Hyperspectral Image Super-Resolution Meets Deep Learning: A Survey and Perspective

Xinya Wang ⓘD, Qian Hu ⓘD, Yingsong Cheng ⓘD, and Jiayi Ma ⓘD, *Member, IEEE*

Abstract—Hyperspectral image super-resolution, which refers to reconstructing the high-resolution hyperspectral image from the input low-resolution observation, aims to improve the spatial resolution of the hyperspectral image, which is beneficial for subsequent applications. The development of deep learning has promoted significant progress in hyperspectral image super-resolution, and the powerful expression capabilities of deep neural networks make the predicted results more reliable. Recently, several latest deep learning technologies have made the hyperspectral image super-resolution method explode. However, a comprehensive review and analysis of the latest deep learning methods from the hyperspectral image super-resolution perspective is absent. To this end, in this survey, we first introduce the concept of hyperspectral image super-resolution and classify the methods from the perspectives with or without auxiliary information. Then, we review the learning-based methods in three categories, including single hyperspectral image super-resolution, panchromatic-based hyperspectral image super-resolution, and multispectral-based hyperspectral image super-resolution. Subsequently, we summarize the commonly used hyperspectral dataset, and the evaluations for some representative methods in three categories are performed qualitatively and quantitatively. Moreover, we briefly introduce several typical applications of hyperspectral image super-resolution, including ground object classification, urban change detection, and ecosystem monitoring. Finally, we provide the conclusion and challenges in existing learning-based methods, looking forward to potential future research directions.

Index Terms—Deep learning, hyperspectral image, image fusion, image super-resolution, survey.

I. INTRODUCTION

HYPERSPECTRAL (HS) sensors can record rich reflection information covering a continuous and wide spectrum range. Thus, HS data is usually equipped with a higher spectral resolution, having dozens or hundreds of spectral bands, which reflects the subtle spectral properties of different objects [1]. Because of the superior capacity of discriminating key features of the captured scene [2], [3], HS images have been widely applied in many fields, such as mineral exploration [4], vegetation classification [5] and change detection [6], to name a few. Constrained by imaging devices, there

is always a trade-off between spatial and spectral resolution of HS data. Compared with the natural image, the HS image has a low spatial resolution to keep an acceptable signal-to-noise ratio. As a result, the development of some applications is hampered, limiting the further exploration of HS data. Due to the theoretical and technical limitations of hardware equipment, it is vital to investigate algorithm techniques to overcome the spatial resolution restrictions.

Super-resolution (SR) task aims to improve the spatial resolution of the input data, which has received intensive research in the HS field. Before the prevalence of deep learning, the early methods leveraged the traditional approaches including Bayesian model [7]–[9], tensor representation [10]–[14] and matrix decomposition [15]–[19], etc., to solve the ill-posed problem. Typically, these methods highly depend on the prior that is sophisticatedly designed, such as self-similarity, sparsity, or low rank of the HS image, to regularize the SR reconstruction process. However, the drawbacks of these methods have been increasingly apparent. On the one hand, the traditional heuristic models have limited expression ability. Thus, they usually fail to restore the high-frequency details. On the other hand, the handcraft priors are simple and rough so they do not satisfy the complex real scenes, resulting in limited reconstruction performance. The motivation for introducing deep learning into HS image SR is to overcome the limitations of traditional methods. First, driven by data, deep learning models have powerful expressive capabilities with the help of external datasets. Second, mature network structures in deep models can better explore the spatial-spectral correlation of HS images, so as to obtain more valuable features. Benefiting from these advantages, deep learning promotes tremendous progress in HS image SR, achieving outstanding performance far exceeding traditional methods.

From the perspective of source data, existing learning-based methods for the HS image SR task can be divided into two categories: single HS image SR and fusion-based HS image SR. The former only utilizes a single low-resolution (LR) input to recover the high-resolution (HR) result. In recent years, single HS image SR has received a lot of attention due to extensive deep learning investigations in the natural image SR task. In the single HS image SR task, researchers are devoted to exploiting rich spectral information and processing the high-dimension features, generating the HR HS image with spatial-spectral consistency. With additional auxiliary information, the latter methods fuse the HS image with auxiliary images, including panchromatic (PAN), RGB, and multi-
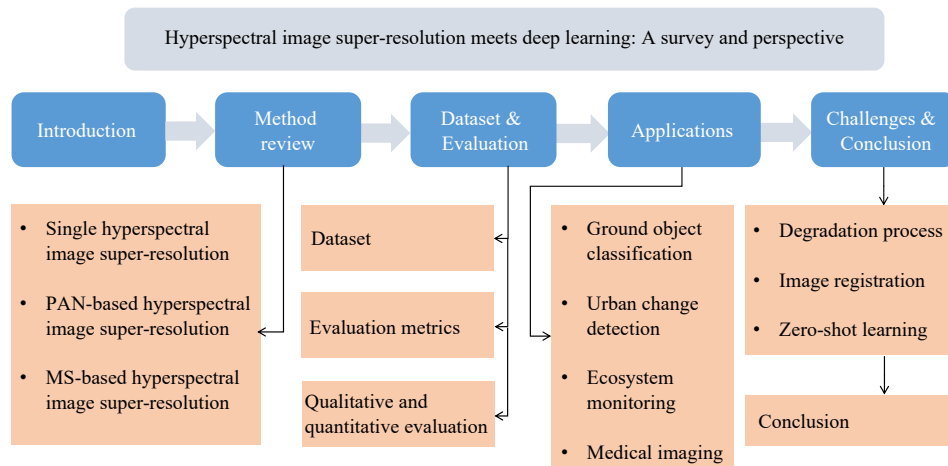
Fig. 1.   Overall framework of the survey.

spectral (MS) images, to improve the spatial resolution of the observed HS image. The goal of this fusion-based HS image SR is to produce the HS data with finer spatial resolution than HS sensors, instead of showing a higher spectral resolution. Generally, these auxiliary images captured at the same scene as the LR HS image have a higher spatial resolution. On the one hand, the simultaneous acquisition of HS data with either MS or PAN images in some existing satellites such as Prisma mission2 [20] and Gaofen 5-024 [21] activates the potentiality in the design and development of these fusion-based methods. On the other hand, due to the excellent characteristics of fused images, fusion-based HS image SR has also been flourishing with deep learning. These fusion methods are dedicated to establishing the relationship between two inputs, serving for feature extraction and fusion. Because of the different emphases in various HS image SR tasks, it is appropriate to review the deep learning-based HS image SR methods according to the above categories.

The motivation for this work is as follows. On the one hand, in the past decades, HS image SR has prospered by virtue of deep learning, and the research on deep SR models has entered a new stage. Whereas, the early surveys [1], [22] only present the traditional fusion methods with the absence of recent findings. On the other hand, existing surveys [23]–[26] mainly focus on the specific HS image fusion task, For example, these works [24]–[26] provide the review and guideline of the HS and MS image fusion. [23] contains some state-of-the-art HS-MS and HS-PAN approaches. Few works comprehensively review the latest technologies from the HS image SR perspective. To this end, we propose a comprehensive survey for the state-of-the-art on the use of deep learning in the HS image SR task, which could help relevant researchers better understand the current development status of the HS image SR field.

The overall framework of this survey is shown in Fig. 1. In particular, we concentrate on the deep learning works for improving the spatial resolution of HS data, rather than spatio-spectral SR [27], [28] and spectral SR [29]–[31]. We first review the single HS image SR. Then, we give a review of HS image SR methods with the assistance of PAN images. Since the well-known pansharpening problem [22] is to fuse the MS

image with the PAN image, some works confuse them in naming. Although HS and PAN fusion task is denoted as HS sharpening in [32], to avoid confusion, we regard these methods as "PAN-based hyperspectral image super-resolution". Finally, we comprehensively review the MS-based hyperspectral image super-resolution with the MS or RGB image as auxiliary information as most methods do not distinguish between MS and RGB images on purpose. In natural HS datasets, the MS image is equivalent to the RGB image or contains RGB and near-infrared bands, while in remote sensing HS datasets, the MS image contains multiple bands, such as six bands from Blue to SWIR2 in Indian Pines and Washington DC datasets. From a technical level, there is not much difference between deep learning HS image SR methods based on MS and RGB images. It is worth noting that a few methods [33]–[36] fuse multiple images to super-resolve HS images, which will not be introduced in detail. Moreover, we sum up the commonly used dataset and conduct a brief evaluation of representative deep learning-based methods in each category. Subsequently, some typical applications of HS image SR are introduced, i.e., ground object classification, urban change detection, ecosystem monitoring and medical imaging. In the end, we put forward some prospects for the future study with the challenges that exist in HS image SR and summarize this survey.

The rest of this paper is organized as follows. Section II introduces the notations. Section III is devoted to the introduction of methodologies for single HS image SR, which is further divided into three categories: two-step methods, group-based methods, and others. In Section IV, PAN-based HS image SR approaches are considered from four aspects: model-guided methods, straight fusion methods, multi-stage methods, and other end-to-end deep learning methods. Section V is about MS-based HS image SR algorithms. The classification of techniques in this category is performed by describing the mathematical fundamentals and reviewing the related literature. Section VI supplies a conclusion about the commonly used datasets in the HS image SR task. Section VII introduces the commonly used evaluation metrics in the HS image SR field and three experiments have been conducted to compare the state-of-the-art methods in three categories for

qualitative and quantitative evaluation in Section VIII. Section IX displays some studies of HS image SR applications. Finally, challenges and new guidelines are proposed in Section X and we draw a conclusion in Section XI.

II. NOTATIONS

For a convenient description, we first introduce some notations used in the paper before introducing the various methods. Let the desired HR HS image be denoted as $Y \in \mathbb{R}^{H \times W \times C}$, where C is the number of spectral bands, H and W are the height and width respectively of the HR HS image. The LR HS image is supposed as $X \in \mathbb{R}^{h \times w \times C}$, where $h = H/s$, $w = W/s$ are the height and width respectively of the LR HS image with the scale factor s. We denote the PAN image as $P \in \mathbb{R}^{H \times W \times 1}$ and the MS image is referred to as $Z \in \mathbb{R}^{H \times W \times c}$ with lower spectral resolution ($c < C$).

III. SINGLE HYPERSPECTRAL IMAGE SUPER-RESOLUTION

Single HS image SR does not utilize the auxiliary image, mining only information from the LR image itself to reconstruct the HR image, which has a wider range of application scenarios. According to the development of techniques, learning-based methods can be roughly divided into three categories.

A. Two-Step Methods

Traditional methods for single HS image tasks are based on handcraft features, which only reflect the characteristics of a certain aspect of HS images. And the optimization in the test stage is complicated and time-consuming, resulting in limited performance. With the rise of deep learning in the vision field, the neural network has been introduced into the single HS image SR task. Early, some methods attempt to use the convolutional neural network (CNN) as a part of the algorithm and combine the post-process for further optimization. Li *et al.* [37] and Hu *et al.* [38] made use of a spectral difference convolutional neural network followed by the spatial correction strategy for post-processing. Resorting to dictionary learning, He and Liu [39] applied a Laplacian pyramid network to learn the coarse HR HS image and reconstructed the spectral information by non-negative dictionary learning. Considering that the wavelet could capture image structures in different orientations, in [40], the 3D CNN is applied to predict the wavelet coefficients of HR HS image that can be obtained by inverse wavelet transform. The nonnegative matrix factorization (NMF) can decompose a nonnegative matrix into a product of nonnegative matrices, which is proven to be helpful in the SR task. Thereby, the NMF strategy has been popular as the post-process to enforce the intermediate HR HS image that is obtained by a tailored CNN correlated to the LR one [41]–[44]. With the relatively low spatial resolution, pixels in the HS image are usually mixed with different materials. The spectral curve of each pixel in the HS image is the mixture of different pure material reflectances, and these pure materials are called endmembers. Considering the mathematics simplicity and physical effectiveness, each spectral curve is assumed to be modeled by a linear spectral mixture model. Thus, based on the hypothesis of the linear model, the HS image could be decomposed into the product of the endmember and abundance matrix. To fully exploit the intrinsic properties of HS images, Lu *et al.* [45] designed a dual-branch network with spatial spread transform function for inferring abundance maps, then reconstructing the LR and HR images to characterize the spatial collaborations. As illustrated in Fig. 2(a), this operation that first super-resolves the bands just by their spatial information and then incorporates a manual way to make a further utilization of the spectral information is inflexible and unstable. Meanwhile, the performance of the NMF highly depends on the iteration times and the number of the endmembers. For this type of method, the noise generated by the unmixing operation is inevitable during the mapping operation, which makes a negative influence on the SR process.
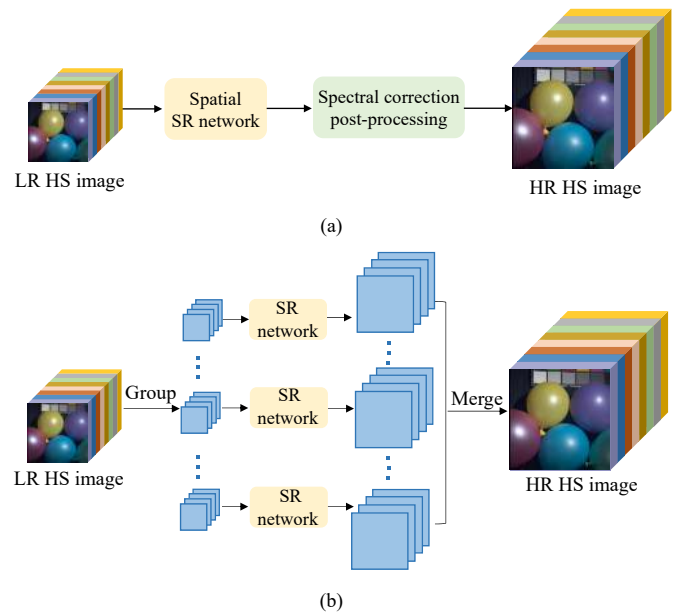

Fig. 2.   Diagrams of single HS image SR: (a) Two-step methods; (b) Group-based methods.

B. Group-Based Methods

Since HS images have dozens or hundreds of spectral channels, rebuilding all spectral bands simultaneously requires a large-scale model. Thereby, some methods concentrate on grouping the whole HS image along the spectral dimension into spectral subgroups for SR, which is displayed in Fig. 2(b). Since the neighboring bands of the HS image are highly correlated in the spectral domain [4], according to the correlation coefficient curve, the key bands are firstly selected and super-resolved by a CNN while the whole HR HS image is reconstructed by further optimization [41]–[43]. To learn an end-to-end map, Li *et al.* [46] developed a group deep recursive residual network (GDRRN) using the group convolution to reduce the number of parameters. Sharing the same insight, SSPSR [47] proposes the group reconstruction strategy. Specifically, SSPSR divides the bands of HS images into groups that are reconstructed by the branch network with shared parameters and finally merges the subgroups for processing to learn spectral correlation, which greatly reduces the number of parameters and achieves excellent performance.

Related to this, RFSR [48] exploits feedback mechanisms to construct interactions between spectral groups and constrain spectral continuity in the HR space by separable 3D convolution. Moreover, Li *et al.* [49] applied a progressive split-merge SR framework with gradient-guided group attention to overcome the inherent resolution limitations. Besides, Zhang *et al.* [50] presented a novel difference curvature multidimensional network in which the input image is channel-wisely split, thus lowering the burden on the device. Instead of explicit grouping as input, in [51], a novel method with a spectral loss function is developed for group multiscale feature fusion. Considering the block characteristic of the HS image, Liu *et al.* [52] adapted the spectral attention for group convolutions to rescale grouping features with holistic spectral information. Specifically, the spectral attention mechanism is constructed by covariance statistics of features. Since the current band is highly correlated with adjacent bands, Hu *et al.* [53] took the current band and its neighboring LR band as the input to learn feature maps of the current band, its neighboring band, and their spectral difference in a parallel way. Similarly, Wang *et al.* [54] utilized the current band and its two adjacent bands to perform single-band SR by a dual-channel network, in which features that have been extracted from the previous band SR task are fed into the network of the current band. With the feedback mechanism, a novel refined local-global network is adopted in [55] to correct the low-level representation by feedback high-level semantic information, effectively exploring spatial-spectral priors between spectral bands. To alleviate the difficulty of feature extraction and reconstruction, a group-based embedding learning and integration network [56] reconstructs the HR images in a group-by-group manner. Recently, by separating adjacent and extremely similar bands in each group, Wang *et al.* [57] pioneered the exploration of shuffling spectral band sequences to improve the reconstruction effect.

Dividing the high-dimension HS data into group bands, the key of this group-based method is to keep spectral consistency between the super-resolved HS data and the HR one. Although many efforts have been made in existing methods, there is still room for improvement in removing spectral distortion.

C. Other End-to-End Methods

With the prosperity of deep CNN in the RGB image SR field, many end-to-end deep learning models have been proposed to solve the HS image SR problem as

$$Y = \mathcal{F}(X, \theta) \tag{1}$$

where $\mathcal{F}$ represents deep learning networks with the parameter θ.

1) 3D Structure: Due to the three-dimensional structure of the HS image, 3D convolution and its variations became popular in this task. Mei *et al.* [58], [59] used 3D full convolution to construct the network with a sensor-specific mode. As the computational complexity of 3D convolution is tremendous, Li *et al.* [60], [61] proposed a dual-flow 1D-2D spectral-spatial CNN, which employs 1D and 2D convolutions to extract spectral and spatial features in a separated way and then fuse them gradually. Similarly, a separable 3D convolution is fur-

ther introduced in [62] to analyze spatial and spectral information, reducing the model size with the maintenance of the capability of 3D convolution in the spectral domain. To make up for the deficiency of spatial information extraction by separable 3D convolution, MCNet [63], ERCSR [64], SFCSR [54] and method in [65] deploy mix 2D/3D convolution networks and share spatial information to reconstruct spatial details better. Moreover, Fu *et al.* [66] designed a bidirectional 3D quasi-recurrent neural network for the HS image SR with an arbitrary number of bands. By jointly embedding the local and nonlocal attention in a residual 3D CNN, a hybrid local and nonlocal 3D attentive CNN [67] is built for the HS image SR.

