

Tree-Structured Neural Network for Hyperspectral Pansharpener

Lin He , *Member, IEEE*, Hanghui Ye , Dahan Xi , Jun Li , *Fellow, IEEE*, Antonio Plaza , *Fellow, IEEE*, and Mei Zhang 

Abstract—Hyperspectral (HS) pansharpener refers to fusing low spatial resolution HS (LRHS) images with the corresponding panchromatic (PAN) images to create high spatial resolution HS (HRHS) images. Most of the existing HS pansharpener methods overlook the spatial and spectral imbalance of the ground objects of different types in the observed scenes. To address the dilemma, in this article we develop a novel tree-structured neural network (Tree-SNet) to form an adaptive spatial-spectral processing for HS pansharpener. The Tree-SNet method maps a convolutional neural network (CNN) onto a hierarchical tree structure, where routing nodes automatically tune the data distributed to tree paths, which is adaptive to the local characteristics of the data, while spatial enhancement (SpatE) and spectral enhancement (SpecE) modules are dynamically performed in the tree paths to further strengthen the adaptive processing. The proposed Tree-SNet is evaluated on several datasets, and the experimental results verify its superiority.

Index Terms—Adaptive data distribution, convolutional neural network (CNN), dynamic enhancement, hierarchical tree, hyperspectral (HS) images, pansharpener.

I. INTRODUCTION

SATELLITE remote sensing seeks to provide accurate observations of the Earth's surface with sufficient information. However, owing to the inherent physical constraints of satellite sensors, the spatial resolutions and the spectral resolutions of the connected data have an effect on each other. For example, a panchromatic (PAN) image is usually composed of only one band with high spatial resolutions, while a hyperspectral (HS) image comprises hundreds of bands but often with a low spatial resolution. Considering the fact that the HS data with full resolution are required for many practical applications [1], [2],

[3], [4], [5], [6], [7], we have to resort to the HS pansharpener technique, which has drawn much attention in recent years. It is able to reconstruct high-resolution HS (HRHS) images by fusing the matching pairs of low-resolution HS (LRHS) images and PAN images.

The HS pansharpener technique is originally extended from traditional multispectral (MS) pansharpener methods [2], [2], [3], [8], [10], [11], [12], some of which include Gram-Schmidt (GS) [8], intensity-hue-saturation (IHS) [13], principal component analysis (PCA) [14], smoothing filter-based intensity modulation (SFIM) [15], wavelet transform [16], modulation transfer function generalized Laplacian pyramid (MTF-GLP) [17] and its high-pass variant [18], can be approximately categorized into component substitution (CS) and multiresolution analysis (MRA). Some of the other pansharpener approaches can be attributed to the variational optimization (VO)-based processing, which regards pansharpener as an ill-posed inverse problem by variation [9][19], [20], [21]. Representative VO-based methods include matrix factorization [2], [22] and Bayesian estimation [2], [23], [24]. Matrix factorization methods usually decompose low-resolution MS/HS images and high-resolution PANs to obtain basis matrices and coefficient matrices, respectively, and then form high-resolution MS/HS images by the multiplication of the basis matrices and the coefficient matrices. In contrast, Bayesian estimation methods treat MS/HS pansharpener as an optimization problem in the Bayesian framework, which is often regularized by a prior. Since convolutional neural networks (CNNs) posed significant potential for enhancing spatial resolutions of data, CNN-based methods were proposed for the pansharpener task. CNN-based pansharpener is pioneered by pansharpener neural network (PNN) [25], which aims at MS pansharpener and is then followed by some researchers [26], [27], [28], [29]. An HS image is usually characterized by obviously higher spectral resolution and contains much more spectral bands than an MS image. Therefore, CNNs need to be designed more elaborately to deal with HS pansharpener. Zheng et al. [32] designed a CNN structure for HS pansharpener that can make full use of the relationship between spatial information and spectral information through a residual block. Xie et al. [33] discovered deep priors to predict the HRHS images, which are used as a minimum constraint during the HRHS image restoration. He et al. [34], [35] built a spectral prediction structure to enhance the spectral fidelity [34] and proposed an HS pansharpener framework that can produce reliable spatial detail through a physically interpretable skip connection.

