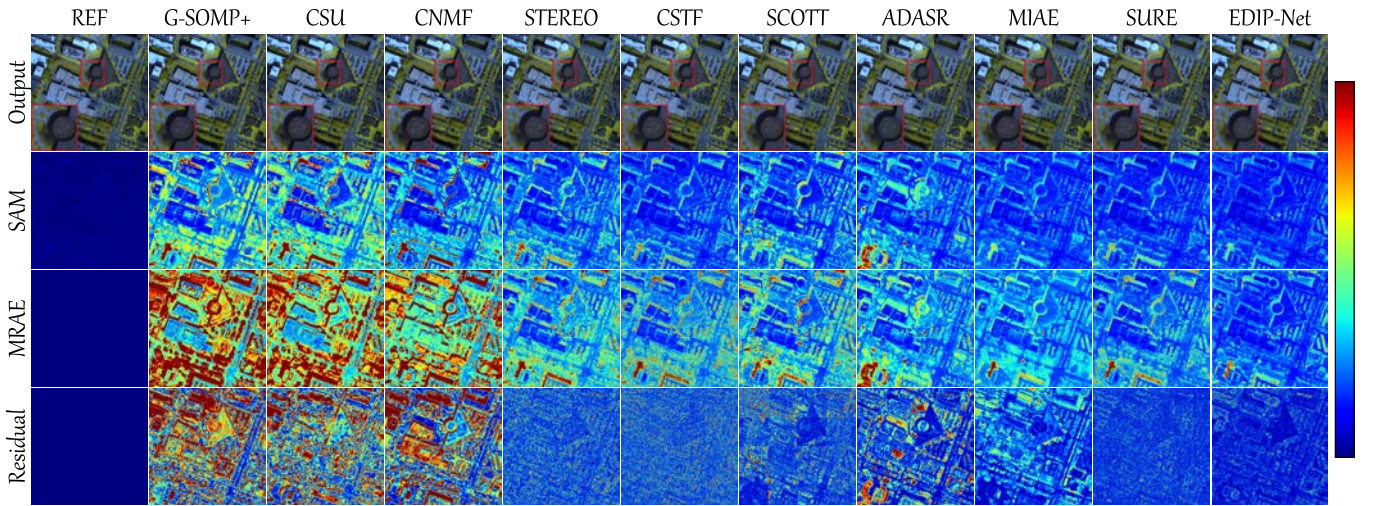


Fig. 8. Visual comparison of all methods in the Houston dataset. (First row) Pseudo-color image of SR outputs (R:46, G:30, B:14). (Second row) Heatmap of SAM error. (Third row) Heatmap of MRAE. (Fourth row) Residual heatmap at band 24.

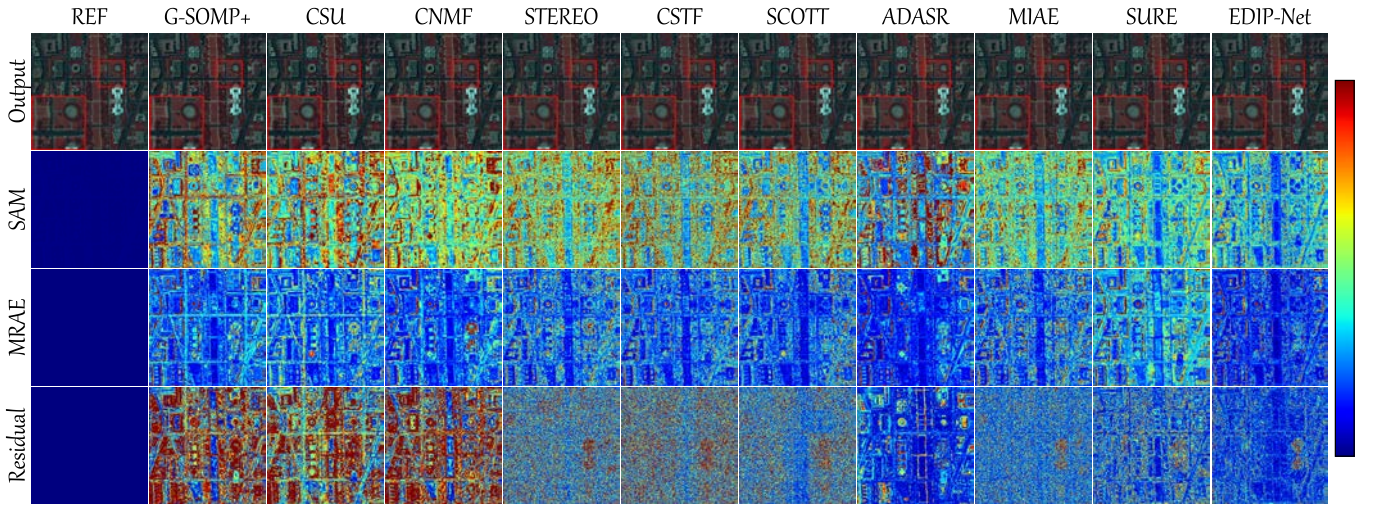


Fig. 9. Visual comparison of all methods in the Washington DC Mall dataset. (First row) Pseudo-color image of SR outputs (R:60, G:27, B:17). (Second row) Heatmap of SAM error. (Third row) Heatmap of MRAE. (Fourth row) Residual heatmap at band 16.