2) GAN Framework: Inspired by the generative adversarial network (GAN), a GAN-based framework [68] is proposed for HS image SR. In [69], SRGAN [70] has been extended with 3D convolution and attention mechanism to not only exploit the spatial features but also preserve spectral properties. Related to this, Li *et al.* [71] presented a novel band attention embedded in the adversarial learning-based SR method, which is beneficial to alleviate the spectral distortion problem. Similarly, Jiang *et al.* [72] employed the 2D-1D GAN framework, which includes two subnetworks, i.e., spatial and spectral subnetworks. This decomposition effectively reduces computational complexity, but there is no joint processing of spatial and spectral information. Moreover, a novel GAN-based method is adopted in [73] with a spatial feature-enhanced network and a spectral refined network, which can effectively enhance coarse high-frequency features of HS images. To alleviate the problem of mode collapse that GAN-based models frequently suffer from, [74] coupled with a latent encoder converts the optimization of GAN models to the problem of learning the distributions of HR HS samples in the latent space.

Since 3D convolution and its variations have become a fundamental part of the network, researchers have turned to other advanced techniques in computer vision. Based on the deep image prior (DIP) algorithm [75], Sidorov and Yngve Hardeberg [76] extended a 3D-convolutional version using intrinsic properties of a CNN without any training for HS image denoising, inpainting, and SR. Absorbing the advantages of DIP, DRPSR algorithm [77] appends a special input processing module onto the HS image SR network to automatically adjust the structure of the input without requiring the training samples so that the choice of network structure is no longer limited. Combining the sophisticated residual structures [70], [78] in natural image SR task, Zhao *et al.* [79] developed a novel gradient-guided residual dense network into the single HS image SR task. In [80], a self-calibrated attention residual block is elaborately designed to fully exploit the spatial information and the correlation between the spectra of the HS data while retaining spectral consistency. Besides, redefining the SR problem in the abundance domain, Wang *et al.* [81] unified the unmixing and SR into an end-to-end method based on the assumption that the LR HS image and HR HS image share the same spectrum endmembers. To learn spatial-spectral characteristics among different spectral segments, a multiscale feature generation and fusion network based on wavelet transform is presented in [82]. By casting the embedding of

Zhang *et al.* [82]
Wang *et al.* [80]
Zhao *et al.*[79]
Wang *et al.* [73]
Yang *et al.* [67]
Fu *et al.* [66]
ERCSR (Li *et al.* [64])
SFCSR (Wang *et al.* [54])
Liu *et al.* [52]
Zhang *et al.* [50]
PSMSR (Li *et al.* [49])
Lu *et al.* [45]

Sidorov and
Hardeberg [76]
Jiang *et al.* [72]
Huang *et al.* [68]
Li *et al.* [60], [61]
Hu *et al.* [43]
DFMF (Xie *et al.* [42])
Yang *et al.* [40]

3DFCNN (Mei *et al.* [59])
Mei *et al.* [58]
Yuan *et al.* [41]
Hu *et al.* [38]
Li *et al.* [37]

Liu *et al.* [84]

Single HS
Image SR

2017   2019   2021   2023

He and Liu [39]
GDDRN (Li *et al.* [46])

Zou and Huang [44]
SSPSR (Jiang *et al.* [47])
Hu *et al.* [51]
IFN (Hu *et al.* [53])
Wang *et al.* [62]
MCNet (Li *et al.* [63])
Dou *et al.* [69]
Li *et al.* [71]

RFSR (Wang *et al.* [48])
Tang *et al.* [55]
GELIN (Wang *et al.* [56])
GSSR (Wang *et al.* [57])
Li *et al.* [65]
LEGAN (Shi *et al.* [74])
DRPSR (Gong *et al.* [77])
aeDPCN (Wang *et al.* [81])
Hou *et al.* [83]
Interactformer (Liu *et al.* [85])
Deng *et al.* [86]

- Two-step methods
- Group-based methods
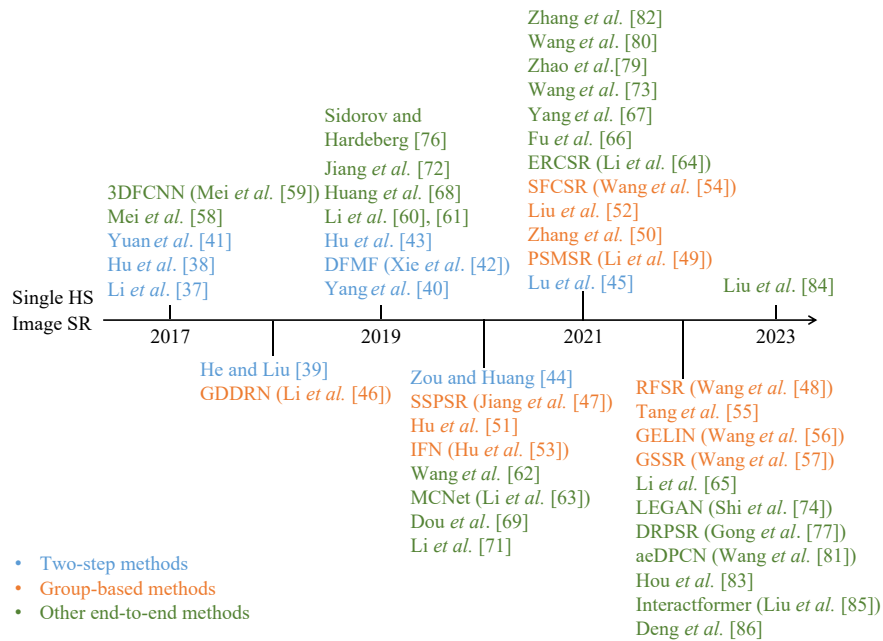- Other end-to-end methods

Fig. 3.   The development of the single HS image SR methods.

the high-dimensional spatial-spectral information of HS images as an approximation to the posterior distribution, Hou *et al.* [83] incorporated the proposed feature embedding scheme into a physically-interpretable deep framework to construct an end-to-end HS image SR method. Similarly, a maximum a posterior (MAP) framework is also unfolded into an interpretable multi-stage network [84] to explicitly impose the degradation model constraint. Recently, introducing the Transformer structure into this task, Interactformer [85] interacts with global and local features extracted by the Transformer and 3D CNN branches for SR, achieving state-of-the-art performance. To fit the real degradation circumstance, a novel framework [86] of multiple frame splicing strategy is put forward to improve the HS SR quality with multiple HS degradation models.

All in all, the two-step methods combining the traditional algorithm as a post-process could inevitably bring about error accumulation, failing to produce satisfactory results. Although the group-based methods alleviate the dilemma caused by the hyperspectral dimension of HS data to some extent, they still have shortcomings in maintaining spectral consistency. With the sophisticated structure and well-designed training strategy, the end-to-end deep learning methods have obtained an increasing performance in simulated experiment data. Until now, few single HS image SR methods could cope with the real degradation scene, and in the future, the blind SR for the HS image task may receive more attention. Besides, as shown in Fig. 3, we provide a chart about the timeline to represent more clearly the developments in this field. It can be seen that the two-step methods appear in the early stage and the full end-to-end methods play a dominant role in recent years.

IV.   PAN-BASED HYPERSPECTRAL IMAGE SUPER-RESOLUTION

In order to achieve superior SR performance, investigations have explored relatively HR data acquired by panchromatic sensors to assist HS image SR. PAN-based HS image SR is to produce HR HS image by fusing the LR HS image with the PAN image. Due to the similarity of the problems, some pan-sharpening methods originally designed for fusing MS images and PAN images are generic for both tasks. Owing to the end-to-end nature of deep learning methods, pansharpening methods can be directly applied to PAN-based HS image SR by changing the number of the input channel. Although these tasks all fuse images to generate HR images, the different source images carry different information, making it inappropriate to directly apply pansharpening methods for PAN-based HS image SR.

Since PAN images have a higher spatial resolution while HS images have a higher spectral resolution, the key to this task lies in how to fuse the information from both to generate superior results in terms of spatial structure and spectral correlation. According to the different ways of fusion, the existing deep learning methods can be broadly classified into four categories, say model-guided methods, straight fusion methods, multi-stage methods, and others.

A. Model-Guided Methods

Various traditional techniques have developed in the last decades to enhance the spatial resolution of HS imagery [1], such as multiresolution analysis (MRA), component substitution (CS), matrix factorization, etc. However, the effectiveness of these traditional methods is limited by the weak learning ability of models. In addition, these methods often require sophisticated manual designs.

MRA and CS-based methods for PAN-based HS image SR can be generalized as Fig. 4(a). The detailed image is first obtained by subtracting the low-pass blurred image from the HR PAN image, and the resulting HR HS image is achieved by gain matrix injecting the detailed image into the upsampled LR HS image, which can be summarized as
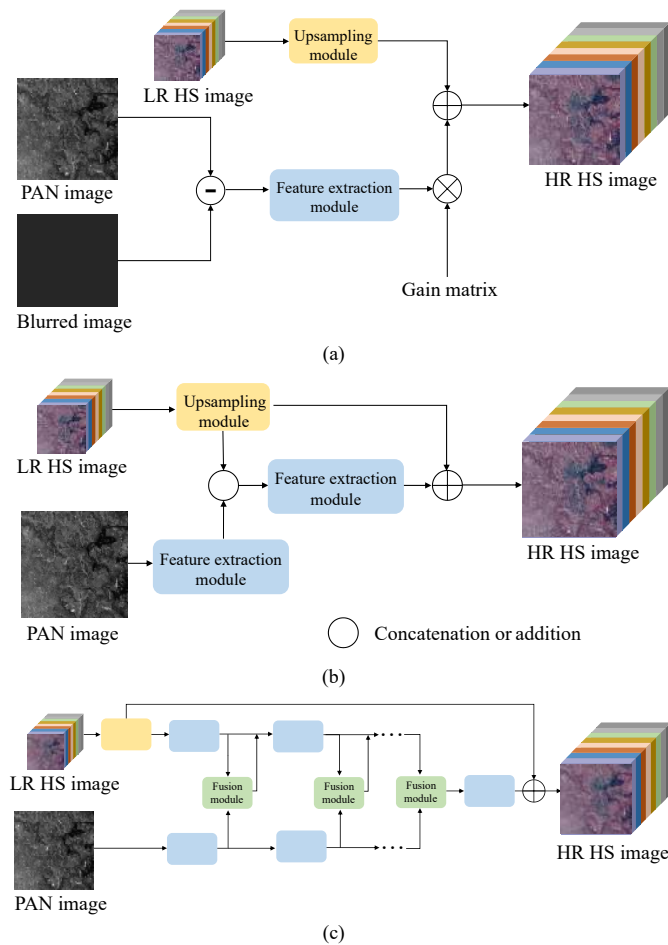
$$Y = X\uparrow + G \otimes (P - O_L) \qquad (2)$$

Fig. 4.   Diagrams of PAN-based HS image SR: (a) Model-guided methods; (b) Straight fusion methods; (c) Multistage methods.

where Y, X and P represent HR HS image, LR HS image, and PAN image, respectively, $\uparrow$ is the upsampling operation, G is the gain matrix, and O_L refers to the low-pass blurred image, the generation of which is different between MRA and CS algorithms. MRA methods obtain blurred images by multi-scale decomposition of PAN images, which could well preserve the spectral information by extracting details from the PAN image, but suffer from spatial distortions [22]. On the other hand, CS methods generate blurred images from LR HS images by substituting the component that contains the spatial information with the PAN image. Thus, the CS-based approaches are advocated for their high fidelity in rendering spatial details [87] but produce significant spectral distortion due to the spectral mismatch between the PAN and HS images [88]. In addition, according to (2), the gain matrix highly depends on handcraft design, making these methods not only time-consuming but also not generalizable.

With the popularity of deep neural networks, researchers have introduced deep networks to substitute modules in the aforementioned framework, achieving excellent success. Early, learning from the sophisticated networks in natural image SR, DDLPS method [89] combines progressive upsample strategy in LapSRN [90] to build a learnable upsampling module. Dong *et al.* [91] further replaced the upsampling and feature extraction modules with deep neural networks while

using the Laplacian Pyramid structure to obtain multi-resolution PAN images. In [92], [93], to alleviate the spectral distortion caused by spectral mismatch, networks are designed to process LR HS images to preserve the spectral correlation. Moreover, GASN algorithm [94] tailors a component network that can be plugged into MRA and CS-based methods.

Other model-guided methods have also achieved promising results. Zheng *et al.* [95] introduced a guided filter to inject spatial information from the PAN image into the LR HS image. Mog-DCN method [96] adopts convolutional networks to simulate spatial degradation and spectral degradation as well as their inverse transformations. Especially, Lu *et al.* [97] decomposed the SR problem into an MS and PAN image fusion task and an MS and HS image fusion task by generating simulated MS images from LR HS images. In this way, well-designed methods for these two fusion tasks can be introduced to facilitate the development of PAN-based HS image SR.

B. Straight Fusion Methods

With deep learning demonstrating superior learning capabilities in the field of computer vision, it is natural to directly use deep learning to learn the HR HS image from LR HS and PAN sources. As shown in Fig. 4(b), the basic concept of these methods is to derive spatial information from PAN images and spectral information from LR HS images. Subsequently, the convolutional network is designed to fuse these two types of information to obtain desired images. Some methods employ an element-wise addition combining spatial and spectral features [98], [99], while others use concatenation, which is further divided into row-level concatenation [100]–[102] and feature-level concatenation [103], [104]. In addition to direct fusion, some new techniques are developed in this field. In [100], [101], deep image prior network [75] is selected as the upsampling module. With the popularity of the attention mechanism, the spatial-spectral consistent features are generated by attention in [100], [102]. Meanwhile, HPGAN method [104] uses GAN to generate more realistic results for HS images.

C. Multistage Methods

To reduce information loss in deep neural networks as much as possible, researchers have favored multistage feature fusion, which fuses features in multiple stages to increase the fusion capability of models.

It has been demonstrated that shallow and deep network layers extract different levels of features. Specifically, the shallow features focus on low-level information such as edges, while the deep features mainly reflect the high-level semantic information of the image. Therefore, fusing features extracted from PAN and LR HS images at multiple network layers allows the model to be more flexible in combining spatial and spectral information from different layers to generate superior results.

As displayed in Fig. 4(c), multi-stage fusion methods are designed with ingenious feature fusion modules to fully merge different levels of spectral and spatial features. Specifically, various kinds of attention including local and global attention

Li *et al.* [61]   Zheng *et al.* [109]
HyperPNN (He *et al.* [103])   HPGAN (Xie *et al.* [104])
He *et al.* [93]   CCNNF (Lu *et al.* [97])
DDLPS (Li *et al.* [89])   Mog-DCN (Dong *et al.* [96])

PAN-based
HS image SR   2019   2021   2023

Xie *et al.* [92]
Zheng *et al.* [95]
Zheng *et al.* [98]   Dong *et al.* [91]   Shang *et al.* [111]
DHP-DARN (Zheng *et al.* [100])   GASN (Dong *et al.* [94])   Hypertransformer (Bandara and Patel [112])
He *et al.* [123]   Wu *et al.* [99]   MSSL (Qu *et al.* [114])
                              Bandara *et al.* [101]   Zhao and Du, *et al.* [115], [116]
                              Zhuo *et al.* [102]   DBDENet (Qu *et al.* [117])
• Model-guided methods   Hou *et al.* [105]   Dong *et al.* [118]
• Straight fusion methods   Guan and Lam [106]   He *et al.* [119], [120]
• Multistage methods   RHDN (Qu *et al.* [107])   Dong *et al.* [121]
• Other methods   Yang *et al.* [108]   RE-RANet (Nie *et al.* [122])
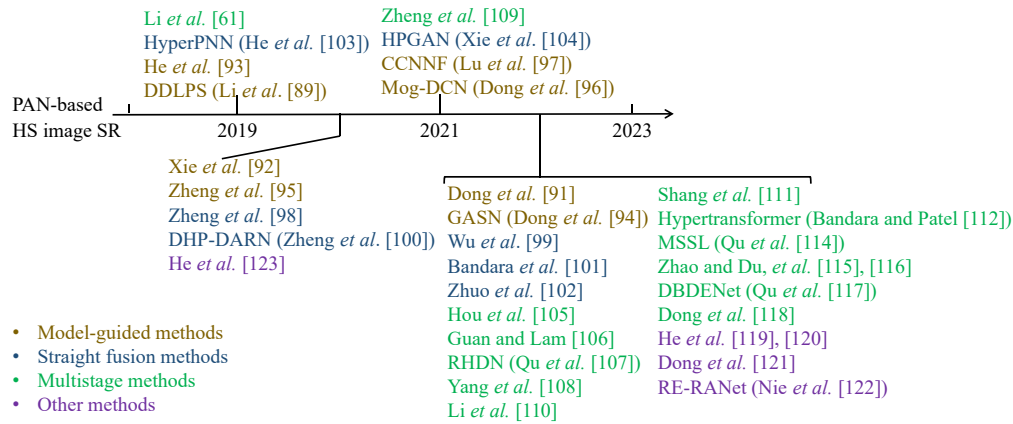   Li *et al.* [110]

Fig. 5.   The development of the PAN-based HS image SR methods.

[105], dual-attention [106], spatial and spectral attention [107], cross attention [108] and 1D-2D attention [61] are proposed to construct formidable fusion modules.

Based on the idea of multi-layer fusion, some works employ feature modulation to adaptively inject PAN features [109], [110]. Transformer has also been introduced to the PAN-based HS image SR task because of its success in the field of computer vision [111]–[113].