Manuscript received 30 July 2023; revised 16 October 2023 and 26 November 2023; accepted 7 December 2023. Date of publication 18 December 2023; date of current version 8 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62071184, Grant 61571195, and Grant 42030111, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515011615 and Grant 2023A1515011887, and in part by the Guangzhou Science and Technology Program under Grant 202002030395. (Corresponding authors: Lin He; Jun Li.)

Lin He, Hanghui Ye, Dahan Xi, and Mei Zhang are with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: helin@scut.edu.cn; 872726477@qq.com; 2574242594@qq.com; zhangmei@scut.edu.cn).

Jun Li is with the School of Computer Science, China University of Geosciences, Wuhan 430078, China, and also with the Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430078, China (e-mail: lijuncug@cug.edu.cn).

Antonio Plaza is with the Hyperspectral Computing Laboratory, University of Extremadura, 10003 Caceres, Spain (e-mail: aplaza@unex.es).

Digital Object Identifier 10.1109/JSTARS.2023.3344117

Existing HS pansharpening CNNs have achieved great progress [25], [26], [30], [31], [32], [33], [34], [35], [36], [42], which usually accomplish better quality indicators in pansharpening processing than traditional methods. However, the spatial and the spectral attributes of an HS image and their coupling may be spatially different in the real-world situation, while most of the HS pansharpening methods did not explicitly consider such a spatial varying characteristic of HS data, which may cause distortion to the pansharpened HRHS image, especially when large-scale networks are absent. Intuitively, scenes with more spatial information such as residential areas and scenes with more spectral information like vegetations should require variable processing of recovering spatial information or spectral information. Inspired by this, we propose a novel HS pansharpening CNN to incorporate CNN into a tree structure, where spectral-spatial processing is automatically adjusted through different pansharpening paths. Major innovative contributions of our work can be identified as follows.

- 1) A tree-structured neural network (Tree-SNet). Tree-SNet maps a CNN onto a hierarchical tree structure. Routing nodes automatically tune input data to different paths corresponding to spatially local characteristic of the data, fulfilling an adaptive spatial adjustment, while in tree paths the data go through additional dynamic processings to reinforce the adaptiveness.
- 2) A spatial enhancement (SpatE) module and a spectral enhancement (SpecE) module. SpatE and SpecE modules are designed in the paths of the tree to further enhance spatial components and spectral components of the data, respectively, with the involvement of dynamic convolutions along the spatial dimension and the spectral dimension.
- 3) A patch integration rule connected to the data routing. After spatial and spectral processings on tree paths, we design a self-adjustable integration rule to fuse all the processed patches from the paths. The rule is associated with the data routing on the tree and collaboratively accomplishes the adaptive spatial and spectral processing on the tree.

The rest of this article is organized as follows. Section II proposes our pansharpening neural network, i.e., Tree-SNet. Section III shows our experimental results in detail. Finally, Section IV concludes this article.

II. METHODOLOGY

To handle the spatially varying characteristic of HS data in the duration of pansharpening, a feasible way is to design a pansharpening procedure capable of adaptively or automatically adjusting the spectral-spatial processing. More specifically, in this section, we seek to exploit an innovative Tree-SNet to deal with the problem.

A. Overall Design

HS pansharpening CNNs are usually connected to a learning process to reconstruct HRHS images by fusing LRHS images and PAN images, which consists of a training stage and a subsequent inferring stage. The two stages can be formulated as

$$\theta^{opt} = \arg \min_{\theta} l \left(\Phi \left(\tilde{\mathbf{H}}_t^{\uparrow}, \tilde{\mathbf{P}}_t; \theta \right), \mathbf{H}_t \right) \quad (1)$$