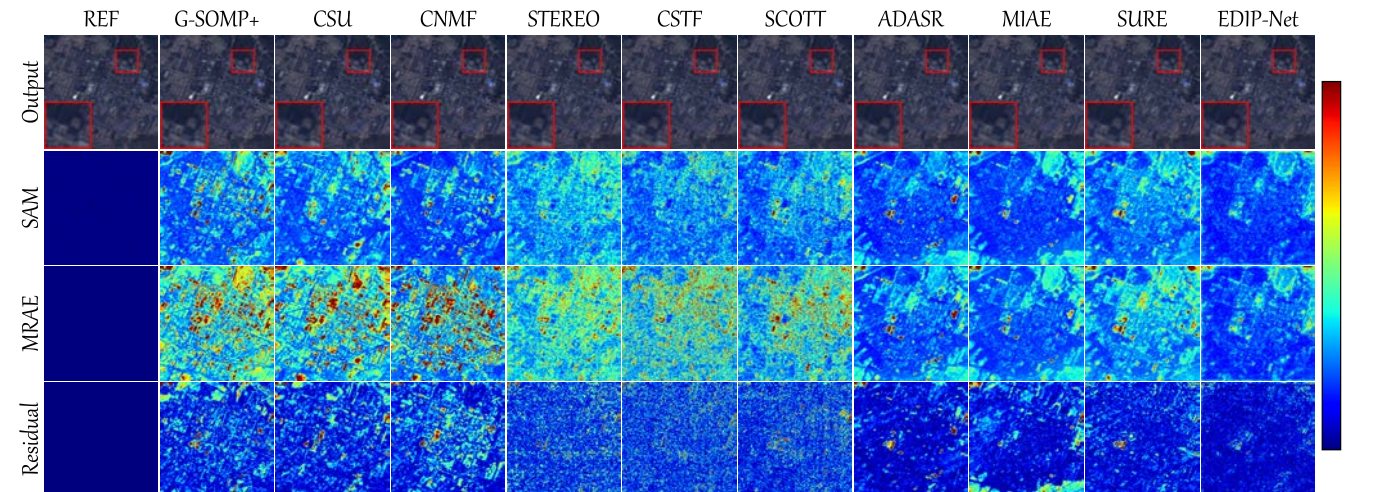


Fig. 10. Visual comparison of all methods in the TG dataset. (First row) Pseudo-color image of SR outputs (R:30, G:20, B:7). (Second row) Heatmap of SAM error. (Third row) Heatmap of MRAE. (Fourth row) Residual heatmap at band 44.

by ADASR and SURE. At the same time, other traditional methods are generally inferior to the abovementioned ones, in which tensor representation achieves more reliable results than matrix decomposition. Moreover, the PSNR curve in



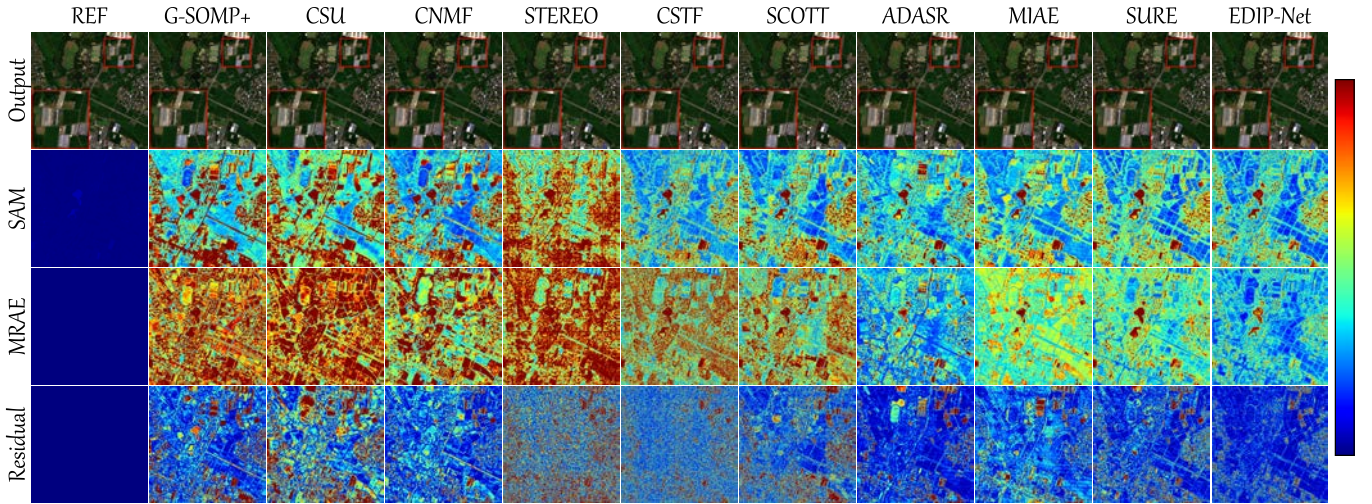


Fig. 11. Visual comparison of all methods in the Chikusei dataset. (First row) Pseudo-color image of SR outputs (R:57, G:37, B:17). (Second row) Heatmap of SAM error. (Third row) Heatmap of MRAE. (Fourth row) Residual heatmap at band 7.