It can be noticed that features obtained from images of different resolutions also contain diverse information. Some multi-stage methods fuse multi-resolution information to generate richer features [114]–[116]. Differently from the previous methods, Qu *et al.* [117] customized two bi-directional branches and cross attention, which can efficiently utilize the hierarchical features from PAN and LR HS images. Besides, a two-branch cross-feedback dense network with context-aware guided attention is proposed in [118] to yield more effective spatial-spectral transfer.

D. Other Methods

While standard PAN-based HS image SR seeks to obtain the best performance on experimental data, some other works are devoted to making the field more practical. In applications, HS SR is usually involved as a pre-processing phase to assist some practical remote sensing tasks like object detection. However, the required spatial resolutions in some real-world applications may not be the same as that of the PAN image, which means standard PAN-based HS image SR is obviously incapable of dealing with such a situation. Therefore, arbitrary resolution HS image SR based on the PAN image, which is expected to sharpen LR HS images to any user-customized resolutions, comes into being. A standard arbitrary resolution HS SR method can be divided into two main phases, which are fusion and arbitrary scale SR [119], [120].

Since HR HS images are not available, the supervised methods are generally trained on the simulated image pairs, which consist of the degraded images and the original HS images. Specifically, the degraded image is obtained by implementing spectral degradation and spatial degradation on the original HS image. However, such a model trained in simulation may not be suitable for real scenarios, which has led scholars to focus on unsupervised PAN-based HS image SR. Unsupervised training can be performed by constraining the spatial and spectral coherence of the HR HS image and LR HS image [121], [122]. Another solution is provided in [123], which is downsampling the image into LR space to train the upsampling model and then applying it directly to the original image to obtain the HR one.

The model-guided method generally introduces deep learning into the traditional framework, thereby improving performance while retaining the interpretability of the method. However, they tend to contain some numerical steps as pre-processing or post-processing, resulting in uncompetitive performance. Although straight fusion methods achieve end-to-end training and retain spatial and spectral information efficiently through neural networks, direct addition or concatenation does not fuse spatial and spectral information well, thus affecting the quality of final images. By contrast, fusing the HS and PAN images in a multi-stage way exhibits excellent performance, which is much favored. Besides, some other methods are devoted to dealing with real scenarios. Even though these methods have not worked well for the time being, they are more promising. In the meanwhile, the developments of these methods are demonstrated in Fig. 5. In the early days, there were fewer PAN-based SR methods, and this type of method has not received extensive attention until 2022.

V. Ms-Based Hyperspectral Image Super-Resolution

The multispectral image is another commonly used auxiliary image for fusion-based MS image SR. Compared with PAN images, the MS image has a relatively high spectral resolution, and therefore, the spectral consistency of the reconstructed result could be better preserved. According to the unique technical route, generally, the MS-based HS image SR methods can be divided into three categories: model-based methods, decomposition-based methods, and other deep learning methods.

A. Model-Based Methods

A traditional approach to fuse auxiliary MS image and LR HS image is the model-based method, which aims at increasing the spatial resolution of LR HS image by designing hand-crafted prior and solving objective function. Generally, as

shown in Fig. 6(a), with the LR HS input $X \in \mathbb{R}^{h \times w \times C}$ and the auxiliary MS image $Z \in \mathbb{R}^{H \times W \times c}$, the model-based problem can be formulated as the minimization of the following objective function:

$$\mathcal{L}_Y = \|X - YBS\|_F^2 + \|Z - RY\|_F^2 + \varphi(Y) \tag{3}$$

where the $\|\cdot\|_F^2$ refers to the Frobenius norm, $B \in \mathbb{R}^{N \times N}$ is the spatial blurring matrix with $N = H \times W$, $S \in \mathbb{R}^{N \times n}$ denotes the point spread function (PSF) with $n = h \times w$, $R \in \mathbb{R}^{c \times N}$ represents the spectral response function (SRF) of the MS sensors, and $\varphi(Y)$ represents the predefined sophisticated prior. Based on the above objective function, several optimization meth-

Fig. 6. Diagrams of MS-based HS image SR: (a) Model-based methods; (b) Decomposition-based methods; (c) Other deep learning methods.

ods have been proposed and achieved good results. Although these methods are highly based on rigorous theoretical analysis, their reconstruction performance is quite limited due to the lack of end-to-end optimization and the inaccurate prior for real-world images.

In recent years, inspired by the great success of deep learning, several methods have attempted to combine deep learning with the traditional model-based method for better reconstruction performance. Considering the advantage of end-to-end training, several methods unfold the optimization function and use deep networks to solve sub-problems. Wei *et al.* [124] reformulated the optimization problem into three sub-optimization problems by half quadratic splitting (HQS) algorithm. Then, the one related to the regularizer is modeled by a deep denoising network, while the other two subproblems are further represented by network structure modules for end-to-end training. A CNN denoiser [125] is plugged into the alternating direction method of multipliers (ADMM) algorithm to estimate the coefficients, which is generalizable and flexible to cope with different HS datasets. Taking the low-rankness along the spectrum into consideration, MHF-net [126] designs an iterative algorithm to solve the optimization problem, and further unfolds the algorithm into the deep network. Although the domain knowledge is introduced into the HS observation model in MHF-net, it only uses the deep denoising network to regularize the intermediate HR MS images, leading to limited reconstruction performance. To overcome this drawback, MoG-DCN [96] proposes a deep convolutional network based on the back-projection technique for exploiting the spatial and spectral features to refine the HR HS image estimates. Moreover, the MoG-DCN method also approximates the spatial and spectral degradation operators by convolutional layers, which enhances the flexibility of the method. For the scenarios where the PSF and SRF are unknown, Wang *et al.* [127] proposed a deep blind iterative fusion network (DBIF), which achieves observation model estimation and fusion process iteratively and alternatively. The DBIF is further optimized by making the fusion process invertible in [128], which reduces the spectral and spatial distortion. Similarly, through learned PSF and SRF, the method in [129] generates the latent HR HS image by a generative neural network and attempts to degenerate it to the observed LR HS image and HR MS image.

Since recovering the HR HS image from the LR HS image is an ill-posed problem, many hand-crafted priors are introduced to regularize the solution space, such as nonlocal self-expressive prior, low-rank prior, and sparsity prior. However, these heuristic priors cannot well-represent the characteristics of the desired HR HS image, resulting in inferior reconstruction quality. With the rise of deep learning, many methods adopt deep neural networks to exploit the characteristics of the latent HR HS image. In [130], [131], a deep convolutional network is used to learn the data-driven deep prior, and the final output should obey both the physical model and the deep prior simultaneously. To regularize the final output HR HS image with the deep prior, many regularizers including the Euclidean distance in [130], [131], total variation in [132], and spatial-spectral gradient deviation in [133] are explored and have achieved considerable performance improvements. To
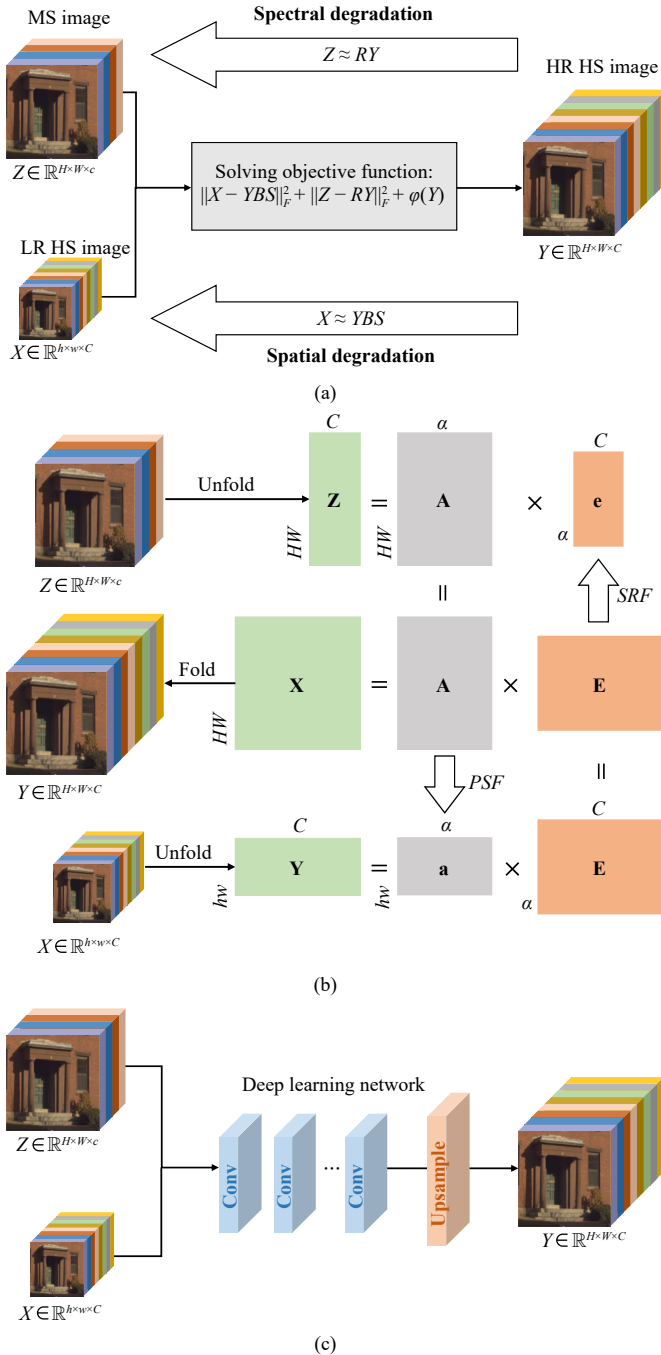
further balance the contribution of the physical model and the deep prior, [134] leverages the minimum distance criterion for seeking the optimal hyperparameters of the regularization term $\varphi(Y)$. Combining the degradation model and data prior, a variational fusion model [135] is regularized implicitly. To make full use of the information of patch under subspace representation, an interpretable patch-aware network [134] is deployed to unfold the subspace-based optimization model with two regularization terms representing pixel localization and texture. Most recently, Wan *et al.* [136] reconstructed the HS image under the guidance of RGB images through TV^3 minimization, which encourages the output HR HS image to be similar to both the deep prior and HR RGB guidance, and stay smooth simultaneously. Resorted to a novel spatio-spectral regularization, the DHIF method [137] enforces each pixel to be predicted by its spatial-spectral neighbors, which makes (3) differentiable. In this way, this approach is capable of optimizing iterative spatio-spectral regularization by the multi-stage network in an end-to-end manner.

The above methods solve the objective function in a supervised manner, where numerous labeled data are required. Differently from the supervised learning scheme, it is the first time that the unsupervised coupled CNN [138] was developed with learnable PSF and SRF for the fusion task and the only information the proposed method requires is the spectral coverage of the MS image and HS image, which is easy to obtain from the data provider. In [139], Gao *et al.* operated RGB-HS image fusion in a self-supervised manner, which gets rid of the dependence on the training dataset. In an unsupervised manner, the DSSP method [140] proposes an hourglass architecture to generate the HR HS image from a noisy input and regulates the output through the available HR RGB and LR HS images. To handle the unknown spatial degradation, UAL [141] designs a two-stage network to recover the latent HR HS image in a coarse-to-fine scheme, where a general image prior is first learned on synthetic data and further adapted to the specific HS image under unknown degradation in an unsupervised manner. Considering that the SRF is crucial for HS image SR, Fu *et al.* [142] designed an SRF optimization layer to automatically choose the best SRF, or to model the optimal SRF under specific physical constraints. Wei *et al.* [143] obtained the deep image prior through a 3D convolution network, and the pixel-aware refinement mechanism is introduced to generate intermediate reconstruction results for self-supervised learning. The unsupervised scheme can handle the dilemma that the HR HS image is unavailable, which has great potential and needs further exploration.

B. Decomposition-Based Methods

Based on the assumption that the HS image can be decomposed into a series of pure spectral vectors (endmembers) and the corresponding proportional coefficients (abundance) based on the linear spectral model, some methods have been proposed to super-resolve the HS image by estimating the abundance and endmembers of the HS image, the process of which is demonstrated in Fig. 6. Typically, with the unfolded HR HS image as $Y \in \mathbb{R}^{HW \times C}$, the corresponding HR MS image as $Z \in \mathbb{R}^{HW \times c}$, and the corresponding LR HS image as $X \in \mathbb{R}^{hw \times C}$, the

framework of decomposition-based methods can be formulated as follows:

$$Y^{HW \times C} = A^{HW \times \alpha} E^{\alpha \times C} \tag{4}$$

$$X^{HW \times c} = A^{HW \times \alpha} e^{\alpha \times c}, \quad e = ER \tag{5}$$

$$Z^{hw \times C} = a^{hw \times \alpha} E^{\alpha \times C}, \quad a = SA \tag{6}$$

where A and a refer to the abundance of the images, while E and e refer to the endmembers, which are related to the complexity of the images. R and S denote the SRF and the PSF respectively. Noticing that the HR HS image and HR MS image share the same abundance A, while the HR HS image and LR image share the same endmembers E, the key for decomposition-based methods lies in the exploitation of the shared information.

Based on the above observation, uSDN [144] first proposes an unsupervised coupled encoder-decoder structure. The two encoders aim to learn the abundance of the input HR MS image and LR HS image, respectively, and encourage the abundance to follow a Dirichlet distribution where the sum-to-one property is satisfied. The decoder is served as the endmember E and is shared by the two modalities. Based on the framework of uSDN, several methods have been designed for more realistic scenarios [145]–[147]. FusionNet [145] combines a variational probabilistic autoencoder to describe the nonlinear spectral mixture process and gives the fusion problem a meta-learning explanation. In [146], an unsupervised and unregistered network is proposed for the scenarios where the two input modalities are not well-registered. Zheng *et al.* [147] integrated the dense registration and the SR task into a unified model. Specifically, the nonrigid registration network acts as a link that re-establishes the spatial relations between the abundances extracted by the spectral unmixing. Most recently, the method in [148] improves the SR results by learning the SRF with the assistance of the endmember, which further relieves the SRF dependence for most current methods. Similarly, based on matrix decomposition, an unsupervised approach [149] that involves one implicit autoencoder is trained to estimate the HR prediction as well as PSF and SRF. Considering the nonlinear mixture, a nonlinear variational probabilistic generative model is tailored in [150] for the unsupervised fusion task, leading to an expressive end-to-end mapping to obtain the HR HS image. Attributed to the scalability of the inference and optimization of the unsupervised nature, this model is unsupervisedly trained based on the historical datasets in advance and processes the incoming test data in real-time.

C. Other Deep Learning Methods

Except for the above two categories, there are other methods that attempt to solve the MS-assisted HS image SR problem by designing network structures or modules for effectively exploiting the intrinsic data characteristic. Learning-based methods generally learn the non-linear mapping function, $\mathcal{F}_\theta$, linking the HR (target) image, Y, with the observed HS image, X, and the MS image, Z, i.e., $Y = \mathcal{F}_\theta(X, Z)$, where θ indicates the network parameters.

In the early stage, it is natural to apply the 3D CNN for the

HS data and [151] leverages dimensionality reduction to significantly reduce the computational time of the 3D CNN. With the rapid development of deep learning, some powerful architectures are introduced, including residual learning [152], pyramid structure [153], [154], dense connection [155], [156] and attention mechanism [157]–[163]. Since HS images and MS images differ in resolution, the multi-scale architecture is of great prosperity to exploring the rewarding information from different spatial sizes. Specifically, Han *et al.* [164] integrated the multi-level outputs into the cost function for more robust reconstruction results. The GDD method [165] learns semantic features from RGB images by a U-shaped network (UNet), and the semantic features are further used to guide the reconstruction of HR results. A dual-UNet structure is also adopted in [160], concentrating on different scales and depth features from MS images and injecting these features into HS images for better reconstruction. Related to this, Wang *et al.* [166] proposed a full-scale linked UNet with spatial-spectral joint perceptual attention for this fusion task. In [167], the LR HS image is fused with the HR MS image at different scales to not only fully utilize both low-level and high-level semantic information, but also alleviate the reconstruction difficulty. With the HR RGB guidance, GuidedNet [168] effectively predicts the HR residual details that are added to the upsampled HSI to simultaneously improve spatial quality and preserve spectral information. Additionally, the proposed approach is also suitable for other resolution enhancement tasks, such as pansharpening and single image SR. According to the assumption that the acquired HR MS image and the reconstructed HR HS image share similar underlying characteristics, Ha *et al.* [169] induced an unsupervised auxiliary task for multiscale fusion. Besides, an error feedback network coupled with multiscale feature learning is designed in [170] to iteratively learn the image fusion and degeneration.

In view of the spectral correlation, [171] extracts the features from the spectrum of each pixel in the LR HS image and its corresponding spatial neighborhood in the MS image. To fully exploit the rich spectral information of LR HS image and high-frequency edge details of MS images, the preservation of the color and structure information has gained much attention recently [159], [172]. Related to this, the SSR-NET method [173] is constrained under spatial edge loss and spectral edge loss. And the edge map of MS image generated via transfer learning is served as guidance for the edge-conditioned network to exploit low-level edge detail information [174]. In SCPNet [159], two powerful modules are employed to learn the structure details and spectral consistency, and a cross attention based strategy is used to preserve both valuable information jointly. The SFPN [172] method proposes a symmetrical feature propagation network to super-resolve the HR HS image in a course-to-fine manner. Since the commonly used L_1 loss function takes no gradient and spectral information into consideration, an RAP (RMSE, Angle, and Laplacian) loss is proposed in [167] to boost the final results.

Considering a pixel-wise spectral mapping, the spectral relationship is first learned between LR MS images and LR HS images in [175] and then a self-supervised fine-tuning strategy is exploited to transfer the learned spectral mapping to generate HR HS images. Related to this, considering the spectral SR task, the high-resolution HS image is learned from the spectral mapping between the HR MS image and the ground-truth HR HS one [176]. Sharing the insight that the spectrally degraded version of the HS image should be equivalent to the spatially degraded version of the MS image, Li *et al.* [177] adaptively learned the unsupervised fusion task through a three-stage procedure. The first two stages are designed to learn the degradation model and a coarse HR HS image, and the last stage uses the learned degradation model to further refine the HR result. Following an unsupervised way, [178] constructs a multi-attention-guided network without training data and a novel network structure is proposed as a regularizer for the unsupervised fusion problem. Besides, a physics GAN-based model is also introduced in [179] to estimate the degradation.