TABLE I
NOTATION

Symbol	Description
\mathbf{P}	Observed PAN image
\mathbf{H}	Observed LRHS image
\mathbf{H}^{\uparrow}	Pre-interpolated version of the observed LRHS image
$\tilde{\mathbf{P}}$	Synthetic reduced-resolution PAN image
$\tilde{\mathbf{H}}$	Synthetic reduced-resolution LRHS image
$\tilde{\mathbf{H}}^{\uparrow}$	Pre-interpolated version of the reduced-resolution LRHS image
\mathbf{P}_t	PAN image for training phase
\mathbf{H}_t	LRHS image for training phase
$\tilde{\mathbf{P}}_t$	Synthetic reduced-resolution PAN image for training phase
$\tilde{\mathbf{H}}_t$	Synthetic reduced-resolution LRHS image for training phase
$\tilde{\mathbf{H}}_t^{\uparrow}$	Pre-interpolated version of the synthetic reduced-resolution
$\hat{\mathbf{H}}$	Pansharpened HRHS image

and

$$\hat{\mathbf{H}} = \Phi \left(\mathbf{H}^{\uparrow}, \mathbf{P}; \theta^{opt} \right) \quad (2)$$

respectively, $\Phi(\cdot)$ represents the CNN pansharpening processing from a preinterpolated LRHS image and a PAN image to the pansharpened HRHS image through the CNN parameters θ , $l(\cdot)$ denotes the pansharpening loss in the training stage, which imposes constraints on the differences between the pansharpened HRHS image and the referenced ground-truth, and the meanings of other relative symbols can be found in Table I. In the training stage, the CNN pansharpening learns pansharpening parameters θ^{opt} , which is applied in the inferring stage to infer the pansharpened HRHS image.

The overall design of our pansharpening CNN, i.e., Tree-SNet, is graphically presented in Fig. 1. As shown in the figure, we utilize a hierarchical tree to guide the construction of Tree-SNet. The input PAN and preinterpolated LRHS are first stacked along the spectral dimension and then cropped to form patch samples for the training stage or the inferring stage. In the tree, routing nodes are employed to dynamically tune patch data, i.e., distribute the data to the paths of different spectral and spatial processings with adaptive weights. Thus, adaptive tuning of the pansharpening is achieved. In the practical fulfillment of the training stage, patch samples of training data are directly used as samples of adaptive moment estimation (Adam) optimizer, which trains the pansharpening parameters θ^{opt} with respect to a loss; and once θ^{opt} is determined, the pansharpening model $\Phi(\cdot)$ with fixed parameters θ^{opt} would be used to infer the pansharpening of the test or the incoming data.

B. Routing Nodes

Routing nodes $[\mathcal{N}_i^j]$ of the Tree-SNet tree serve as routers to determine how to distribute input data into paths of the tree. More specifically, routing nodes would measure spatial components and spectral components of the observed scene, and accordingly tune the settlement of the succeeding spatial-spectral processing with enhancement on tree paths. In such a way, the spatial details and spectral fidelities of the scene can be enhanced while keeping them balanced on the whole.