TABLE VI

QUANTITATIVE PERFORMANCE IN WASHINGTON DC MALL DATASET

Method	SAM	PSNR	ERGAS	CC	RMSE	UIQI
G-SOMP+	3.1131	33.2740	1.8321	0.9642	0.0074	0.9804
CSU	3.0727	33.4366	1.7942	0.9646	0.0075	0.9825
CNMF	3.0457	33.3241	1.9946	0.9538	0.0068	0.9803
STEREO	3.0152	34.2118	2.0439	0.9528	0.0055	0.9745
CSTF	3.1262	34.2838	1.7547	0.9638	0.0065	0.9838
SCOTT	2.8191	34.5812	1.7959	0.9628	0.0058	0.9802
ADASR	2.9305	35.7708	2.4451	0.9427	0.0075	0.9735
MIAE	2.6461	35.2858	2.2137	0.9474	0.0046	0.9698
SURE	2.2931	35.1149	5.3071	0.9434	0.0040	0.9486
Ours	<b>2.0441</b>	<b>37.1012</b>	<b>1.7315</b>	<b>0.9648</b>	<b>0.0037</b>	<b>0.9856</b>

TABLE VII

QUANTITATIVE PERFORMANCE IN TIANGONG-1 DATASET

Method	SAM	PSNR	ERGAS	CC	RMSE	UIQI
G-SOMP+	1.4152	42.9887	0.3375	0.9909	0.0056	0.9988
CSU	1.2499	43.0198	0.3353	0.9909	0.0055	0.9990
CNMF	1.0916	43.6647	0.3074	0.9952	0.0051	0.9989
STEREO	1.4409	45.5809	0.2623	0.9928	0.0042	0.9993
CSTF	1.2765	45.3448	0.2667	0.9929	0.0042	0.9993
SCOTT	1.2721	45.5515	0.2590	0.9932	0.0041	0.9993
ADASR	1.0836	47.9547	0.2119	0.9952	0.0034	0.9995
MIAE	1.0646	49.3634	0.1889	0.9965	0.0029	0.9996
SURE	1.2525	47.0820	0.2123	0.9947	0.0035	0.9995
Ours	<b>0.9543</b>	<b>50.0133</b>	<b>0.1528</b>	<b>0.9968</b>	<b>0.0026</b>	<b>0.9998</b>

TABLE VIII

QUANTITATIVE PERFORMANCE IN CHIKUSEI DATASET

Method	SAM	PSNR	ERGAS	CC	RMSE	UIQI
G-SOMP+	2.2594	40.0128	0.6842	0.9924	0.0086	0.9937
CSU	2.3780	38.4987	0.8128	0.9866	0.0093	0.9894
CNMF	1.8029	42.3297	0.4825	0.9955	0.0075	0.9941
STEREO	2.6728	43.0539	0.6270	0.9892	0.0073	0.9914
CSTF	1.4870	44.2189	0.4503	0.9946	0.0052	0.9963
SCOTT	1.6273	44.5831	0.5491	0.9920	0.0054	0.9958
ADASR	1.4052	46.1057	0.4291	0.9945	0.0047	0.9982
MIAE	1.4049	46.4614	0.4070	0.9959	0.0054	0.9974
SURE	1.4839	46.4847	0.3720	0.9958	0.0042	0.9980
Ours	<b>1.3322</b>	<b>48.5270</b>	<b>0.3168</b>	<b>0.9967</b>	<b>0.0038</b>	<b>0.9990</b>

achieved by these deep learning-based methods, and fewer reconstruction errors exist in the boundary and farmland areas. Among all of them, our proposed EDIP-Net still gives more accurate fusion results with the lowest error map.

The evaluation metrics reported in Table VIII prove the conclusions made in the visual inspection. Specifically, EDIP-Net still gets the first place in all indices, showing an advantage in the PSNR value. Also, the other three deep learning-based methods give competitive results with relatively high scores. However, traditional methods generally suffer from serious distortion and struggle to restore the original spatial structure and spectral information. On the other hand, Fig. 12 shows the PSNR value of each spectral band, which firmly demonstrates the reliability of our proposed method in the spatial domain.

Fig. 12 further indicates our competitive results in the spatial reconstruction.

4) *Chikusei*: The visual performance is displayed in Fig. 11. As we can see, traditional methods show limited ability in spatial recovery and spectral preservation, especially in the building areas. On the contrary, more reliable results can be

#### E. Analysis of Model Complexity

Table IX analyzes the model complexity of DL-based methods in the TG dataset, including running time, model parameter, and floating-point operations (FLOPs). It can be observed that the double U-shape network leads to a heavy computational burden due to the two isolated U-Net structures.