In addition, motivated by the classic wavelet decomposition-based method, the PZRes-Net method [156] learns the zero-centric residual image which contains high-frequency spatial details and proposes the mean-value invariant upsampling to avoid distortion. Although the above methods have achieved promising performance, one of their fundamental limitations is that the model trained on a specific dataset cannot be directly applied to another dataset, since HS images from different datasets have different band numbers due to the diversity of cameras. Thus, Zhang *et al.* [180] proposed a meta-learning SR framework (MLSR) to solve the dilemma. MLSR achieves the HS image SR for a wide variety of input-output band settings, which has great potential for further development. Recently, Hu *et al.* [181] first attempted to apply the Transformer architecture, estimating the spatial residual between the upsampled LR MS image and the desired HR HS image. In [182], an efficient Transformer-based network is also designed to capture interactions of spatial regions and interspectra dependencies. Similarly, considering the strong correlation among different patches in HS data and complementary information among patches with high similarity, a novel pyramid shuffle-and-reshuffle Transformer [183] supports the information interaction among global patches in an efficient manner for HS and MS image fusion.

For the real scene, the MS images may have non-overlapped bands of the original LR HS data. Thus, a novel CS-based method by combing neural network [184] is presented to construct the complicated nonlinear relationship between overlapped and non-overlapped bands of the original LR HS data and the trained network is mapped to the fused overlapped HR HS bands to estimate the non-overlapped HR HS bands.

In general, the model-based methods injecting deep network into the model-based algorithms significantly improve the reconstruction performance and provide good interpretability for the deep model, which is thus widely preferred. Similarly, decomposition-based methods also have better interpretability. The major defect is that the nonnegative matrix factorization is identically inaccurate. Thereby, incorporating the extra task into HS image SR inevitably brings about error accumulation. For other deep learning-based methods, the introduction of more powerful models leads to better
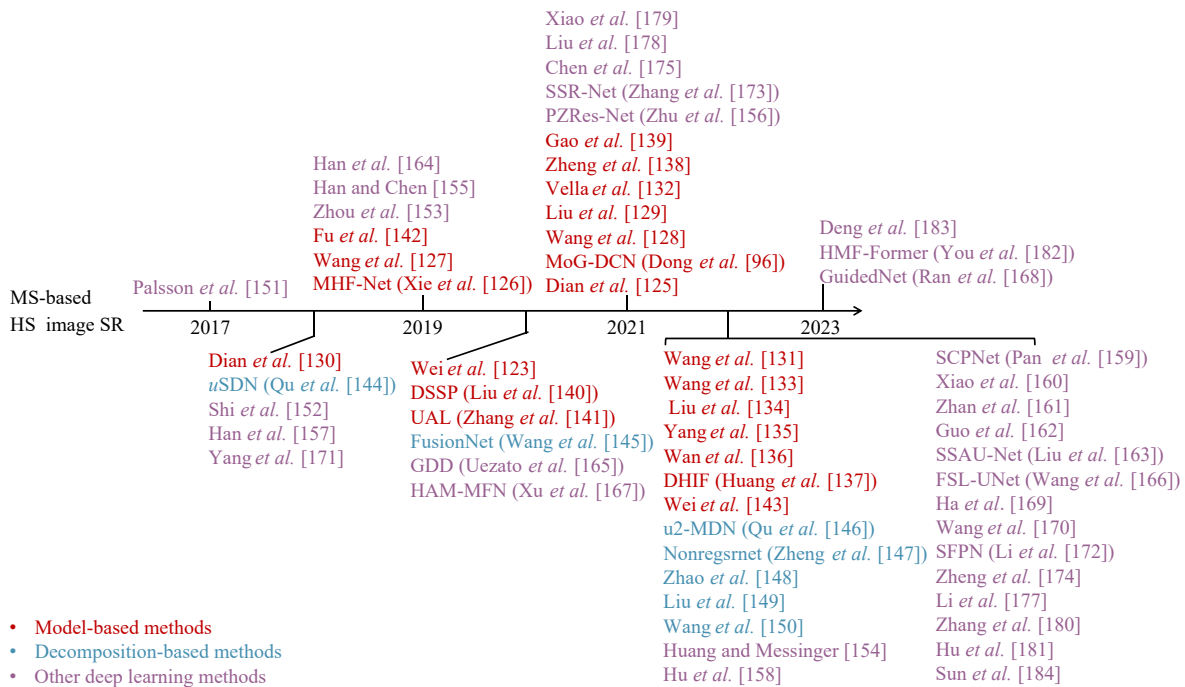
Fig. 7. The development of the MS-based HS image SR methods.

TABLE I
SUMMARY OF THE DATASETS IN THE HS IMAGE SR TASK IN TERMS OF THE DOWNLOAD LINK, COMMONLY USED CATEGORY (IN WHICH TASK USUALLY USE), NUMBER (THE NUMBER OF IMAGES IN THE DATASET), SIZE, AND WAVELENGTH RANGE.

Dataset	Download link	Commonly used category	Number	Size	Wavelength range (nm)
ICVL	https://icvl.cs.bgu.ac.il/hyperspectral/	MS	200	$1392 \times 1300 \times 31$	400−700
CAVE	https://www.cs.columbia.edu/CAVE/databases/multispectral/	Single, PAN, MS	32	$512 \times 512 \times 31$	400−700
Chikusei	https://naotoyokoya.com/Download.html	Single, PAN, MS	1	$2517 \times 2335 \times 128$	363−1018
Harvard	http://vision.seas.harvard.edu/hyperspec/d2x5g3/	Single, MS	50	$1040 \times 1392 \times 31$	400−700
Houston	https://hyperspectral.ee.uh.edu	PAN, MS	1	$349 \times 1905 \times 144$	380−1050
Pavia Centre	https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes	Single, PAN, MS	1	$1096 \times 715 \times 102$	430−830
Pavia University	https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes	Single, PAN	1	$610 \times 340 \times 103$	430−830
WDC Mall	https://rslab.ut.ac.ir/data	PAN, MS	1	$1208 \times 307 \times 191$	400−2400
Botswana	https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes	PAN	1	$1496 \times 256 \times 145$	400−2500
Cuprite	https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes	MS	1	$512 \times 614 \times 224$	370−2480

reconstruction performance with increasing computational complexity. Besides, we use a chart about the timeline in Fig. 7 to represent more clearly the developments in this field. Asshown in Fig. 7, model-based methods have been always receiving certain attention. With the development of deep learning structures, methods focusing on network structure design are flourishing.

VI. DATASETS

In this section, we will conclude the HS image datasets commonly used in the HS image SR task, the result of which is listed in Table I. According to the way of our classification, we report the category in which these datasets are commonly used. The "Number" column indicates the number of images in the dataset. It can be observed that depending on a large external dataset, the single HS image SR methods have pre-ferred the natural scene HS image dataset containing more than one imagery. Sharing similar scenarios and nearly over-lapping spectral ranges, the CAVE and Harvard datasets are combined together to train the deep model. For the PAN and MS-based SR tasks, since the auxiliary data can be simulated from the HS data, these methods are more flexible in select-ing datasets, varying in different settings. We also demon-strate some representative datasets in Fig. 8.

A. ICVL Hyperspectral Dataset

ICVL HS dataset [185] was acquired by the International Computational Vision Laboratory (ICVL) including 200 images. All the HS images in the ICVL dataset were acquired using a Specim PS Kappa DX4 HS camera and a rotary stage for spatial scanning. The database images were acquired by a Specim PS Kappa DX4 HS camera and a rotary stage for spa-
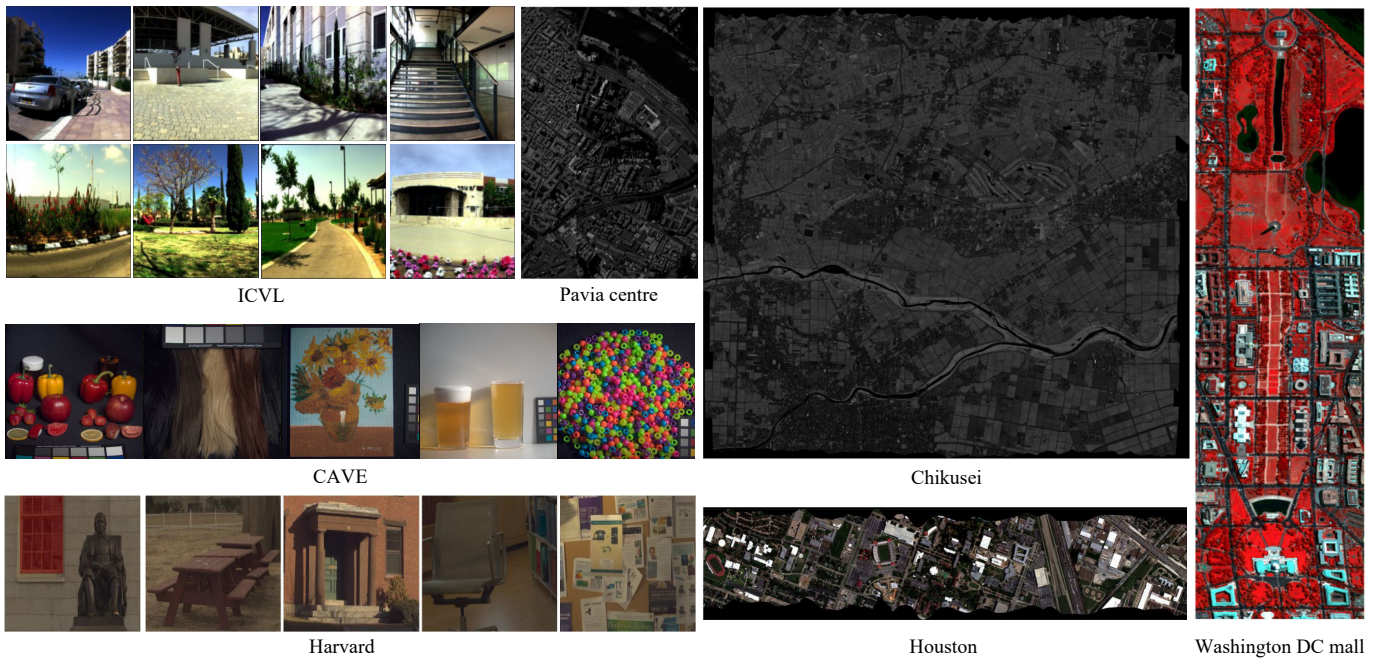
Fig. 8. Samples in the commonly used datasets. RGB images in ICVL, CAVE, and Harvard are provided. One band in Pavia Centre is selected for visualization as well as that in Chikusei. The Houston and Washington DC Mall are given in pseudo-RGB versions.

tial scanning. HS images were collected with the spatial resolution of 1392×1300 over 519 spectral bands in the raw format ($400 - 1000$ nm at roughly 1.25 nm increments). For convenience, the ICVL laboratory downsampled these images to 31 spectral channels from 400 nm to 700 nm at 10 nm increments, provided in the mat file. This dataset presents a variety of natural scenes including urban and rural areas, and many HS images are overlapped as they are captured under the same scene with just 1-minut or 2-minute interval. Some RGB descriptions are exhibited in Fig. 8.

B. CAVE Dataset

The CAVE dataset [186] was captured by a Cooled CCD camera, involving a wide variety of real-world materials and objects. The HS camera records full spectral resolution reflectance information from 400 nm to 700 nm at 10 nm steps. This dataset consists of 32 HS scenes with a size of $512 \times 512 \times 31$ pixels, which can be divided into 5 sections, namely food and drinks, skin and hair, paints, real and fake, and stuff (e.g., feathers, flowers, superballs, etc.). Each scene also contains a single representative color image, displayed using sRGB values rendered under a neutral daylight illuminance. We select 5 image from 5 classes to show this dataset in Fig. 8.

C. Hyperspec Chikusei Dataset

The Chikusei dataset [187] was taken by the Headwall Hyperspec-VNIR-C imaging sensor over agricultural and urban areas in Chikusei, Ibaraki, Japan. The HS data has 128 bands in the spectral coverage from 363 nm to 1018 nm. The image scene consists of 2517×2335 pixels and the ground sampling distance was 2.5 m. Since the lack of edge information, we do not visualize the HS dataset. For one scene dataset, methods usually cut the edge of the original HS image

and split the big image into a training set and a test set in the experiment. Therefore, performance evaluation on this dataset largely depends on the dividing way.

D. Harvard Dataset

The Harvard dataset [188] collected a database of fifty images under daylight illumination, both outdoor and indoor scenes featuring a diversity of objects, and materials. These HS images were captured by a commercial HS camera, each with approximately 10 nm bandwidth and centered at steps of 10 nm from the wavelength range of 400 nm to 700 nm. This dataset has 50 HS images of $1040 \times 1392 \times 31$ size, five RGB samples of which are exhibited in Fig. 8.

E. University of Houston Dataset

The Houston dataset was provided by the Hyperspectral Image Analysis group and the NSF Funded Center for Airborne Laser Mapping (NCALM) at the University of Houston. The dataset was originally used for the scientific purposes of the 2013 IEEE GRSS Data Fusion Contest. The HS imagery was acquired by the ITRES CASI-1500 sensor over the University of Houston campus with a ground sampling distance of 2.5 m, including both landcover and urban regions. The HS image has $349 \times 1905 \times 144$ pixels taken in the 380 nm to 1050 nm wavelength range.

F. Pavia Dataset

The Pavia dataset was acquired by the reflective optics system imaging spectrometer (ROSIS) during a flight campaign over Pavia, Northern Italy, composed of two scenes: Pavia Centre and Pavia University (PaviaU). The Pavia Centre has 1096×1096 pixels in the spatial resolution, but some regions contain no information and have been discarded before analysis, leaving 1096×715 pixels. It was initially composed of 115

bands that have been reduced to 102 bands after removing the water vapor absorption and noise bands. The PaviaU dataset was taken by the same sensor over the University of Pavia, Italy, with a ground sampling distance of 1.3 m. The PaviaU is commonly cropped to 610×340 with 103 spectral bands taken within the wavelength range of $430 - 830$ nm since the empty regions are eliminated.

G. Washington DC Mall Dataset

The Washington DC (WDC) Mall Dataset was acquired by the hyperspectral digital image collection experiment (HYDICE) sensor, taken over the National Mall in Washington, DC, USA. The sensor system used in this case measured pixel response in 210 bands in the 400 to 2400 nm region of the visible and infrared spectrum. Bands in the 900 nm and 1400 nm region where the atmosphere is opaque have been omitted from the data set, leaving 191 bands covering the wavelength range from 400 nm to 2400 nm. The whole image contains 1208 scan lines with 307 pixels in each scan line. In Fig. 8, we show the simulated color view of imagery.

H. Botswana Dataset

This data taken over Okavango Delta, Botswana was acquired by Hyperion sensor on NASA EO-1 satellite at 30 m pixel resolution over a 7.7 km strip in 242 bands covering the $400 - 2500$ nm portion of the spectrum in 10 nm windows. To mitigate the effects of bad detectors, inter-detector miscalibration, and intermittent anomalies, the HS image was first pre-processed, and un-calibrated and noisy bands that cover water absorption features were removed, reducing to 145 bands with a spatial size of 1496×256.

I. Cuprite Dataset

This Cuprite set was retrieved from AVIRIS over the Cuprite mining district in Nevada, USA. The HS imagery has $512 \times 512 \times 224$ pixels covering the wavelength range of 370-2480 nm with 10 nm interval. After removing the noisy channels $(1 - 2$ and $221 - 224)$ and water absorption channels $(104 - 113$ and $148 - 167)$, this dataset remains 188 channels. Cuprite is the benchmark dataset for HS unmixing research.

VII. EVALUATION METRICS

To introduce the commonly used evaluation metrics, we denote $\hat{Y}$ and Y $(\hat{Y}, Y \in \mathbb{R}^{H \times W \times C})$ as the prediction and corresponding ground-truth image, respectively.

The peak signal-to-noise ratio (PSNR) index is the mean ratio between the maximum power of the image and the power of the residual errors of the spectral bands. For the i-th spectral band, the PSNR is calculated by

$$\text{PSNR}(Y_i, \hat{Y}_i) = 10 \cdot \log_{10}\left(\frac{\max(Y_i)^2}{mean\|Y_i - \hat{Y}_i\|_2^2}\right). \quad (7)$$

A higher PSNR value indicates a better image quality of the reconstructed HS image. The average value of all spectral bands is the PSNR for the HS image.

The spectral angle mapper (SAM) [189] is used to quantify the spectral information preservation at each pixel. More precisely, this index calculates the angle between two vectors of the estimated and reference spectra to represent spectral similarity. The SAM values near zero indicate high spectral similarity with no spectral distortion. The SAM is defined as

$$\text{SAM}(Y, \hat{Y}) = \arccos\left(\frac{Y \cdot \hat{Y}}{\|Y\|_2 \|\hat{Y}\|_2}\right). \quad (8)$$

The correlation coefficient (CC) is another widely used indicator measuring the spectral quality of the predicted images. It calculates the correlation coefficient between the prediction and the corresponding reference image as

$$\text{CC}(Y_i, \hat{Y}_i = \frac{\sum_{w=1}^{W}\sum_{h=1}^{H}(Y_i - \mu_{Y_i})(\hat{Y}_i - \mu_{\hat{Y}_i})}{\sqrt{\sum_{w=1}^{W}\sum_{h=1}^{H}(Y_i - \mu_{Y_i})^2 \sum_{w=1}^{W}\sum_{h=1}^{H}(\hat{Y}_i - \mu_{\hat{Y}_i})^2}} \quad (9)$$

where W and H are the width and height of the image, μ indicates mean value of an image. CC ranges from -1 to 1, and the best value is 1. The average value of all spectral bands is the CC for the HS image.