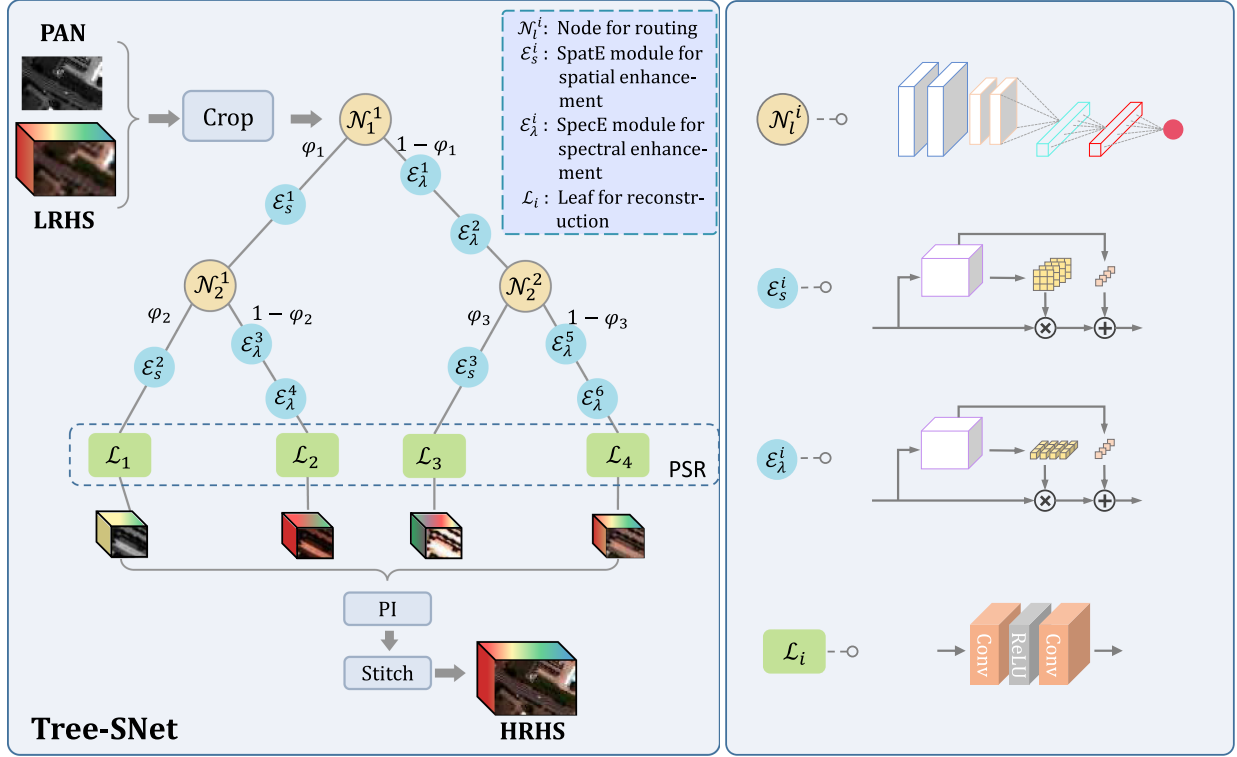


Fig. 1. HS pansharping based on the proposed Tree-SNet.

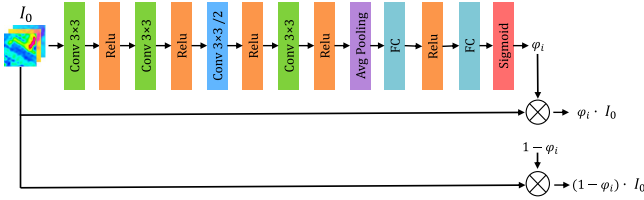


Fig. 2. Structure of a routing node in the proposed Tree-SNet.

More specifically, Fig. 2 graphically shows the architecture of a routing node. The routing node comprises a main path and two skip paths. The main path is a convolutional path, which is composed of consecutive convolutional layers, one of which is a two-stride layer, and followed by a global average pooling, several full connected (FC) layers and a sigmoid activation. The final FC layer is connected to a sigmoid activation to produce a φ_i value in $(0, 1)$. The output of the main path, i.e., φ_i , is used as a factor to tune the skip paths. Once the convolution parameters of the main path are determined, the factor φ_i would adaptively follow the characteristic of the input data of the routing node. The mathematical representation associated with the main path is as follows:

$$\begin{aligned} \mathbf{O}_1 &= \mathcal{F}(\mathbf{W}_1 * \mathbf{I}_0 + \mathbf{b}_1) \\ \mathbf{O}_4 &= \mathcal{F}(\mathbf{W}_4 * \mathcal{F}(\mathbf{W}_3 * \mathcal{F}(\mathbf{W}_2 * \mathbf{O}_1 + \mathbf{b}_2) + \mathbf{b}_3) + \mathbf{b}_4) \\ \varphi_i^j &= \mathcal{S}(\mathcal{FC}(\mathcal{F}(\mathcal{FC}(\mathcal{AP}(\mathbf{O}_4)))))) \end{aligned} \quad (3)$$

where \mathbf{I}_0 denotes the input of the node, \mathbf{O}_d ($d = 1, 2, 3, 4$) are the convolutional outputs of four layers, \mathbf{W}_d and \mathbf{b}_d ($d = 1, 2, 3, 4$)

represent the parameters of these layers, \mathcal{F} and \mathcal{S} stand for the Relu function and the Sigmoid function, respectively, and φ_i is the output of the routing node. This process for the main path of the i th node can be formulated as

$$\varphi_i = \mathcal{N}(\mathbf{I}_0; \boldsymbol{\theta}^N) \quad (4)$$

for brevity, where $\boldsymbol{\theta}^N$ presents the parameters of all layers in this process.