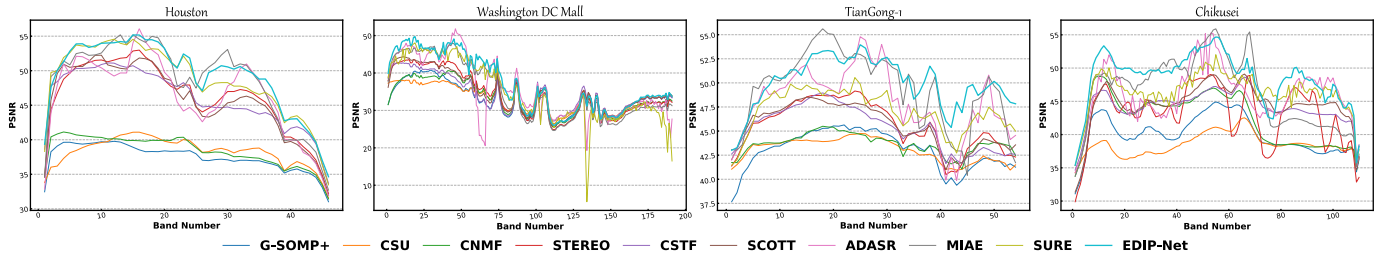


Fig. 12. PSNR value along the spectral band in four datasets.

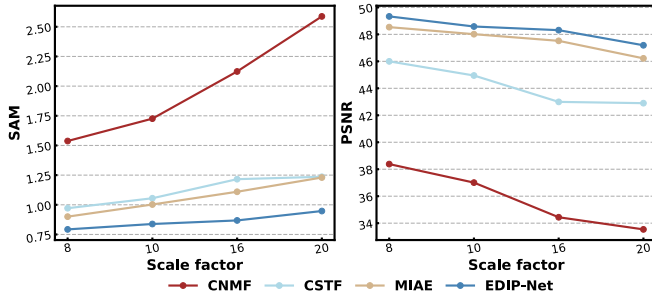


Fig. 13. PSNR and SAM metrics under the scale factor of 8, 10, 16, and 20 in the Houston dataset.

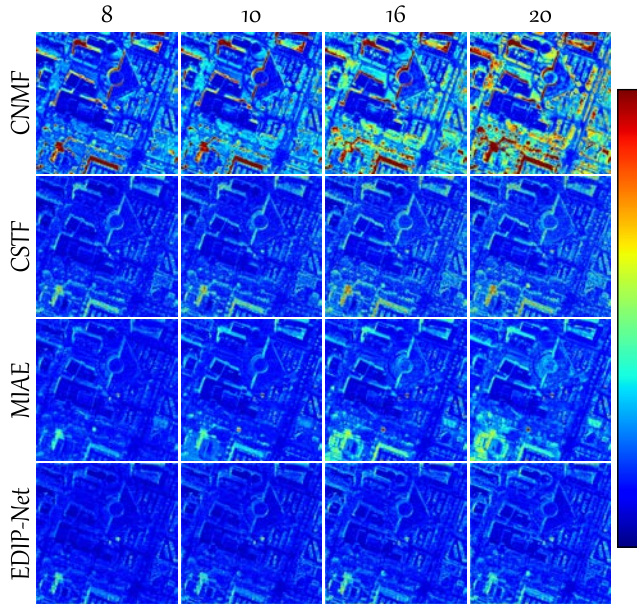


Fig. 14. SAM heatmap under four different scale factors in the Houston dataset.

However, the other three parts of our model show competitive results, with the lowest model complexity. Considering the fusion quality, our proposed EDIP-Net still falls within an acceptable range.

#### F. Robustness Analysis

In addition to the conventional evaluation methods, we examine their performance in more challenging conditions. First, we test our robustness against different scale factors, i.e., 8, 10, 12, and 16, in the Houston dataset. Then, we further apply our EDIP-Net in a real-world dataset to verify its application potential.

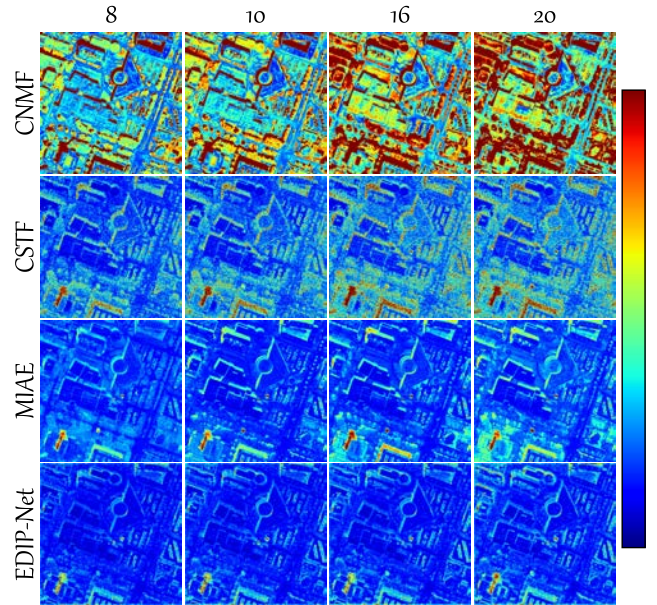


Fig. 15. MRAE heatmap under four different scale factors in the Houston dataset.

1) *Scale Factor*: In this section, three representative methods are selected as competitors, including CNMF, CSTF, and MIAE. From Fig. 13, we can observe that the increase of the scale factor harms the reconstruction result of all methods. Specifically, CNMF is more sensitive to scale variation and falls behind other methods by a large margin. Deep learning-based methods perform relatively stable in different scale settings, in which our proposed EDIP-Net outperforms all competitors in SAM and PSNR metrics.