ERGAS measures the band-wise normalized root of mean square error (RMSE) between the reference HS image Y and the reconstructed HS $\hat{Y}$, with the best value at 0. It is defined as

$$\text{ERGAS}(Y, \hat{Y}) = \frac{100}{s}\sqrt{\frac{1}{C}\sum_{i=1}^{C}\frac{mean\|Y_i - \hat{Y}_i\|_2^2}{mean(Y_i)^2}} \quad (10)$$

where s is the scale factor between the LR HS and HR HS, and C is he number of spectral bands in HS data.

The structural similarity (SSIM) index is to calculate the structural similarity of two images. For the i-th spectral band, it is defined as

$$\text{SSIM}(Y_i, \hat{Y}_i) = \frac{(2\mu_{Y_i}\mu_{\hat{Y}_i} + c_1)(2\sigma_{Y_i, \hat{Y}_i} + c_2)}{(\mu_{Y_i}^2 + \mu_{\hat{Y}_i}^2 + c_1)(\sigma_{Y_i}^2 + \sigma_{\hat{Y}_i}^2 + c_2)} \quad (11)$$

where μ_{Y_i} and $\mu_{\hat{Y}_i}$ are the means of Y_i and $\hat{Y}_i$, respectively, σ_{Y_i} and $\sigma_{\hat{Y}_i}$ are the variances of Y_i and $\hat{Y}_i$, respectively, $\sigma_{Y_i, \hat{Y}_i}$ is the covariance of Y_i and $\hat{Y}_i$, c_1 and c_2 are constants set as 0.0001 and 0.0009, respectively. The best value of SSIM is 1. The mean SSIM is estimated by averaging the SSIM values of all bands in HS data.

To overcome some disadvantages of RMSE, the universal image quality index (UIQI) [190] is proposed to estimate the global spectral quality as

$$\text{UIUQ}(Y_i, \hat{Y}_i) = \frac{4\mu_{Y_i}\mu_{\hat{Y}_i}}{\mu_{Y_i}^2\mu_{\hat{Y}_i}^2} \cdot \frac{\sigma_{Y_i\hat{Y}_i}^2}{\sigma_{Y_i}^2\sigma_{\hat{Y}_i}^2} \quad (12)$$

in which σ and μ represent the variance and mean, respectively. The average value of all spectral bands is the UIQI for the HS image. The larger value of UIQI means better estimated results.

VIII. EXPERIMENTS

A. Experiment on Single HS Image SR

In this section, seven methods are selected as the representative methods for comparison: 3DFCNN [59], GDRRN [46],

| Method	Scale	Parameter	Time (s)	PSNR ↑	SAM ↓	CC ↑	RMSE ↓	ERGAS ↓
3DFCNN [59]	4	39 K	0.291	34.647	3.624	0.9417	0.0142	5.454
GDRRN [46]	4	219 K	0.095	35.027	3.565	0.9501	0.0139	5.286
aeDPCN [81]	4	580 K	0.084	35.312	3.229	0.9683	0.0126	5.107
SSPSR [47]	4	12 875 K	0.306	35.795	2.905	0.9869	0.0087	4.756
RFSR [48]	4	1603 K	0.257	35.984	3.012	0.9835	0.0092	4.828
SFCSR [54]	4	1233 K	1.129	35.601	3.145	0.9746	0.0107	4.933
Interactformer [85]	4	2407 K	1.408	36.121	2.898	0.9875	0.0089	4.741
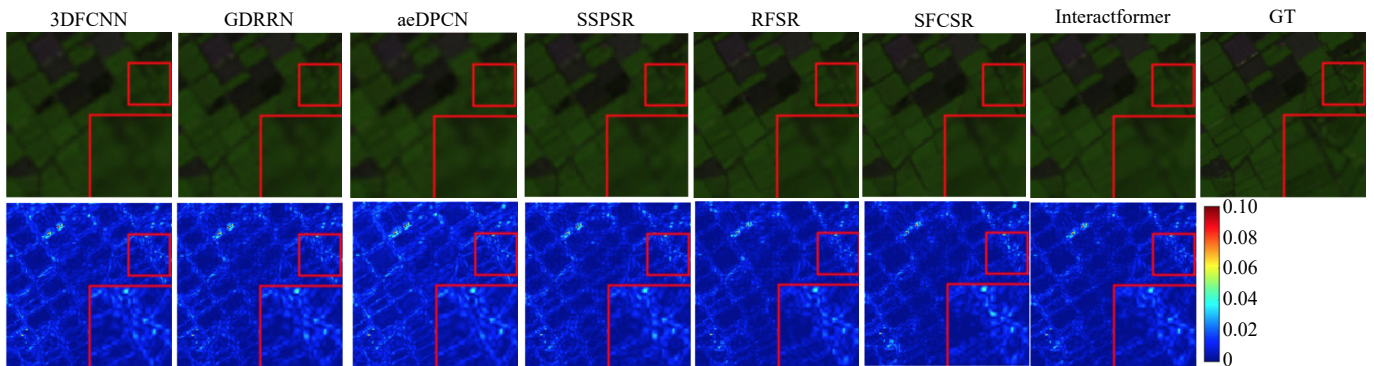
Fig. 9. Qualitative results of representative methods for single HS image SR task.

aeDPCN [81], SSPSR [47], RFSR [48], SCFSR [54] and Interactformer [85]. Among them, GDRRN, SSPSR, and RFSR are group-based methods. The aeDPCN method is a two-step approach. 3DFCNN and SCFSR are of 3D structures. Besides, Interactformer is based on the Transformer structure. For evaluation, the simulated experiment is conducted on the Chikusei dataset [187] with the scale factor 4. The HS imagery is cropped into image patches without overlapping to form the training and testing datasets. The original image patches are treated as ground-truth HR images, and the LR inputs are produced by the Bicubic interpolation. The comparison methods are evaluated by five indexes: PSNR, SAM [189], CC, RMSE, and ERGAS [191], the result of which is listed in Table II. We also report the number of parameters and computation time of comparing methods on the Tesla P100. The qualitative results of the Chikusei dataset are shown in Fig. 9 as well as the error maps.

As shown in Table II, the Interactformer is superior to other methods in terms of all evaluation metrics but with more parameters. It can be observed that 3DFCNN and GDRRN fail in obtaining competitive results due to the shallow structure. For effective feature extraction, 3DFCNN applies 3D convolution to fully explore the spatial-spectral correlation. However, the computational complexity of 3D convolution is huge, especially in the HR space, resulting in lower computation efficiency. GDRRN first introduces the group strategy to process the high-dimensional HS image. Subsequently, the group-based methods (SSPSR and RFSR) toiler the deep model for the group data, realizing considerable performance improvement. It can be seen that the group-based methods

have short running times even with large parameters. Since the 2D convolution does not have enough ability to extract spectral information while processing spatial information, SFCSR uses a mixture of 2D and separable 3D convolution to extract spectral and spatial features. Interactformer achieves remarkable results benefiting from the Transformer and 3D convolution network but it has a longer computation time due to the complex structures. We can find that the aeDPCN method does not perform well. The main reason may be that the unmixing task introduced in the HS image SR problem would bring about unexpected errors.

As demonstrated in Fig. 9, for early models, such as 3DFCNN and GDRRN, they would produce blurry results with large reconstruction errors. Due to the shallow structure, these methods have limited expression ability. Meanwhile, there are some unexpected artifacts that appear in the prediction of the aeDPCN method, which may be caused by the unmixing task. For these group-based methods (SSPSR, RFSR, and SFCSR) with little difference in performance, the difference in reconstruction results is small, but there is still a gap with the ground truth. Owing to the Transformer structure, the advanced Interactformer method could recover more details and have a comparatively small error map.

B. Experiment on PAN-Based HS Image SR

In this section, eight methods are selected as the representative methods for comparison: RE-RANet [122], DDLPS [89], CCNNF [97], HyperPNN [103], DHP-DARN [100], DBDE-Net [117], RHDN [107] and HyperTransformer [112]. Specifically, DDLPS and CCNNF are model-guided methods. Hyper-

TABLE III
QUANTITATIVE COMPARISONS ON DOWN-RESOLUTION AND FULL-RESOLUTION OF REPRESENTATIVE PAN-BASED METHODS EVALUATED
ON BOTSWANA DATASET AT THE SCALE FACTOR 3 IN TERMS OF COMPUTATION TIME, PSNR, SAM, CC, RMSE AND ERGAS INDEXES.

| Method	d	Time (s)	PSNR ↑	SAM ↓	CC ↑	RMSE ↓	ERGAS ↓
RE-RANet [122]	3	0.745 / 3.524	26.3064 / 24.5396	6.0458 / 6.1548	0.9012 / 0.8675	0.0263 / 0.0307	1.2676 / 1.3343
DDLPS [89]	3	0.848 / 4.350	27.3945 / 25.6895	4.7763 / 4.4095	0.9311 / 0.9378	0.0270 / 0.0360	1.0768 / 1.1660
CCNNF [97]	3	0.580 / 2.154	26.0061 / 24.2957	3.0517 / 2.7004	0.8984 / 0.8684	0.0253 / 0.0297	1.3612 / 1.3773
HyperPNN [103]	3	0.675 / 2.341	33.9561 / 33.5801	1.5755 / 1.2883	0.9808 / 0.9813	0.0105 / 0.0111	0.5978 / 0.5375
DHP-DARN [100]	3	0.658 / 2.381	34.9803 / 34.1990	1.4413 / 1.2110	0.9846 / 0.9837	0.0096 / 0.0104	0.5281 / 0.4976
DBDENet [117]	3	0.885 / 2.549	34.9656 / 31.8786	1.3472 / 1.6113	0.9847 / 0.9727	0.0086 / 0.0147	0.5399 / 0.6234
RHDN [107]	3	2.087 / 4.821	32.3060 / 30.3349	1.4044 / 1.5538	0.9410 / 0.9411	0.0125 / 0.0153	1.6686 / 1.3597
HyperTransformer [112]	3	1.709 / 5.087	35.7493 / 30.0509	1.2287 / 1.7605	0.9871 / 0.9642	0.0078 / 0.0169	0.5022 / 0.7117
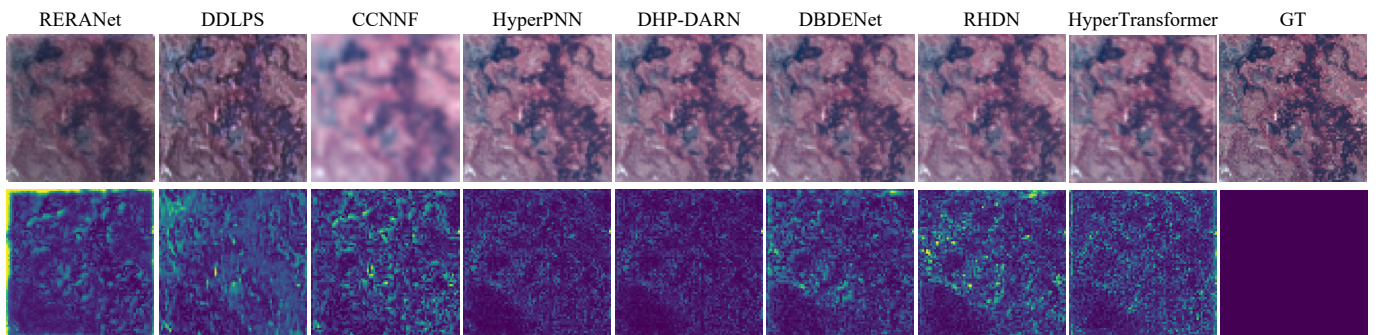
Fig. 10.   Qualitative full-resolution results of representative methods for PAN-based SR.

PNN and DHP-DARN reconstruct the HR HS image in a straight fusion manner. DBDENet, RHDN, and HyperTransformer belong to multistage methods. RE-RANet is an unsupervised algorithm. We conduct the full-resolution and down-resolution experiments on the Botswana dataset and the patch with the size of 216×216 is cropped as full-resolution data. The PAN image is generated by averaging the first thirty bands. All deep models are trained on the down-resolution data and tested on the full-resolution and down-resolution, respectively. The scale factor is 3. The comparison methods are evaluated in terms of computation time, PSNR, SAM, CC, RMSE, and ERGAS.

The evaluation performance is reported in Table III. According to the result, model-guided methods (DDLPS and CCNNF) have limited reconstruction performance. Since only several steps are replaced by deep learning models, the whole framework still involves some traditional steps, such as unmixing in CCNNF and guided filter in DDLPS, leading to slower inference times, especially in full-resolution testing. It can be seen that the straight fusion methods (HyperPNN and DHP-DARN) perform well in both full-resolution and reduced-resolution scales with good generalization performance because of the end-to-end training. Through multistage fusion (DBDENet, RHDN, and HyperTransformer), such approaches improve the SR performance, while in full-resolution testing, they suffer from a decrease. Due to the complex structures, RHDN and HyperTransformer have lower computation efficiency. To date, a few works have paid attention to the unsupervised algorithm. In terms of quantitative indexes, RE-RANet fails in acquiring competitive results.

Since the spectrum of HR prediction is regularized with that of the LR HS image in RE-RANet, the spectral characteristics have changed in the sampling process. Thus, it is difficult to avoid spectral distortion.

Figs. 10 and 11 show the full-resolution and down-resolution super-resolved images of the eight representative methods as well as the error maps. From a visual point of view, in the down-resolution scale, it can be seen from Fig. 11, the multistage methods (DBDENet, RHDN, and HyperTransformer) behave best in reconstructing results. The model-guided methods (DDLPS and CCNNF) have obvious reconstruction errors. Whereas, as for the full-resolution results in Fig. 10, the straight fusion (HyperPNN and DHP-DARN) model realizes superior performance. As shown in Fig. 10, the unsupervised and model-guided algorithms all generate blurry images, while large errors occur in multistage methods.

C. Experiment on MS-Based HS Image SR

In this section, seven methods are selected as the representative methods for comparison: uSDN [144], MHF-net [126], Wei et al. [143], PZRes-Net [156], UAL [141], MoG-DCN [96] and u^2-MDN [146]. Among them, MHF-net, UAL and Wei et al. [143] are model-based methods. For decomposition-based approaches, we choose uSDN and u^2-MDN for comparison. And PZRes-Net is selected as a representative for other deep learning methods. We perform the experiment on the Harvard [188] dataset at the scale factor 32. The corresponding RGB image for the HR MS image is obtained by Canon EOS 5D Mark II coupled with the HR HS image.

Fig. 12 shows the super-resolved images of the 7 represen-
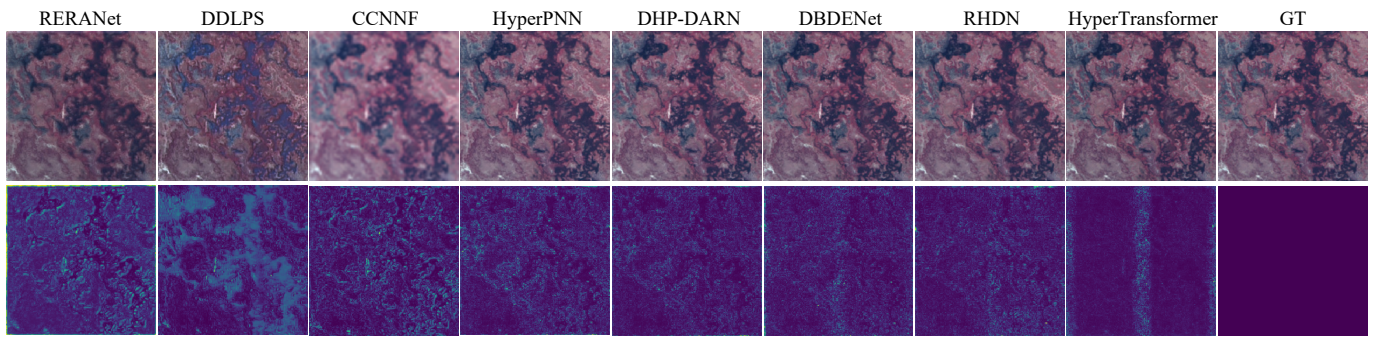
Fig. 11.   Qualitative down-resolution results of representative methods for PAN-based SR.
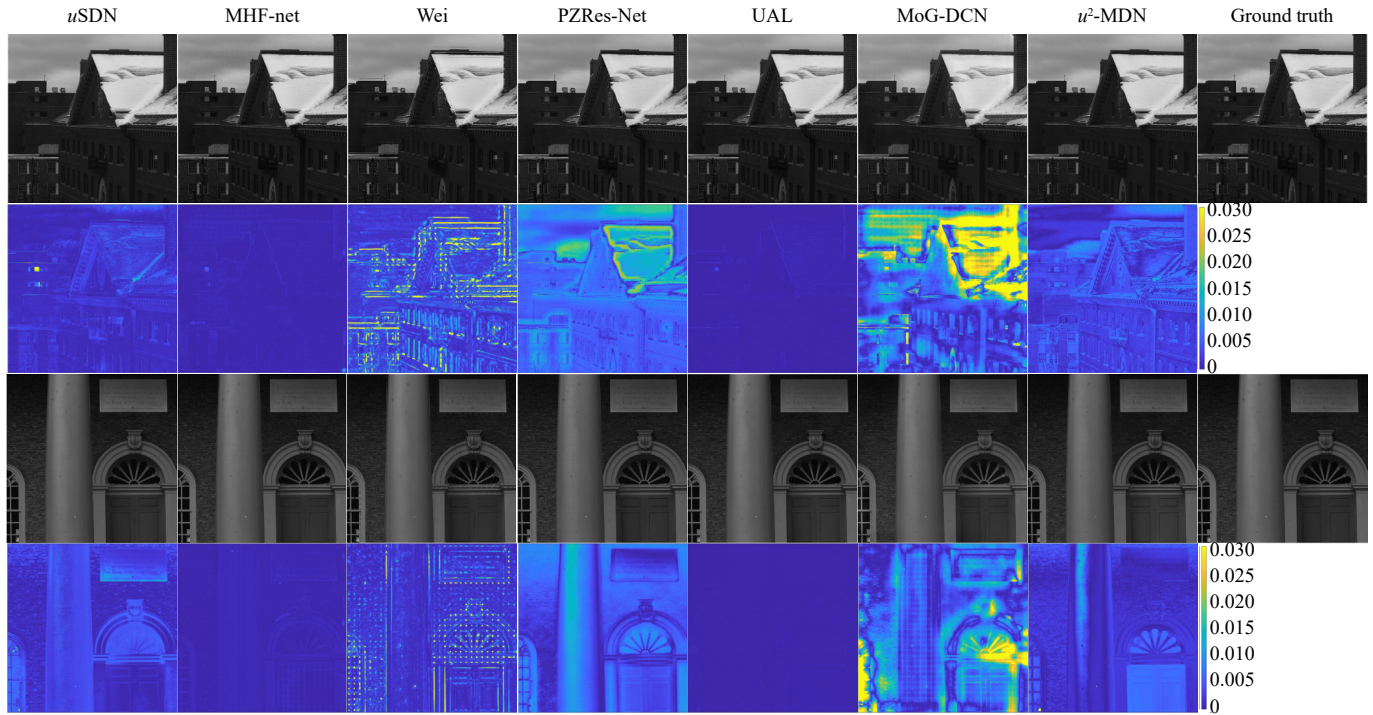

Fig. 12.   Qualitative results of representative methods for MS-based HS image SR task.