As regards the two skip paths, one skip path directly guides the input patch data \mathbf{I}_0 to multiply the factor φ_i ; while the other skip path guides the input data \mathbf{I}_0 to multiply the factor $(1 - \varphi_i)$. The outputs of two skip paths, i.e., $\varphi_i \cdot \mathbf{I}_0$ and $(1 - \varphi_i) \cdot \mathbf{I}_0$, will be fed into the subsequent spatial enhancement path and the spectral enhancement path, respectively, which implies adaptively tuning the spatial and the spectral processings of the local patch. Then, we have

$$\begin{aligned} \mathbf{F}_{\text{in1}} &= \varphi_i \cdot \mathbf{I}_0 \\ \mathbf{F}_{\text{in2}} &= (1 - \varphi_i) \cdot \mathbf{I}_0 \end{aligned} \quad (5)$$

where \mathbf{F}_{in1} and \mathbf{F}_{in2} are outputs of two skip paths, meanwhile, denote the inputs of the spatial enhancement path and the spectral enhancement path, respectively.

C. Spatial Enhancement and Spectral Enhancement Paths

After the adaptive data distribution by routing nodes, we perform further enhancing of the spatial and the spectral information. As shown in Fig. 1, in the left downward path of a node (i.e., a spatial enhancement path), the spatial information is restored by a spatial enhancement (SpatE) module, while the

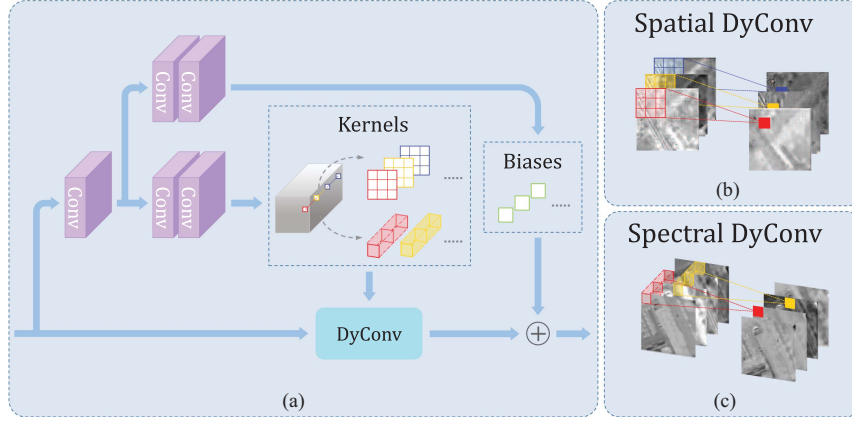


Fig. 3. SpatE module and SpecE module. (a) Overall structure. (b) Spatial dynamic convolution in SpatE. (c) Spectral dynamic convolution in SpecE.

spectral information is restored by two spectral enhancement (SpecE) modules in the right path (i.e., a spectral enhancement path). The SpatE and the SpecE modules are constructed as follows.

- 1) *SpatE Module*: The module aims to enhance the spatial components of observed scenes via automatic spatial adjustment. We resort to spatially dynamic convolution to achieve such a spatial adjustment. More specifically, the proposed SpatE module learns a mapping function $\mathcal{M}_1: \mathbf{F}_{in1} \rightarrow (\mathbf{W}, \mathbf{b})$ that transforms an input feature map \mathbf{F}_{in1} into the weights \mathbf{W} s and biases \mathbf{b} s of kernels associated with every pixel. The learned parameters are then applied to adaptively modulate the upcoming input feature maps along the spatial dimension to strengthen spatial details. The related mathematical representation is of the form

$$\begin{aligned} \mathbf{F}_{out1} &= \mathbf{b} + \mathbf{W} * \mathbf{F}_{in1} \\ (\mathbf{W}, \mathbf{b}) &= \mathcal{M}_1(\mathbf{F}_{in1}; \theta^S) \end{aligned} \quad (6)$$

where θ^S denotes the parameters of \mathcal{M}_1 . Fig. 3(a) and (b) shows the structure of the SpatE module and the implementation of the corresponding spatially dynamic convolution. In the module, several consecutive convolutional layers with ReLU activations are used to perform the mapping \mathcal{M}_1 , which outputs kernel parameters and then they are reshaped as weights and biases to fulfill spatial convolution. Spatial enhancement is conducted by applying these kernels and biases to the input feature maps along the spatial dimension in a channelwise and pixelwise manner, as shown in Fig. 3(b).