Moreover, SAM and MRAE error maps of each method are displayed in Figs. 14 and 15, respectively. With the rise of the scale factor, the error of all methods gradually increases, especially in the shaded area. Despite the inevitable influence caused by scale variation, our method still gives the lowest reconstruction error in all cases, which firmly reflects our stability under different circumstances.

2) *Real-World Dataset*: To evaluate the robustness of our method on real-world data, we conduct experiments in one public dataset called Liao Ning-01 [111].<sup>5</sup> Specifically, it contains an LrHSI of size  $300 \times 300 \times 166$  and an HrMSI of size  $900 \times 900 \times 8$ , which is captured by the ZY-1 02D satellite over Dalian, Liaoning, China. Here, a local area

<sup>5</sup><https://drive.google.com/drive/folders/1JLCCB6ld5R49HDLN5SsMISx1d0fuqRjO>



TABLE IX  
MODEL COMPLEXITY OF DL-BASED METHODS IN THE TG DATASET

Method	EDIP-Net				ADASR	MIAE	SURE
	Degradation learning network	Interactive spectral learning network	Double U-shape network	Degradation-aware decision fusion			
Running Time (S)	16	132	5731	2	265	730	33458
Parameters (K)	0.57	17.58	12251.17	\	14673.26	203.76	195.42
FLOPs (G)	0.0016	0.0067	300.09	\	0.32	11.30	44.85

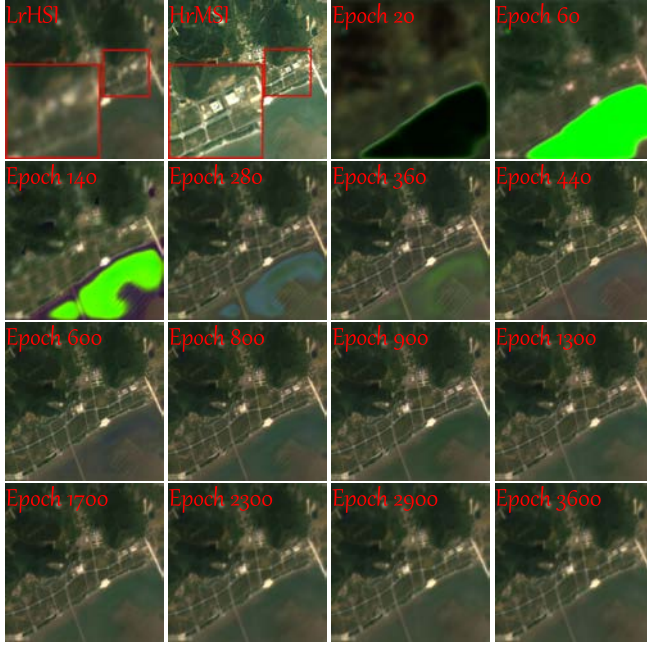


Fig. 16. Observed HrMSI–LrHSI input and corresponding fusion results under different training epochs in Liao Ning-01 dataset.

consisting of  $100 \times 100 \times 147$  LrHSI and  $300 \times 300 \times 8$  HrMSI is cropped for studying. Considering the unavailability of GT HrHSI, we display the observed HrMSI–LrHSI pair and our fusion results in Fig. 16. Moreover, the middle outputs in different training epochs are selected to intuitively exhibit the reconstruction process. It can be observed that our method fails to give a satisfactory result in the early training epochs, and noticeable distortion exists in the water areas. As the training continues, the color difference and spatial blurring gradually disappear. At the epoch of 800, our method achieves the best visual quality, containing similar spectral features to LrHSI and abundant spatial details as HrMSI. However, artifacts begin to occur when training continues, where the entire image appears hazy and the details become blurry. With this observation, our proposed EDIP-Net exhibits the potential ability to tackle real-world data under careful parameter settings.

## VI. CONCLUSION

Despite the remarkable achievements made from the supervised paradigm, the intrinsic data-intensive manner becomes an obstacle to its further application in the HISR. Considering this drawback, we propose an EDIP-Net to realize unsupervised image generation without external data assistance. To further unleash the potential of DIP for the HISR task, our model is built with a two-stage framework with a ZSL

stage to initially generate image-coarse estimations and a DIG stage for final image reconstruction. Specifically, we create paired training samples from the input observations by a degradation learning network supported by a cross-scale spectral relationship. With these samples, we further train an interactive spectral learning network to generate two image-coarse estimations in a ZSL manner. By feeding two estimations into the DIG stage, a double U-shape network can capture their latent hyperspectral prior to produce two HrHSI candidates. Under this premise, a degradation-aware decision fusion strategy is proposed to integrate the optimal results in a pixel-to-pixel manner. We perform extensive ablation experiments to demonstrate the effectiveness of proposed key components. Besides, nine state-of-the-art methods ranging from traditional to deep learning-based category are chosen to verify our superior fusion performance in four public datasets. Last, robustness analysis is conducted to examine our potential in more challenging cases.

Despite these encouraging results, there still exist some limitations that hinder our further development. First, the model performance is susceptible to noise interference, leading to limited robustness under adverse circumstances. Second, the heavy computational cost becomes an obstacle for resource-constrained environments. Therefore, designing noise-resistant and lightweight models will be the focus of future research.

## ACKNOWLEDGMENT

The authors would like to thank all the researchers who kindly provide their code and datasets, especially Yao Liu and Prof. Haoyang Yu for providing the raw ZY-1 02D data and Liao Ning-01 dataset.