TABLE IV
QUANTITATIVE COMPARISONS OF REPRESENTATIVE MS-BASED HS IMAGE SR METHODS EVALUATED ON HARVARD DATASET AT THE
SCALE FACTOR 32 IN TERMS OF COMPUTATION TIME, PSNR, SAM, CC, RMSE AND ERGAS INDEXES.

Method	d	Time (s)	PSNR ↑	SAM ↓	CC ↑	RMSE ↓	ERGAS ↓
*u*SDN [144]	32	43.921	38.7311	5.2048	0.9821	4.3133	0.4513
MHF-net [126]	32	51.653	42.2384	3.9076	0.9900	3.2211	0.3310
Wei *et al.* [143]	32	0.493	36.9394	4.0163	0.9759	4.5486	0.5448
PZRes-Net [156]	32	0.055	39.9231	3.5939	0.9889	3.2994	0.3621
UAL [141]	32	0.793	44.3841	2.6643	0.9927	2.2581	0.2609
MoG-DCN [96]	32	0.743	36.6360	3.2957	0.9839	4.7215	0.4675
u^2-MDN [146]	32	46.207	38.3421	5.2618	0.9742	4.2218	0.4836

tative methods, and the corresponding average evaluation performance as well as the computation time are shown in Table IV. Based on the experimental results, some important observations can be drawn as follows: *u*SDN and u^2-MDN are two representative unsupervised decomposition-based methods, both of them take a single image as input, without rely-

ing on the external dataset. Therefore, these methods should be reoptimized on each input, resulting in a longer running time. Since their models are not trained on the entire HS dataset, the reconstruction performance is not stable, and some high-frequency edge details cannot be well recovered. For the model-based method, MoG-DCN can generate results with
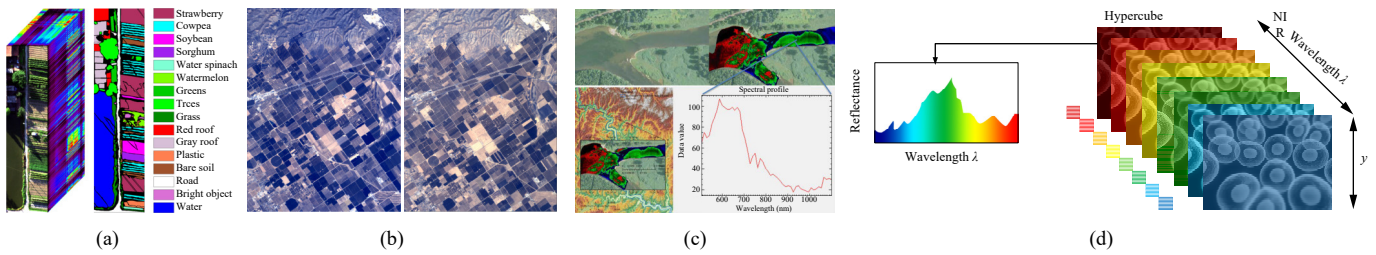
Fig. 13.   Examples of three applications: (a) Ground object classification of the WHU-Hi-HanChuan dataset [192]; (b) Urban changes, taken on the years 2013 and 2015 with the AVIRIS sensor over the Santa Barbara region, California[1]; (c) Ecosystem monitoring of submerged aquatic vegetation and the possible invasive algae[2]; (d) Hyperspectral medical imaging.

relatively good spectral fidelity, however, there are unfavorable artifacts in reconstruction results, which is obvious in error maps exhibited in Fig. 12. The super-resolved images by Wei *et al.* [143] contain fake lines and dots, resulting in poor visual effects. MHF-net carefully designs a fusion network with high interpretability according to the unfolded optimization algorithm, and the spatial information of the image is well reconstructed. Since the MHF-net involves iterative optimization, the inference time is long. UAL sufficiently utilizes the knowledge from both the HS images and RGB images through the proposed self-guiding module, leading to both spatial-spectral consistent results. Unlike the two types of methods mentioned above, PZRes-Net fails to recover competitive results because it realizes the RGB and HS image fusion from the perspective of the wavelet decomposition. The zero-mean normalization applied in the network changes the pixel value distribution, resulting in significant global deviations between the reconstructed image and the ground-truth image.

IX. Applications

HS image SR can effectively improve the spatial resolution and provide more high-frequency details for high-level computer vision tasks, so as to improve the performance of these applications. We will briefly introduce several typical applications, e.g., ground object classification, urban change detection, ecosystem monitoring and medical imaging, which can intuitively demonstrate the importance of HS image SR. Examples of three applications are exhibited in Fig. 13.

A. Ground Object Classification

Ground object classification and mapping are essential in forestry, agriculture, mineral, and other environmental studies, including plant classification [193], soil composition estimation [194], mineral exploration [195] and many other applications. Specifically, remote sensing and land-use vegetation classification is seldom performed without ground truth or the collection of reference data. Spectrometers used in forestry investigations can obtain spectral signatures for the classification and mapping of vegetation, ecosystem productivity, and crop type, or yield. The soil composition controls the dynamics of various agrochemical processes in the soil. Thereby,

monitoring, mapping, and describing the spatial variability of soil composition are the key prerequisites for understanding the effects of agricultural practices on soil composition changes. In earth science, geologic scientists have utilized the advantages of HS imaging in different geological applications, such as the mineral industry, water quality determination, oil, and gas industries. HS imaging at various scales can be diagnostic in mineralogic and lithological mapping in different climate and tectonic settings.

Due to distinct spectral response features, ground objects are distinguishable through their spectral patterns. Therefore, by recording the subtle reflection information in the spectrum, HS imagery as one of the available data sources in remote sensing can classify the ground object within large areas, which should be provided in a sufficient spatial resolution. Nowadays, some investigations have demonstrated benefits derived from the improvement of spatial resolution through SR, which motivated the development of a large number of techniques for HS image SR.

B. Urban Change Detection

Urban change detection uses multiperiod remote sensing satellite image technology to judge whether changes have occurred, to determine where changes have occurred, and to identify the types of changes. Timely and accurate change detection in the urban is rather significant for urban development and resource management [6]. Remote sensing technology provides a large-scale view of the landscape over a long period of time and has been demonstrated to be an efficient method for change detection. HS sensors measure radiance by a large number of bands covering a wide spectral range. Owing to the fine spectral resolution, HS data supply more detailed information on spectral changes than MS data, improving the change detection performance. An important issue to be addressed is the effect of different spectral and spatial resolutions on prediction. Besides, the signal-to-noise ratio (SNR) of sensors is another important issue that affects detection ability. Fortunately, the HS image SR techniques could overcome the spatial resolution limitation, further promoting the development of urban change detection [196].

C. Ecosystem Monitoring

Ecosystem monitoring refers to the observation and investigation of land, sea, and vegetation by analyzing HS images to prevent disasters, such as flooding [197] and [198], fire, biological invasion, etc. However, with the wealth of spectral

---

[1] https://citius.usc.es/investigacion/datasets/hyperspectral-change-detection-dataset.
[2] https://eros.usgs.gov/doi-remote-sensing-activities/2015/usgs/hyperspectral-analysis-submerged-aquatic-vegetation.

information, the spatial resolution is sacrificed in HS data. The spatial property of the HS image depends on the design of the sensor in terms of its field of view and the altitude at which it operates above the surface. Each of the detectors in an HS sensor measures the energy received from a finite patch on the ground surface, which is hard to upgrade. In other words, the low spatial resolution of HS images restricts the further improvement of the accuracy of ecosystem monitoring. The HS image SR is evolving as a robust technique to improve spatial resolution. With or without the auxiliary images, the SR techniques for HS images can get reliable results from image processing [199].

D. Medical Imaging

HS imaging can obtain broad-area images of tissues and provide spatial and spectral information about some tissue samples. It reflects the quality features such as the size and shape of these samples, as well as their internal texture structure and composition differences, which plays an important role in medical diagnostic and surgical guidance [200], [201].

However, since the deep-learning techniques in HS image processing are still at the stage of theoretical development and technical exploration, hyperspectral medical diagnosis is limited by the bottleneck of HS image processing. How to extract richer information at high spectral resolution and spatial resolution without losing some details, is still a challenge to be tackled in spectral image processing. To this end, HS image SR could provide fine spatial resolution for HS medical diagnosis, improving diagnostic accuracy. Since hyperspectral images are acquired in a large number of bands, the images contain a lot of useful information with superfluous information such as background and electrical noise, which makes the analysis of the images difficult. Thereby, HS image SR is of great significance in HS image pre-processing. As HS imaging continues to evolve, more studies refine the algorithm to ensure the accountability of HSI analysis for routine clinical use [202].

X. CHALLENGES AND GUIDELINES

The HS image SR has made significant progress in the past decade. However, there still remain some challenges. We present the existing challenges and new guidance for the HS image SR.

A. Degradation Process

A general solution to the HS image SR problem is assuming that the degradation process (PSF or SRF) is known and fixed, e.g., bicubic downsampling and Gaussian blurring, and then constructing a deep neural network and performing SR reconstruction through data-driven. Thus, in the case of the same degradation in the simulated experiment, these methods have realized increasing objective SR performance based on the perfectly-known PSF or SRF. However, in real-world scenarios, the image degradation models are usually diverse and unknown due to the imaging and transmission process. There is often a mismatch between the simple assumptions adopted in existing HS image SR algorithms and the degradation of real inputs. Thus, even if the degradation process is slightly

different, the performance of these methods will be severely decreased, which brings difficulties to the practical application of these data-driven methods in real data. In recent years, a few scholars have paid attention to the practical HS image SR problem where the degradation process is unknown, trying to estimate the degradation process in advance [129].

B. Image Registration

The popular HS image SR methods are fusion-based approaches that super-resolve the LR HS image with the assistance of auxiliary images of high spatial resolution. However, the introduction of auxiliary images inevitably brings a new problem: image registration, which is also a difficult task to be solved. The image registration aims to geometrically align two images of the same scene acquired by different sensors or at different times or from different views. Since the auxiliary image is requested to be captured at the same scene as the observed one, image registration is a crucial pre-processing step, especially in the real scenario where an auxiliary image exists. Although image registration approaches [203]–[205] have been extensively studied, most fusion-based HS image SR methods assume that the auxiliary image and the input LR HS image are well aligned. On the contrary, there is no case that remote sensing satellites provide data from multiple sources without deformation. Recently, several approaches consider image registration with the fusion task to reduce the distortion, which needs further development [146], [147].

C. Zero-Shot Learning

Zero-shot learning intends to train the deep learning model on the data to be super-resolved. In the training phase, the LR image is learned from the observation, which conforms to the real scenarios. In this way, the paired LR-HR training data is scene-specific and degradation-specific, benefiting the testing procedure that the observed image is used as the input of the SR model. Besides, large amounts of paired data are no longer required to drive deep models. In view of the properties of HS images, zero-shot learning has two advantages in the HS image SR task. First, most of the HS image dataset consists of one HS imagery according to Table 8. The insufficient data is suitable for zero-learning methods. Second, since the observational models, spectral range, recording scene, and the number of spectral bands differ in different datasets, a well-trained SR model derived from one dataset can not be used for the other dataset. Consequently, the existing learning-based HS image SR methods for HS images have poor generalization ability. In this case, zero-shot learning can be easily adaptive for any HS image dataset. Therefore, zero-shot learning would be an important topic in future HS image SR research.

XI. CONCLUSIONS

As an algorithm technique, HS image SR has played a critical role in various fields, including the classification and mapping, change detection and remote sensing monitoring. In recent years, an increasing number of HS image SR methods based on deep learning have been presented to improve spatial resolution of the HS image. In order to provide effective reference and understanding for relevant researchers and tech-

nicians, this survey comprehensively analyzes the latest deep learning-based methods in the field of HS image SR. We discuss the characteristics of various methods with or without auxiliary information and directions for further improvement are also provided. In addition, we also conduct performance comparisons between representative approaches in each category, and introduce some typical applications of HS image SR.

REFERENCES

[1] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29–56, 2017.

[2] I. Mariotto, P. S. Thenkabail, A. Huete, E. T. Slonecker, and A. Platonov, "Hyperspectral versus multispectral crop-productivity modeling and type discrimination for the hyspiri mission," *Remote Sensing Environment*, vol. 139, pp. 291–305, 2013.

[3] M. Marshall and P. Thenkabail, "Advantage of hyperspectral eo-1 hyperion over multispectral ikonos, geoeye-1, worldview-2, landsat etm+, and modis vegetation indices in crop biomass estimation," *ISPRS J. Photogrammetry Remote Sensing*, vol. 108, pp. 205–218, 2015.

[4] Y. Li, W. Xie, and H. Li, "Hyperspectral image reconstruction by deep convolutional neural network for classification," *Pattern Recognition*, vol. 63, pp. 371–383, 2017.

[5] G. A. Carter, K. L. Lucas, G. A. Blossom, C. L. Lassitter, D. M. Holiday, D. S. Mooneyhan, D. R. Fastring, T. R. Holcombe, and J. A. Griffith, "Remote sensing and mapping of tamarisk along the colorado river, USA: A comparative use of summer-acquired hyperion, thematic mapper and quickbird data," *Remote Sensing*, vol. 1, no. 3, pp. 318–329, 2009.

[6] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geoscience Remote Sensing Magazine*, vol. 7, no. 2, pp. 140–158, 2019.

[7] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2015, pp. 3631–3640.

[8] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geoscience Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2014.

[9] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based bayesian fusion of multispectral and hyperspectral images," *IEEE Trans. Geoscience Remote Sensing*, vol. 47, no. 11, pp. 3834–3843, 2009.

[10] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2017, pp. 5344–5353.

[11] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Processing*, vol. 27, no. 8, pp. 4118–4130, 2018.

[12] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Processing*, vol. 28, no. 6, pp. 3034–3047, 2019.

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

[14] R. Dian, S. Li, L. Fang, T. Lu, and J. M. Bioucas-Dias, "Nonlocal sparse tensor factorization for semiblind hyperspectral and multispectral image fusion," *IEEE Trans. Cybernetics*, vol. 50, no. 10, pp. 4469–4480, 2020.

[15] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2011, pp. 2329–2336.

[16] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geoscience Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2012.

[17] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Computer Vision*, 2015, pp. 3586–3594.

[18] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Processing*, vol. 25, no. 5, pp. 2337–2352, 2016.

[19] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Processing*, vol. 28, no. 10, pp. 5135–5146, 2019.

[20] R. Loizzo, R. Guarini, F. Longo, T. Scopa, R. Formaro, C. Facchinetti, and G. Varacalli, "Prisma: The italian hyperspectral mission," in *Proc. IEEE Int. Geoscience Remote Sensing Symposium*, 2018, pp. 175–178.

[21] Y.-N. Liu, D.-X. Sun, X.-N. Hu, X. Ye, Y.-D. Li, S.-F. Liu, K.-Q. Cao, M.-Y. Chai, J. Zhang, Y. Zhang, *et al.*, "The advanced hyperspectral imager: Aboard china's gaofen-5 satellite," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 4, pp. 23–32, 2019.

[22] L. Loncan, L. B. De Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geoscience Remote Sensing Magazine*, vol. 3, no. 3, pp. 27–46, 2015.

[23] H. Zhang, H. Xu, X. Tian, J. Jiang, and J. Ma, "Image fusion meets deep learning: A survey and perspective," *Information Fusion*, vol. 76, pp. 323–336, 2021.

[24] R. Dian, S. Li, B. Sun, and A. Guo, "Recent advances and new guidelines on hyperspectral and multispectral image fusion," *Information Fusion*, vol. 69, pp. 40–51, 2021.

[25] D. Sara, A. K. Mandava, A. Kumar, S. Duela, and A. Jude, "Hyperspectral and multispectral image fusion techniques for high resolution applications: A review," *Earth Science Informatics*, vol. 14, no. 4, pp. 1685–1705, 2021.

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

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

[28] ——, "Deep unfolding network for spatiospectral image super-resolution," *IEEE Trans. Computational Imaging*, vol. 8, pp. 28–40, 2022.

[29] Y. Cai, J. Lin, X. Hu, H. Wang, X. Yuan, Y. Zhang, R. Timofte, and L. Van Gool, "Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2022, pp. 17502–17511.