- 2) *SpecE Module*: In contrast to the SpatE module, the SpecE module is introduced to improve spectral fidelities of observed scenes. This module also learns a mapping function $\mathcal{M}_2: \mathbf{F}_{in2} \rightarrow (\mathbf{W}, \mathbf{b})$ that transforms the input feature maps \mathbf{F}_{in2} into kernel parameters. But different from SpatE, SpecE uses these kernel parameters to form spectral kernels of the spectral convolution operation to strengthen spectral fidelities. The connected formulation

is as follows:

$$\begin{aligned} \mathbf{F}_{out2} &= \mathbf{b} + \mathbf{W} * \mathbf{F}_{in2} \\ (\mathbf{W}, \mathbf{b}) &= \mathcal{M}_2(\mathbf{F}_{in2}; \theta^\lambda) \end{aligned} \quad (7)$$

where θ^λ presents the parameters of \mathcal{M}_2 . Fig. 3(a) and (c) displays the structure of the SpecE module and the implementation of its associated spectral convolution. As can be seen, the SpecE module is built on a similar structure to the SpatE module, except the dynamic convolution is performed with $1 \times 1 \times k$ spectral kernels.

In the spatial enhancement and the spectral enhancement paths, SpatE and SpecE modules use spatially and spectrally dynamic convolutions to lift spatial and spectral information, which to some degree are also spatially self-adjustable and thus collaborate with the data distribution by routing nodes and the connected reconstruction to achieve the spatial adaptiveness of our Tree-SNet.

D. Final Reconstruction

After the data distribution via routing nodes and the associated spatial and spectral enhancements on paths, a process of final reconstruction needs to be fulfilled to achieve the desired pansharpened HS image. The final reconstruction is composed of three sequential phases: Path-based spectral reconstruction (PSR), patch integration (PI), and stitching (as shown in Fig. 1).

In PSR, the output of each path is processed with two 1×1 convolutional layers (as shown on the bottom right part of Fig. 1) to restore the spectral information of each patch to yield pansharpened patches, which can be formulated as

$$\mathbf{L} = \mathcal{L}(\mathbf{F}_{out}; \theta^L) \quad (8)$$

where \mathcal{L} denotes two convolutional layers that reconstruct the spectral information and θ^L represents the parameters of the above layers. \mathbf{L} stands for the final output features of the patch.

PI seeks to integrate a set of patches from all the paths to form a final patch by a fusion rule, which corresponds to the data distribution by routing nodes. The associated mathematical

representation is as follows:

$$\begin{aligned} \mathbf{N}_{\text{out}} = & \varphi_1 \cdot \varphi_2 \cdot \mathbf{L}_1 + \varphi_1 \cdot (1 - \varphi_2) \cdot \mathbf{L}_2 \\ & + (1 - \varphi_1) \cdot \varphi_3 \cdot \mathbf{L}_3 + (1 - \varphi_1) \cdot (1 - \varphi_3) \cdot \mathbf{L}_4 \end{aligned} \quad (9)$$

where φ_i^j and \mathbf{N}_{out} stand for, respectively, the weight of spatial enhancement and the final HRHS patch by weighting the final output features.

Stitching aims at reconstructing the pansharpened HRHS image by combining all final patches according to their corresponding positions, which can be formulated as

$$\mathbf{H}_{\text{out}} = \mathcal{ST}(\mathbf{N}_{\text{out}}) \quad (10)$$

where \mathcal{ST} denotes the process of stitching by average the sum of all final patches according to the matching positions. In brief, the final reconstruction can be formulated as

$$\mathbf{H}_{\text{out}} = \mathcal{FR}(\mathbf{F}_{\text{out}}, \phi; \theta^L) \quad (11)$$

where \mathcal{FR} represents all the three consecutive operations of path-based spectral reconstruction, patch integration, and stitching, which collaboratively fulfill reconstructing the desired HRHS image. Especially, in the patch integration, we design a combining rule in a subtle way (as shown in Fig. 1) to integrate outputs of all paths, which corresponds to the routing of distributed data on the tree.