## 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 et al., "Deep feature aggregation network for hyperspectral anomaly detection," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–16, 2024.
- [3] Z. Han et al., "Multisource collaborative domain generalization for cross-scene remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5535815.
- [4] Y. Wen, T. Gao, J. Zhang, Z. Li, and T. Chen, "Encoder-free multiaxis physics-aware fusion network for remote sensing image dehazing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4705915.
- [5] H. Guo, Q. Shi, A. Marinoni, B. Du, and L. Zhang, "Deep building footprint update network: A semi-supervised method for updating existing building footprint from bi-temporal remote sensing images," *Remote Sens. Environ.*, vol. 264, Oct. 2021, Art. no. 112589.
- [6] Z. Liu, Y. Zhong, G. Ma, X. Wang, and L. Zhang, "A deep temporal-spectral-spatial anchor-free Siamese tracking network for hyperspectral video object tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5539216.



- [7] Z. Han, J. Yang, L. Gao, Z. Zeng, B. Zhang, and J. Chanussot, "Dual-branch subpixel-guided network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5521813.
- [8] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag. Replaces Newsletter*, vol. 5, no. 2, pp. 29–56, Jun. 2017.
- [9] G. Vivone, "Multispectral and hyperspectral image fusion in remote sensing: A survey," *Inf. Fusion*, vol. 89, pp. 405–417, Jan. 2023.
- [10] R. Dian, S. Li, B. Sun, and A. Guo, "Recent advances and new guidelines on hyperspectral and multispectral image fusion," *Inf. Fusion*, vol. 69, pp. 40–51, May 2021.
- [11] X. Zhao, K. Liu, K. Gao, and W. Li, "Hyperspectral time-series target detection based on spectral perception and spatial-temporal tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5520812.
- [12] H. Zhang, C. Zhang, F. Xie, and Z. Jiang, "A closed-loop network for single infrared remote sensing image super-resolution in real world," *Remote Sens.*, vol. 15, no. 4, p. 882, Feb. 2023.
- [13] H. Zhang, P. Wang, and Z. Jiang, "Nonpairwise-trained cycle convolutional neural network for single remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4250–4261, May 2021.
- [14] 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.
- [15] 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.
- [16] X. Wang, C. Cheng, S. Liu, R. Song, X. Wang, and L. Feng, "SS-INR: Spatial-spectral implicit neural representation network for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5525214.
- [17] S. Liu, S. Liu, S. Zhang, B. Li, W. Hu, and Y.-D. Zhang, "SSAU-Net: A spectral-spatial attention-based U-Net for hyperspectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5542116.
- [18] 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. 5531916.
- [19] T. Xu et al., "A coupled tensor double-factor method for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5515417.
- [20] T. Xu, T.-Z. Huang, L.-J. Deng, X.-L. Zhao, and J. Huang, "Hyperspectral image superresolution using unidirectional total variation with Tucker decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4381–4398, 2020.
- [21] 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.
- [22] Q. Ma, J. Jiang, X. Liu, and J. Ma, "Multi-task interaction learning for spatio-spectral image super-resolution," *IEEE Trans. Image Process.*, vol. 31, pp. 2950–2961, 2022.
- [23] 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. 5501713.
- [24] 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.
- [25] C. He, Y. Xu, Z. Wu, and Z. Wei, "Connecting low-level and high-level visions: A joint optimization for hyperspectral image super-resolution and target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5514116.
- [26] J. Wang, T. Lu, X. Huang, R. Zhang, and X. Feng, "Pan-sharpening via conditional invertible neural network," *Inf. Fusion*, vol. 101, Jan. 2024, Art. no. 101980.
- [27] 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.
- [28] X. Cao, Y. Lian, J. Li, K. Wang, and C. Ma, "Unsupervised multi-level spatio-spectral fusion transformer for hyperspectral image super-resolution," *Opt. Laser Technol.*, vol. 176, Sep. 2024, Art. no. 111032.
- [29] Q. Liu, X. Meng, S. Zhang, X. Li, and F. Shao, "A temporally insensitive spatio-temporal fusion method for remote sensing imagery via semantic prior regularization," *Inf. Fusion*, vol. 117, May 2025, Art. no. 102818.
- [30] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.
- [31] T. Uezato, D. Hong, N. Yokoya, and W. He, "Guided deep decoder: Unsupervised image pair fusion," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2020, pp. 87–102.
- [32] S. Liu, S. Miao, J. Su, B. Li, W. Hu, and Y. 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.
- [33] 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.
- [34] 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.
- [35] 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, Sep. 2024, Art. no. 102395.
- [36] Z. Wang, M. K. Ng, J. Michalski, and L. Zhuang, "A self-supervised deep denoiser for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5520414.
- [37] K. Zhang, M. Wang, and S. Yang, "Multispectral and hyperspectral image fusion based on group spectral embedding and low-rank factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1363–1371, Mar. 2017.
- [38] X. Han, B. Shi, and Y. Zheng, "Self-similarity constrained sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5625–5637, Nov. 2018.
- [39] X. Fu, H. Liang, and S. Jia, "Mixed noise-oriented hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5526916.
- [40] 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.
- [41] Y. Xu, Z. Wu, J. Chanussot, P. Comon, and Z. Wei, "Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 348–362, Mar. 2019.
- [42] K. Zhang, M. Wang, S. Yang, and L. Jiao, "Spatial-spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1030–1040, Apr. 2018.
- [43] R. Dian, S. Li, L. Fang, T. Lu, and J. M. Bioucas-Dias, "Non-local sparse tensor factorization for semiblind hyperspectral and multispectral image fusion," *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 4469–4480, Oct. 2020.
- [44] R. Dian, Y. Liu, and S. Li, "Hyperspectral image fusion via a novel generalized tensor nuclear norm regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 8, 2024, doi: 10.1109/TNNLS.2024.3385473.
- [45] H. Xu, C. Fang, Y. Ge, Y. Gu, and J. Zheng, "Cascade-transform-based tensor nuclear norm for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5540616.
- [46] Z.-C. Wu, Y.-J. Li, T.-Z. Huang, L.-J. Deng, and G. Vivone, "CroDoSR: Tensor cross-domain rank for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5530315.
- [47] S. Mei, R. Jiang, X. Li, and Q. Du, "Spatial and spectral joint super-resolution using convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4590–4603, Jul. 2020.
- [48] S. Mei et al., "Lightweight multiresolution feature fusion network for spectral super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5501414.
- [49] S. Mei, Y. Geng, J. Hou, and Q. Du, "Learning hyperspectral images from RGB images via a coarse-to-fine CNN," *Sci. China Inf. Sci.*, vol. 65, no. 5, pp. 1–14, May 2022.
- [50] N. Wang, S. Mei, Y. Zhang, M. Ma, and X. Zhang, "Hyperspectral image reconstruction from RGB input through highlighting intrinsic properties," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5525613.