[30] X. Hu, Y. Cai, J. Lin, H. Wang, X. Yuan, Y. Zhang, R. Timofte, and L. Van Gool, "Hdnet: High-resolution dual-domain learning for spectral compressive imaging," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2022, pp. 17542–17551.

[31] R. Dian, T. Shan, W. He, and H. Liu, "Spectral super-resolution via model-guided cross-fusion network," *IEEE Trans. Neural Networks and Learning Systems*, 2023.

[32] X. Fu, W. Wang, Y. Huang, X. Ding, and J. Paisley, "Deep multiscale detail networks for multiband spectral image sharpening," *IEEE Trans. Neural Networks Learning Systems*, vol. 32, no. 5, pp. 2090–2104, 2021.

[33] X. Meng, X. Meng, Q. Liu, J. Shu, F. Shao, G. Yang, and W. Sun, "Integrated fusion for panchromatic, multispectral, hyperspectral remote sensing images with different swath widths," *IEEE Geoscience Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[34] X. Tian, W. Zhang, Y. Chen, Z. Wang, and J. Ma, "Hyperfusion: A computational approach for hyperspectral, multispectral, and panchromatic image fusion," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–16, 2022.

[35] X. Tian, W. Zhang, D. Yu, and J. Ma, "Sparse tensor prior for hyperspectral, multispectral, and panchromatic image fusion,"

IEEE/CAA J. Autom. Sinica, vol. 10, no. 1, pp. 284–286, 2023.

[36] K. Li, W. Zhang, D. Yu, and X. Tian, "Hypernet: A deep network for hyperspectral, multispectral, and panchromatic image fusion," *ISPRS J. Photogrammetry Remote Sensing*, vol. 188, pp. 30–44, 2022.

[37] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 29–41, 2017.

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

[39] Z. He and L. Liu, "Hyperspectral image super-resolution inspired by deep laplacian pyramid network," *Remote Sensing*, vol. 10, no. 12, p. 1939, 2018.

[40] J. Yang, Y.-Q. Zhao, J. C.-W. Chan, and L. Xiao, "A multi-scale wavelet 3D-CNN for hyperspectral image super-resolution," *Remote Sensing*, vol. 11, no. 13, p. 1557, 2019.

[41] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Selected Topics Applied Earth Observations Remote Sensing*, vol. 10, no. 5, pp. 1963–1974, 2017.

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

[43] J. Hu, M. Zhao, and Y. Li, "Hyperspectral image super-resolution by deep spatial-spectral exploitation," *Remote Sensing*, vol. 11, no. 10, p. 1229, 2019.

[44] C. Zou and X. Huang, "Hyperspectral image super-resolution combining with deep learning and spectral unmixing," *Signal Processing: Image Communication*, vol. 84, p. 115833, 2020.

[45] X. Lu, D. Yang, J. Zhang, and F. Jia, "Hyperspectral image super-resolution based on spatial correlation-regularized unmixing convolutional neural network," *Remote Sensing*, vol. 13, no. 20, p. 4074, 2021.

[46] Y. Li, L. Zhang, C. Dingl, W. Wei, and Y. Zhang, "Single hyperspectral image super-resolution with grouped deep recursive residual network," in *Proc. IEEE Int. Conf. Multimedia Big Data*, 2018, pp. 1–4.

[47] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *IEEE Trans. Computational Imaging*, vol. 6, pp. 1082–1096, 2020.

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

[49] Y. Li, Z. Du, S. Wu, Y. Wang, Z. Wang, X. Zhao, and F. Zhang, "Progressive split-merge super resolution for hyperspectral imagery with group attention and gradient guidance," *ISPRS J. Photogrammetry Remote Sensing*, vol. 182, pp. 14–36, 2021.

[50] C. Zhang, M. Zhang, Y. Li, X. Gao, and S. Qiu, "Difference curvature multidimensional network for hyperspectral image super-resolution," *Remote Sensing*, vol. 13, no. 17, p. 3455, 2021.

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

[52] D. Liu, J. Li, and Q. Yuan, "A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 59, no. 9, pp. 7711–7725, 2021.

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

[54] Q. Wang, Q. Li, and X. Li, "Hyperspectral image superresolution using spectrum and feature context," *IEEE Trans. Industrial Electronics*, vol. 68, no. 11, pp. 11 276–11 285, 2021.

[55] Z. Tang, Q. Xu, P. Wu, Z. Shi, and B. Pan, "Feedback refined local-global network for super-resolution of hyperspectral imagery," *Remote Sensing*, vol. 14, no. 8, p. 1944, 2022.

[56] X. Wang, Q. Hu, J. Jiang, and J. Ma, "A group-based embedding

learning and integration network for hyperspectral image super-resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–16, 2022.

[57] X. Wang, Y. Cheng, X. Mei, J. Jiang, and J. Ma, "Group shuffle and spectral-spatial fusion for hyperspectral image super-resolution," *IEEE Trans. Computational Imaging*, vol. 8, pp. 1223–1236, 2022.

[58] S. Mei, X. Yuan, J. Ji, S. Wan, J. Hou, and Q. Du, "Hyperspectral image super-resolution via convolutional neural network," in *Proc. IEEE Int. Conf. Image Processing*, 2017, pp. 4297–4301.

[59] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3D full convolutional neural network," *Remote Sensing*, vol. 9, no. 11, p. 1139, 2017.

[60] J. Li, R. Cui, B. Li, Y. Li, S. Mei, and Q. Du, "Dual 1d-2d spatial-spectral cnn for hyperspectral image super-resolution," in *Proc. IEEE Int. Geoscience Remote Sensing Symposium*, 2019, pp. 3113–3116.

[61] J. Li, R. Cui, B. Li, R. Song, Y. Li, and Q. Du, "Hyperspectral image super-resolution with 1D–2D attentional convolutional neural network," *Remote Sensing*, vol. 11, no. 23, p. 2859, 2019.

[62] Q. Wang, Q. Li, and X. Li, "Spatial-spectral residual network for hyperspectral image super-resolution," arXiv preprint arXiv: 2001. 04609, 2020.

[63] Q. Li, Q. Wang, and X. Li, "Mixed 2D/3D convolutional network for hyperspectral image super-resolution," *Remote Sensing*, vol. 12, no. 10, p. 1660, 2020.

[64] Q. Li, Q. Wang, and X. Li, "Exploring the relationship between 2D/3D convolution for hyperspectral image super-resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 59, no. 10, pp. 8693–8703, 2021.

[65] Q. Li, Y. Yuan, and Q. Wang, "Hyperspectral image super-resolution via multi-domain feature learning," *Neurocomputing*, vol. 472, pp. 85–94, 2022.

[66] Y. Fu, Z. Liang, and S. You, "Bidirectional 3D quasi-recurrent neural network for hyperspectral image super-resolution," *IEEE J. Selected Topics Applied Earth Observations Remote Sensing*, vol. 14, pp. 2674–2688, 2021.

[67] J. Yang, L. Xiao, Y.-Q. Zhao, and J. C.-W. Chan, "Hybrid local and nonlocal 3-D attentive CNN for hyperspectral image super-resolution," *IEEE Geoscience Remote Sensing Letters*, vol. 18, no. 7, pp. 1274–1278, 2021.

[68] Q. Huang, W. Li, T. Hu, and R. Tao, "Hyperspectral image super-resolution using generative adversarial network and residual learning," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Processing*, 2019, pp. 3012–3016.

[69] X. Dou, C. Li, Q. Shi, and M. Liu, "Super-resolution for hyperspectral remote sensing images based on the 3D attention−SRGAN network," *Remote Sensing*, vol. 12, no. 7, p. 1204, 2020.

[70] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2017, pp. 4681–4690.

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

[72] R. Jiang, X. Li, A. Gao, L. Li, H. Meng, S. Yue, and L. Zhang, "Learning spectral and spatial features based on generative adversarial network for hyperspectral image super-resolution," in *Proc. IEEE Int. Geoscience Remote Sensing Symposium*, 2019, pp. 3161–3164.

[73] B. Wang, S. Zhang, Y. Feng, S. Mei, S. Jia, and Q. Du, "Hyperspectral imagery spatial super-resolution using generative adversarial network," *IEEE Trans. Computational Imaging*, vol. 7, pp. 948–960, 2021.

[74] Y. Shi, L.-Q. Han, L. Han, S. Chang, T. Hu, and D. Dancey, "A latent encoder coupled generative adversarial network (LE-GAN) for efficient hyperspectral image super-resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–19, 2022.

[75] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2018, pp. 9446–9454.

[76] O. Sidorov and J. Yngve Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE Int. Conf. Computer Vision Workshops*, 2019.

[77] Z. Gong, N. Wang, D. Cheng, X. Jiang, J. Xin, X. Yang, and X. Gao, "Learning deep resonant prior for hyperspectral image super-resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–14, 2022.

[78] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2018, pp. 2472–2481.

[79] M. Zhao, J. Ning, J. Hu, and T. Li, "Hyperspectral image super-resolution under the guidance of deep gradient information," *Remote Sensing*, vol. 13, no. 12, p. 2382, 2021.

[80] B. Wang, S. Mei, Y. Feng, and Q. Du, "Hyperspectral imagery super-resolution based on self-calibrated attention residual network," in *Proc. IEEE Int. Geoscience Remote Sensing Symposium*, 2021, pp. 3896–3899.

[81] X. Wang, J. Ma, J. Jiang, and X.-P. Zhang, "Dilated projection correction network based on autoencoder for hyperspectral image super-resolution," *Neural Networks*, vol. 146, pp. 107–119, 2022.

[82] J. Zhang, M. Shao, Z. Wan, and Y. Li, "Multi-scale feature mapping network for hyperspectral image super-resolution," *Remote Sensing*, vol. 13, no. 20, p. 4180, 2021.

[83] J. Hou, Z. Zhu, J. Hou, H. Zeng, J. Wu, and J. Zhou, "Deep posterior distribution-based embedding for hyperspectral image super-resolution," *IEEE Trans. Image Processing*, vol. 31, pp. 5720–5732, 2022.

[84] D. Liu, J. Li, Q. Yuan, L. Zheng, J. He, S. Zhao, and Y. Xiao, "An efficient unfolding network with disentangled spatial-spectral representation for hyperspectral image super-resolution," *Information Fusion*, vol. 94, pp. 92–111, 2023.

[85] Y. Liu, J. Hu, X. Kang, J. Luo, and S. Fan, "Interactformer: Interactive transformer and cnn for hyperspectral image super-resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–15, 2022.

[86] C. Deng, X. Luo, and W. Wang, "Multiple frame splicing and degradation learning for hyperspectral imagery super-resolution," *IEEE J. Selected Topics Applied Earth Observations Remote Sensing*, vol. 15, pp. 8389–8401, 2022.

[87] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of ms + pan data," *IEEE Trans. Geoscience Remote Sensing*, vol. 45, no. 10, pp. 3230–3239, 2007.

[88] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geoscience Remote Sensing*, vol. 46, no. 5, pp. 1301–1312, 2008.

[89] K. Li, W. Xie, Q. Du, and Y. Li, "Ddlps: Detail-based deep Laplacian pansharpening for hyperspectral imagery," *IEEE Trans. Geoscience Remote Sensing*, vol. 57, no. 10, pp. 8011–8025, 2019.

[90] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2017, pp. 624–632.

[91] W. Dong, T. Zhang, J. Qu, S. Xiao, J. Liang, and Y. Li, "Laplacian pyramid dense network for hyperspectral pansharpening," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–13, 2022.

[92] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Networks Learning Systems*, vol. 31, no. 5, pp. 1529–1543, 2020.

[93] G. He, J. Zhong, J. Lei, Y. Li, and W. Xie, "Hyperspectral pansharpening based on spectral constrained adversarial autoencoder," *Remote Sensing*, vol. 11, no. 22, p. 2691, 2019.

[94] W. Dong, Y. Yang, J. Qu, W. Xie, and Y. Li, "Fusion of hyperspectral and panchromatic images using generative adversarial network and image segmentation," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–13, 2022.

[95] Y. Zheng, J. Li, Y. Li, K. Cao, and K. Wang, "Deep residual learning for boosting the accuracy of hyperspectral pansharpening," *IEEE Geoscience Remote Sensing Letters*, vol. 17, no. 8, pp. 1435–1439, 2020.

[96] W. Dong, C. Zhou, F. Wu, J. Wu, G. Shi, and X. Li, "Model-guided deep hyperspectral image super-resolution," *IEEE Trans. Image Processing*, vol. 30, pp. 5754–5768, 2021.

[97] X. Lu, J. Zhang, D. Yang, L. Xu, and F. Jia, "Cascaded convolutional neural network-based hyperspectral image resolution enhancement via an auxiliary panchromatic image," *IEEE Trans. Image Processing*, vol. 30, pp. 6815–6828, 2021.

[98] Y. Zheng, J. Li, Y. Li, Y. Shi, and J. Qu, "Deep residual spatial attention network for hyperspectral pansharpening," in *Proc. IEEE Int. Geoscience Remote Sensing Symposium*, 2020, pp. 2671–2674.

[99] X. Wu, J. Feng, R. Shang, X. Zhang, and L. Jiao, "Multi-objective guided divide-and-conquer network for hyperspectral pansharpening," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–17, 2022.

[100] 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. Geoscience Remote Sensing*, vol. 58, no. 11, pp. 8059–8076, 2020.

[101] 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. Geoscience Remote Sensing*, vol. 60, pp. 1–16, 2022.

[102] Y.-W. Zhuo, T.-J. Zhang, J.-F. Hu, H.-X. Dou, T.-Z. Huang, and L.-J. Deng, "A deep-shallow fusion network with multi-detail extractor and spectral attention for hyperspectral pansharpening," *IEEE J. Selected Topics Applied Earth Observations Remote Sensing*, vol. 15, pp. 7539–7555, 2022.

[103] L. He, J. Zhu, J. Li, A. Plaza, J. Chanussot, and B. Li, "Hyperpnn: Hyperspectral pansharpening via spectrally predictive convolutional neural networks," *IEEE J. Selected Topics Applied Earth Observations Remote Sensing*, vol. 12, no. 8, pp. 3092–3100, 2019.

[104] W. Xie, Y. Cui, Y. Li, J. Lei, Q. Du, and J. Li, "HPGAN: Hyperspectral pansharpening using 3-D generative adversarial networks," *IEEE Trans. Geoscience Remote Sensing*, vol. 59, no. 1, pp. 463–477, 2021.

[105] S. Hou, S. Xiao, W. Dong, and J. Qu, "Multi-level features fusion via cross-layer guided attention for hyperspectral pansharpening," *Neurocomputing*, vol. 506, pp. 380–392, 2022.

[106] P. Guan and E. Y. Lam, "Multistage dual-attention guided fusion network for hyperspectral pansharpening," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–14, 2022.

[107] J. Qu, Z. Xu, W. Dong, S. Xiao, Y. Li, and Q. Du, "A spatio-spectral fusion method for hyperspectral images using residual hyper-dense network," *IEEE Trans. Neural Networks Learning Systems*, pp. 1–15, 2022.

[108] Y. Yang, W. Dong, S. Xiao, T. Zhang, and J. Qu, "Multi-level spatial details cross-extraction and injection network for hyperspectral pansharpening," *Optics Letters*, vol. 47, no. 6, pp. 1371–1374, 2022.

[109] Y. Zheng, J. Li, Y. Li, K. Cao, and K. Wang, "Pansharpening of hyperspectral images with detail guided feature modulation," in *Proc. IEEE Int. Geoscience Remote Sensing Symposium*, 2021, pp. 4456–4459.

[110] Y. Li, Y. Zheng, J. Li, R. Song, and J. Chanussot, "Hyperspectral pansharpening with adaptive feature modulation-based detail injection network," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–17, 2022.

[111] Y. Shang, J. Liu, J. Yang, and Z. Wu, "A model-inspired approach with transformers for hyperspectral pansharpening," *IEEE J. Selected Topics Applied Earth Observations Remote Sensing*, vol. 15, pp. 7187–7202, 2022.

[112] W. G. C. Bandara and V. M. Patel, "Hypertransformer: A textural and spectral feature fusion transformer for pansharpening," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2022, pp. 1767–1777.

[113] J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, and Y. Ma, "Swinfusion: Cross-domain long-range learning for general image fusion via swin transformer," *IEEE/CAA J. Automa. Sinica*, vol. 9, no. 7, pp. 1200–1217, 2022.

[114] J. Qu, Y. Shi, W. Xie, Y. Li, X. Wu, and Q. Du, "MSSL: Hyperspectral and panchromatic images fusion via multiresolution spatial – spectral feature learning networks," *IEEE Trans. Geoscience*

Remote Sensing, vol. 60, pp. 1–13, 2022.

[115] R. Zhao and S. Du, "An encoder–decoder with a residual network for fusing hyperspectral and panchromatic remote sensing images," *Remote Sensing*, vol. 14, no. 9, p. 1981, 2022.

[116] ——, "Spectral-spatial residual network for fusing hyperspectral and panchromatic remote sensing images," *Remote Sensing*, vol. 14, no. 3, p. 800, 2022.

[117] J. Qu, S. Hou, W. Dong, S. Xiao, Q. Du, and Y. Li, "A dual – Branch detail extraction network for hyperspectral pansharpening," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–13, 2022.

[118] W. Dong, J. Qu, T. Zhang, Y. Li, and Q. Du, "Context-aware guided attention based cross-feedback dense network for hyperspectral image super-resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–14, 2022.

[119] L. He, J. Zhu, J. Li, A. Plaza, J. Chanussot, and Z. Yu, "CNN-based hyperspectral pansharpening with arbitrary resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–21, 2022.

[120] L. He, J. Xie, J. Li, A. Plaza, J. Chanussot, and J. Zhu, "Variable subpixel convolution based arbitrary-resolution hyperspectral pansharpening," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–19, 2022.