- [51] J. Liu, S. Li, R. Dian, Z. Song, and L. Tan, "Asymptotic spectral mapping for hyperspectral image fusion," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Nov. 27, 2024, doi: [10.1109/TCSVT.2024.3507860](https://doi.org/10.1109/TCSVT.2024.3507860).
- [52] W. Sun et al., "Domain transform model driven by deep learning for anti-noise hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5500117.
- [53] 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.
- [54] 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. 5531115.
- [55] J.-F. Hu, T.-Z. Huang, L.-J. Deng, T.-X. Jiang, G. Vivone, and J. Chanussot, "Hyperspectral image super-resolution via deep spatiotemporal attention convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7251–7265, Dec. 2022.
- [56] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, Nov. 2018.
- [57] W. He, X. Fu, N. Li, Q. Ren, and S. Jia, "LGCT: Local-global collaborative transformer for fusion of hyperspectral and multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5537114.
- [58] 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. 5503715.
- [59] 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.
- [60] S. Chen, L. Zhang, and L. Zhang, "MSDformer: Multiscale deformable transformer for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5525614.
- [61] 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.
- [62] J. Qu, J. He, W. Dong, and J. Zhao, "S2CycleDiff: Spatial-spectral-bilateral cycle-diffusion framework for hyperspectral image super-resolution," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, 2024, pp. 4623–4631.
- [63] 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.
- [64] H. Zhu et al., "Adaptive dual-path collaborative learning for PAN and MS classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5413115.
- [65] Z. He, K. Xia, P. Ghamisi, Y. Hu, S. Fan, and B. Zu, "HyperViTGAN: Semisupervised generative adversarial network with transformer for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6053–6068, 2022.
- [66] Y. Zhang et al., "A cross-modal feature aggregation and enhancement network for hyperspectral and LiDAR joint classification," *Expert Syst. Appl.*, vol. 258, Dec. 2024, Art. no. 125145.
- [67] J. Yan et al., "Spatial-spectral unfolding network with mutual guidance for multispectral and hyperspectral image fusion," *Pattern Recognit.*, vol. 161, May 2025, Art. no. 111277.
- [68] J. Liu, Z. Wu, and L. Xiao, "A spectral diffusion prior for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5528613.
- [69] 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.
- [70] 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.
- [71] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse Dirichlet-Net for hyperspectral image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2511–2520.
- [72] 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.
- [73] 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.
- [74] 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. 5527016.
- [75] 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.*, vol. 35, no. 9, pp. 13017–13031, Sep. 2023.
- [76] H. Wang, Y. Xu, Z. Wu, and Z. Wei, "Unsupervised hyperspectral and multispectral image blind fusion based on deep Tucker decomposition network with spatial-spectral manifold learning," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 4, 2024, doi: [10.1109/TNNLS.2024.3457781](https://doi.org/10.1109/TNNLS.2024.3457781).
- [77] 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.
- [78] 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.
- [79] 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.
- [80] 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.
- [81] S. Shi, L. Zhang, Y. Altmann, and J. Chen, "Unsupervised hyperspectral and multispectral images fusion based on the cycle consistency," 2023, [arXiv:2307.03413](https://arxiv.org/abs/2307.03413).
- [82] B. Rasti, B. Koirala, P. Scheunders, and P. Ghamisi, "UnDIP: Hyperspectral unmixing using deep image prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5504615.
- [83] C. Zhou and M. R. D. Rodrigues, "Hyperspectral blind unmixing using a double deep image prior," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 11, pp. 16478–16492, Nov. 2024.
- [84] L. Zhuang, M. K. Ng, and X. Fu, "Hyperspectral image mixed noise removal using subspace representation and deep CNN image prior," *Remote Sens.*, vol. 13, no. 20, p. 4098, Oct. 2021.
- [85] Y. Zheng, J. Li, Y. Li, J. Guo, X. Wu, and J. Chanussot, "Hyperspectral pansharpening using deep prior and dual attention residual network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8059–8076, Nov. 2020.
- [86] W. G. C. Bandara, J. M. J. Valanarasu, and V. M. Patel, "Hyperspectral pansharpening based on improved deep image prior and residual reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5520816.
- [87] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3844–3851.
- [88] S. Liu, S. Zhang, S. Liu, B. Li, and Y.-D. Zhang, "UMTF-Net: An unsupervised multiscale transformer fusion network for hyperspectral and multispectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 17221–17238, 2024.
- [89] Y. Fang, Y. Liu, C.-Y. Chi, Z. Long, and C. Zhu, "CS2DIPs: Unsupervised HSI super-resolution using coupled spatial and spectral DIPs," *IEEE Trans. Image Process.*, vol. 33, pp. 3090–3101, 2024.
- [90] J. Li, K. Zheng, L. Gao, L. Ni, M. Huang, and J. Chanussot, "Model-informed multistage unsupervised network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5516117.
- [91] L. Zhang, J. Nie, W. Wei, Y. Li, and Y. Zhang, "Deep blind hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2388–2400, Jun. 2021.
- [92] 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.*, vol. 46, no. 7, pp. 5008–5025, Jul. 2024.
- [93] J. Nie, W. Wei, L. Zhang, C. Ding, and Y. Zhang, "Hybrid pixel-wise registration learning for robust fusion-based hyperspectral image super-resolution," *IEEE Trans. Comput. Imag.*, vol. 10, pp. 915–927, 2024.
- [94] Q. Cao, L.-J. Deng, W. Wang, J. Hou, and G. Vivone, "Zero-shot semi-supervised learning for pansharpening," *Inf. Fusion*, vol. 101, Jan. 2024, Art. no. 102001.