[121] W. Dong, S. Hou, S. Xiao, J. Qu, Q. Du, and Y. Li, "Generative dual-adversarial network with spectral fidelity and spatial enhancement for hyperspectral pansharpening," *IEEE Trans. Neural Networks Learning Systems*, vol. 33, no. 12, pp. 7303–7317, 2022.

[122] J. Nie, Q. Xu, and J. Pan, "Unsupervised hyperspectral pansharpening by ratio estimation and residual attention network," *IEEE Geoscience Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[123] L. He, J. Zhu, J. Li, D. Meng, J. Chanussot, and A. Plaza, "Spectral-fidelity convolutional neural networks for hyperspectral pansharpening," *IEEE J. Selected Topics Applied Earth Observations Remote Sensing*, vol. 13, pp. 5898–5914, 2020.

[124] W. Wei, J. Nie, Y. Li, L. Zhang, and Y. Zhang, "Deep recursive network for hyperspectral image super-resolution," *IEEE Trans. Computational Imaging*, vol. 6, pp. 1233–1244, 2020.

[125] R. Dian, S. Li, and X. Kang, "Regularizing hyperspectral and multispectral image fusion by cnn denoiser," *IEEE Trans. Neural Networks Learning Systems*, vol. 32, no. 3, pp. 1124–1135, 2021.

[126] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by MS/HS fusion net," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2019, pp. 1585–1594.

[127] W. Wang, W. Zeng, Y. Huang, X. Ding, and J. Paisley, "Deep blind hyperspectral image fusion," in *Proc. IEEE Int. Conf. Computer Vision*, 2019, pp. 4150–4159.

[128] W. Wang, X. Fu, W. Zeng, L. Sun, R. Zhan, Y. Huang, and X. Ding, "Enhanced deep blind hyperspectral image fusion," *IEEE Trans. Neural Networks Learning Systems*, pp. 1–11, 2021.

[129] Z. Liu, Y. Zheng, and X.-H. Han, "Deep unsupervised fusion learning for hyperspectral image super resolution," *Sensors*, vol. 21, no. 7, p. 2348, 2021.

[130] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Networks Learning Systems*, vol. 29, no. 11, pp. 5345–5355, 2018.

[131] X. Wang, J. Chen, Q. Wei, and C. Richard, "Hyperspectral image super-resolution via deep prior regularization with parameter estimation," *IEEE Trans. Circuits Systems Video Technology*, vol. 32, no. 4, pp. 1708–1723, 2022.

[132] M. Vella, B. Zhang, W. Chen, and J. F. Mota, "Enhanced hyperspectral image super-resolution via rgb fusion and tv-tv minimization," in *Proc. IEEE Int. Conf. Image Processing*, 2021, pp. 3837–3841.

[133] X. Wang, J. Chen, and C. Richard, "Hyperspectral image super-resolution with deep priors and degradation model inversion," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Processing*, 2022, pp. 2814–2818.

[134] J. Liu, D. Shen, Z. Wu, L. Xiao, J. Sun, and H. Yan, "Patch-aware deep hyperspectral and multispectral image fusion by unfolding subspace-based optimization model," *IEEE J. Selected Topics Applied Earth Observations Remote Sensing*, vol. 15, pp. 1024–1038, 2022.

[135] J. Yang, L. Xiao, Y.-Q. Zhao, and J. C.-W. Chan, "Variational regularization network with attentive deep prior for hyperspectral-ltispectral image fusion," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–17, 2022.

[136] W. Wan, B. Zhang, M. Vella, J. F. Mota, and W. Chen, "Robust rgbguided super-resolution of hyperspectral images via tv^3 minimization," *IEEE Signal Processing Letters*, vol. 29, pp. 957–961, 2022.

[137] T. Huang, W. Dong, J. Wu, L. Li, X. Li, and G. Shi, "Deep hyperspectral image fusion network with iterative spatio-spectral regularization," *IEEE Trans. Computational Imaging*, vol. 8, pp. 201–214, 2022.

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

[139] J. Gao, J. Li, and M. Jiang, "Hyperspectral and multispectral image fusion by deep neural network in a self-supervised manner," *Remote Sensing*, vol. 13, no. 16, p. 3226, 2021.

[140] Z. Liu, Y. Zheng, and X.-H. Han, "Unsupervised multispectral and hyperspectral image fusion with deep spatial and spectral priors," in *Proc. Asian Conf. Computer Vision*, 2020.

[141] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised adaptation learning for hyperspectral imagery super-resolution," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2020, pp. 3073–3082.

[142] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang, "Hyperspectral image super-resolution with optimized RGB guidance," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2019, pp. 11661–11670.

[143] W. Wei, J. Nie, L. Zhang, and Y. Zhang, "Unsupervised recurrent hyperspectral imagery super-resolution using pixel-aware refinement," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–15, 2022.

[144] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse dirichlet-net for hyperspectral image super-resolution," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2018, pp. 2511–2520.

[145] Z. Wang, B. Chen, R. Lu, H. Zhang, H. Liu, and P. K. Varshney, "FusionNet: An unsupervised convolutional variational network for hyperspectral and multispectral image fusion," *IEEE Trans. Image Processing*, vol. 29, pp. 7565–7577, 2020.

[146] Y. Qu, H. Qi, C. Kwan, N. Yokoya, and J. Chanussot, "Unsupervised and unregistered hyperspectral image super-resolution with mutual dirichlet-Net," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–18, 2022.

[147] K. Zheng, L. Gao, D. Hong, B. Zhang, and J. Chanussot, "Nonregsrnet: A nonrigid registration hyperspectral super-resolution network," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–16, 2022.

[148] J. Zhao, Y. Qu, S. Ninomiya, and W. Guo, "Endmember-assisted camera response function learning, toward improving hyperspectral image super-resolution performance," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–14, 2022.

[149] J. Liu, Z. Wu, L. Xiao, and X.-J. Wu, "Model inspired autoencoder for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–12, 2022.

[150] Z. Wang, B. Chen, H. Zhang, and H. Liu, "Unsupervised hyperspectral and multispectral images fusion based on nonlinear variational probabilistic generative model," *IEEE Trans. Neural Networks Learning Systems*, vol. 33, no. 2, pp. 721–735, 2022.

[151] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geoscience Remote Sensing Letters*, vol. 14, no. 5, pp. 639–643, 2017.

[152] Z. Shi, C. Chen, Z. Xiong, D. Liu, Z.-J. Zha, and F. Wu, "Deep residual attention network for spectral image super-resolution," in *Proc. European Conf. Computer Vision Workshops*, 2018.

[153] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Pyramid fully convolutional network for hyperspectral and multispectral image fusion," *IEEE J.*

Selected Topics Applied Earth Observations Remote Sensing, vol. 12, no. 5, pp. 1549–1558, 2019.

[154] S. Huang and D. W. Messinger, "An unsupervised laplacian pyramid network for radiometrically accurate data fusion of hyperspectral and multispectral imagery," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–17, 2022.

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

[156] Z. Zhu, J. Hou, J. Chen, H. Zeng, and J. Zhou, "Hyperspectral image super-resolution via deep progressive zero-centric residual learning," *IEEE Trans. Image Processing*, vol. 30, pp. 1423–1438, 2021.

[157] X.-H. Han, B. Shi, and Y. Zheng, "SSF-CNN: Spatial and spectral fusion with CNN for hyperspectral image super-resolution," in *Proc. IEEE Int. Conf. Image Processing*, 2018, pp. 2506–2510.

[158] J.-F. Hu, T.-Z. Huang, L.-J. Deng, T.-X. Jiang, G. Vivone, and J. Chanussot, "Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks," *IEEE Trans. Neural Networks Learning Systems*, vol. 33, no. 12, pp. 7251–7265, 2022.

[159] B. Pan, Q. Qu, X. Xu, and Z. Shi, "Structure – Color preserving network for hyperspectral image super-resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–12, 2022.

[160] J. Xiao, J. Li, Q. Yuan, and L. Zhang, "A dual-UNET with multistage details injection for hyperspectral image fusion," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–13, 2022.

[161] T. Zhan, Z. Bi, H. Wu, C. Xu, Q. Du, Y. Xu, and Z. Wu, "A novel cross-scale octave network for hyperspectral and multispectral image fusion," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–16, 2022.

[162] Z. Guo, J. Xin, N. Wang, J. Li, and X. Gao, "External-internal attention for hyperspectral image super-resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–14, 2022.

[163] 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. Geoscience Remote Sensing*, vol. 60, pp. 1–16, 2022.

[164] X.-H. Han, Y. Zheng, and Y.-W. Chen, "Multi-level and multi-scale spatial and spectral fusion cnn for hyperspectral image super-resolution," in *Proc. IEEE Int. Conf. Computer Vision Workshops*, 2019.

[165] T. Uezato, D. Hong, N. Yokoya, and W. He, "Guided deep decoder: Unsupervised image pair fusion," in *Proc. European Conf. Computer Vision*, 2020, pp. 87–102.

[166] X. Wang, X. Wang, K. Zhao, X. Zhao, and C. Song, "Fsl-UNet: Full-scale linked UNet with spatial-spectral joint perceptual attention for hyperspectral and multispectral image fusion," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–14, 2022.

[167] S. Xu, O. Amira, J. Liu, C.-X. Zhang, J. Zhang, and G. Li, "HAMMFN: Hyperspectral and multispectral image multiscale fusion network with RAP loss," *IEEE Trans. Geoscience Remote Sensing*, vol. 58, no. 7, pp. 4618–4628, 2020.

[168] 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. Cybernetics*, pp. 1–14, 2023.

[169] V. K. Ha, J. Ren, Z. Wang, G. Sun, H. Zhao, and S. Marshall, "Multiscale spatial fusion and regularization induced unsupervised auxiliary task CNN model for deep super-resolution of hyperspectral images," *IEEE J. Selected Topics Applied Earth Observations Remote Sensing*, vol. 15, pp. 4583–4598, 2022.

[170] W. Wang, W. Zeng, Y. Huang, and X. Ding, "Deep multiscale feedback network for hyperspectral image fusion," *IEEE Geoscience Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[171] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sensing*, vol. 10, no. 5, p. 800, 2018.

[172] Q. Li, M. Gong, Y. Yuan, and Q. Wang, "Symmetrical feature propagation network for hyperspectral image super-resolution," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–12, 2022.

[173] X. Zhang, W. Huang, Q. Wang, and X. Li, "SSR-Net: Spatial – Spectral reconstruction network for hyperspectral and multispectral image fusion," *IEEE Trans. Geoscience Remote Sensing*, vol. 59, no. 7, pp. 5953–5965, 2021.

[174] Y. Zheng, J. Li, Y. Li, J. Guo, X. Wu, Y. Shi, and J. Chanussot, "Edge-conditioned feature transform network for hyperspectral and multispectral image fusion," *IEEE Trans. Geoscience Remote Sensing*, vol. 60, pp. 1–15, 2022.

[175] W. Chen, X. Zheng, and X. Lu, "Hyperspectral image super-resolution with self-supervised spectral-spatial residual network," *Remote Sensing*, vol. 13, no. 7, p. 1260, 2021.

[176] H. Su, H. Jin, and C. Sun, "Deep pansharpening via 3d spectral super-resolution network and discrepancy-based gradient transfer," *Remote Sensing*, vol. 14, no. 17, p. 4250, 2022.

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

[178] S. Liu, S. Miao, J. Su, B. Li, W. Hu, and Y.-D. Zhang, "Umag-net: A new unsupervised multiattention-guided network for hyperspectral and multispectral image fusion," *IEEE J. Selected Topics Applied Earth Observations Remote Sensing*, vol. 14, pp. 7373–7385, 2021.

[179] J. Xiao, J. Li, Q. Yuan, M. Jiang, and L. Zhang, "Physics-based GAN with iterative refinement unit for hyperspectral and multispectral image fusion," *IEEE J. Selected Topics Applied Earth Observations Remote Sensing*, vol. 14, pp. 6827–6841, 2021.

[180] Z. Zhang, Z. Xu, Z. Ahmed, A. Salekin, and T. Rahman, "Hyperspectral image super-resolution in arbitrary input-output band settings," in *Proc. IEEE Winter Conf. Applications Computer Vision*, 2022, pp. 749–759.

[181] J.-F. Hu, T.-Z. Huang, L.-J. Deng, H.-X. Dou, D. Hong, and G. Vivone, "Fusformer: A transformer-based fusion network for hyperspectral image super-resolution," *IEEE Geoscience Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[182] T. You, C. Wu, Y. Bai, D. Wang, H. Ge, and Y. Li, "HMF-former: Spatio-spectral transformer for hyperspectral and multispectral image fusion," *IEEE Geoscience Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

[183] 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. Geoscience Remote Sensing*, vol. 61, pp. 1–15, 2023.

[184] 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. Geoscience Remote Sensing*, vol. 60, pp. 1–13, 2022.

[185] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. European Conf. Computer Vision*, 2016, pp. 19–34.

[186] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Processing*, vol. 19, no. 9, pp. 2241–2253, 2010.

[187] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over chikusei," Space Application Laboratory, University of Tokyo, Japan, Tech. Rep. SAL-2016-05-27, May 2016.

[188] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2011, pp. 193–200.

[189] F. A. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, "The spectral image processing system (SIPS) interactive visualization and analysis of imaging spectrometer data," *Remote Sensing Environment*, vol. 44, no. 2–3, pp. 145–163, 1993.

[190] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[191] Y. Zeng, W. Huang, M. Liu, H. Zhang, and B. Zou, "Fusion of satellite images in URBAN area: Assessing the quality of resulting images," in *Proc. Int. Conf. Geoinformatics*, 2010, pp. 1–4.

[192] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, "WHU-HI: Uav-borne hyperspectral with high spatial resolution (h2) benchmark

datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," *Remote Sensing Environment*, vol. 250, p. 112012, 2020.

[193] J. Behmann, J. Steinrücken, and L. Plümer, "Detection of early plant stress responses in hyperspectral images," *ISPRS J. Photogrammetry Remote Sensing*, vol. 93, pp. 98–111, 2014.

[194] C. Gomez, P. Lagacherie, and G. Coulouma, "Regional predictions of eight common soil properties and their spatial structures from hyperspectral vis–NIR data," *Geoderma*, vol. 189, pp. 176–185, 2012.

[195] F. A. Kruse, J. W. Boardman, and J. F. Huntington, "Comparison of airborne hyperspectral data and eo-1 hyperion for mineral mapping," *IEEE Trans. Geoscience Remote Sensing*, vol. 41, no. 6, pp. 1388–1400, 2003.

[196] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam, "Deep learning-based change detection in remote sensing images: A review," *Remote Sensing*, vol. 14, no. 4, p. 871, 2022.

[197] Z. Shao, H. Fu, D. Li, O. Altan, and T. Cheng, "Remote sensing monitoring of multi-scale watersheds impermeability for Urban hydrological evaluation," *Remote Sensing Environment*, vol. 232, p. 111338, 2019.

[198] N. Rudorff, C. M. Rudorff, M. Kampel, and G. Ortiz, "Remote sensing monitoring of the impact of a major mining wastewater disaster on the turbidity of the doce river plume off the eastern brazilian coast," *ISPRS J. Photogrammetry Remote Sensing*, vol. 145, pp. 349–361, 2018.

[199] C. Pohl and J. van Genderen, "Remote sensing image fusion: An update in the context of digital earth," *Int. J. Digital Earth*, vol. 7, no. 2, pp. 158–172, 2014.

[200] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *J. Biomedical Optics*, vol. 19, no. 1, pp. 010901–010901, 2014.

[201] L. Wei, L. Meng, C. Tianhong, C. Zhaoyao, and T. Ran, "Application of a hyperspectral image in medical field: A review," *Image Graph*, vol. 26, pp. 1764–1785, 2021.

[202] R. Cui, H. Yu, T. Xu, X. Xing, X. Cao, K. Yan, and J. Chen, "Deep learning in medical hyperspectral images: A review," *Sensors*, vol. 22, no. 24, p. 9790, 2022.

[203] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Computer Vision*, vol. 129, pp. 23–79, 2021.

[204] X. Jiang, J. Ma, G. Xiao, Z. Shao, and X. Guo, "A review of multimodal image matching: Methods and applications," *Information Fusion*, vol. 73, pp. 22–71, 2021.

[205] L. Tang, Y. Deng, Y. Ma, J. Huang, and J. Ma, "Superfusion: A versatile image registration and fusion network with semantic awareness," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 12, pp. 2121–2137, 2022.

Xinya Wang received the B.S. degree from the Electronic Information School, Wuhan University, in 2018. She is currently a Ph.D. candidate at the Multi-Spectral Vision Processing Laboratory, Wuhan University. Her current research interests include neural networks, machine learning and image processing.

Qian Hu received the B.S. degree from the Electronic Information School, Wuhan University, in 2022. He is currently a master student with the Electronic Information School, Wuhan University. His current research interests include computer vision and image processing.

Yingsong Cheng received the B.S. degree from the Computer Science and Technology School, Huazhong University of Science and Technology, in 2022. He is currently a master student with the Electronic Information School, Wuhan University. His current research interests include computer vision and image processing.
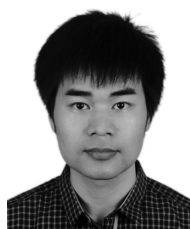
Jiayi Ma (Member, IEEE) received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, in 2008 and 2014, respectively. He is currently a Professor with the Electronic Information School, Wuhan University. He has authored or co-authored more than 300 refereed journal and conference papers, including IEEE TPAMI/TIP, IJCV, CVPR, ICCV, ECCV, etc. He has been identified in the 2019–2022 Highly Cited Researcher lists from the Web of Science Group. He is an Area Editor of *Information Fusion*, an Associate Editor of *IEEE/CAA Journal of Automatica Sinica*, *Neurocomputing*, *Geo-Spatial Information Science*, and *Frontiers in Neuroscience*.

His research interests include computer vision, machine learning, and pattern recognition.