- [95] Y. Mansour and R. Heckel, "Zero-shot noise2noise: Efficient image denoising without any data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 14018–14027.
- [96] A. Shocher, N. Cohen, and M. Irani, "'Zero-shot' super-resolution using deep internal learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 3118–3126.
- [97] 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.
- [98] 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. 2022.
- [99] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
- [100] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2Net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2019.
- [101] 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.
- [102] 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.
- [103] 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.
- [104] 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.
- [105] 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.
- [106] C. Prévost, K. Usvich, 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.
- [107] 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. 5412613.
- [108] 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.
- [109] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Aug. 2002.
- [110] 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, Apr. 2004.
- [111] H. Yu, Z. Ling, K. Zheng, L. Gao, J. Li, and J. Chanussot, "Unsupervised hyperspectral and multispectral image fusion with deep spectral-spatial collaborative constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5534114.



**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 system with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

He has authored over 20 peer-reviewed articles with more than 800 citations in Google Scholar, including five ESI highly cited articles. 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, the President Scholarship for Postgraduate Students in 2024, and the Excellent PAPERS Award at the 2023 International Conference on Remote Sensing, Mapping and Geographic Information Systems. He served as a reviewer for over ten journals.



**Ke Zheng** received the B.S. degree in geographic information system 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 with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, 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, 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 system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, in 2007.

He was a Visiting Scholar with the University of Extremadura, Cáceres, Spain, in 2014, and with 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 by the National Key Research and Development Program of China (2021–2025). He has authored more than 240 peer-reviewed articles, and there are more than 150 journal articles included in Science Citation Index (SCI). He has co-authored three academic books including *Hyperspectral Image Information Extraction*. He holds 30 national invention patents in China. His research focuses on hyperspectral image processing and information extraction.

Dr. Gao was supported by the National Science Foundation for Distinguished Young Scholars of China in 2023, and was awarded the Outstanding Science and Technology Achievement Prize of the CAS in 2016 and the Second Prize of The State Scientific and Technological Progress Award in 2018. He received the 2021 Outstanding Paper Award at the IEEE Workshop on Hyperspectral Image Processing: Evolution in Remote Sensing (WHISPERS). He is an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) and *IET Image Processing*. He is a fellow of the Institution of Engineering and Technology.



**Zhu Han** (Student Member, IEEE) received the B.S. degree in electronic engineering from North China University of Technology, Beijing, China, in 2019. She is currently pursuing the Ph.D. degree in cartography and geographic information system with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing.

From 2022 to 2023, she was a Visiting Ph.D. Student with the Lancaster Environment Centre, Faculty of Science and Technology, Lancaster University, Lancaster, U.K. Her research interests include hyperspectral image processing, deep learning, and artificial intelligence.





**Zhi Li** received the B.Eng. degree in remote sensing science and technology from the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China, in 2021. He is currently pursuing the Ph.D. degree in cartography and geography information system with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.

His research interests include hyperspectral image processing, remote sensing classification, computer vision, and artificial intelligence.



**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 of remote sensing and is 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). He was the Co-Chair of the GRS Data Fusion Technical Committee, from 2005 to 2008, 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.