E. Summary

Our Tree-SNet maps a CNN on a tree structure, where routing nodes are responsible for data distribution corresponding to the spatially local characteristic of the HS image. Meanwhile, dynamic spatial enhancement and dynamic spectral enhancement are built into the paths of the tree to further strengthen the adaptiveness of the spatial-spectral processing, and then the connected reconstruction is designed to form the final pansharpening data. Algorithm II-E outlines the practical implementation of our Tree-SNet.

III. EXPERIMENTS

In this section, experimental results are presented to verify the effectiveness of the proposed Tree-SNet.

A. Experiment Setup

We tested our proposed Tree-SNet on four datasets. Table II shows the rough properties of those datasets.

1) The Pavia University dataset, collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over the University of Pavia, Italy, which contains an HS image with the spatial size of 610×340 and 102 spectral bands covering a spectral range from 0.4–0.9 μm . The bottom left part with the size 200×200 , the right part with the size 410×340 , and the rest parts are selected for test, train and validate, respectively. Totally, 7680 training samples and 768 validation samples with sizes of 20×20 are generated by the use of data augmentation (including rotation and flipping).

2) The Houston University dataset, acquired by the Compact Airborne Spectrographic Imager-1500 (CASI-1500) sensor over

Algorithm 1: Tree-SNet.

Input: $\tilde{\mathbf{H}}_t^\uparrow, \tilde{\mathbf{P}}_t, \mathbf{H}_t, \mathbf{H}^\uparrow, \mathbf{P}$
Output: $\hat{\mathbf{H}}$

```

1 while not done do
2   Crop LR-HR ;
3   for n epochs do
4      $\mathbf{I}_0 = (\tilde{\mathbf{H}}_t^\uparrow, \tilde{\mathbf{P}}_t)$ ;
5     for each node do
6        $\phi = \mathcal{N}(\mathbf{I}_0; \theta^N)$ ;
7        $\mathbf{F}_{in1} = \phi \times \mathbf{I}_0, \mathbf{F}_{out1} = \mathbf{b} + \mathbf{W} * \mathbf{F}_{in1}$ ;
8        $\mathbf{F}_{in2} = (1 - \phi) \times \mathbf{I}_0, \mathbf{F}_{out2} = \mathbf{b} + \mathbf{W} * \mathbf{F}_{in2}$ ;
9     end
10     $\mathbf{H}_{out} = \mathcal{FR}(\mathbf{F}_{out}, \phi; \theta^L)$ ;
11    Update  $\theta^{opt}$  with loss;
12     $\theta^{opt} = (\theta^N, \theta^S, \theta^\lambda, \theta^L)$ 
13  end
14 end
15 Obtain model  $\Phi$  ;
16 return  $\hat{\mathbf{H}} = \Phi(\mathbf{H}^\uparrow, \mathbf{P}; \theta^{opt})$ 

```

the University of Houston, America, which contains an HS image with the spatial size of 340×1325 and 144 spectral bands ranging from 0.4–1.0 μm . The top left part with the size 200×200 , the top right part with the size 410×340 , and the rest part are selected for test, train and validate, respectively. Totally, 7680 training samples and 768 validation samples with sizes of 20×20 are produced.

3) The Merced dataset, gathered by the Hyperion spectrometer and the Advanced Land Imager (ALI) sensor over agricultural and vegetational areas of Merced, California, which contains an HS image with the spatial size of 2100×180 and 134 spectral bands ranging from 0.4–2.5 μm and an accompanied PAN image with the spatial size of 6300×540 . Parts of the HS data with sizes of 150×150 , 600×150 and 400×150 , while the connected PAN data are selected for test, train, and validate, respectively. Totally, 9600 training samples and 960 validation samples with sizes of 20×20 are produced.

4) The Changping dataset, also acquired by the Hyperion spectrometer and the ALI sensor, but over an area of Changping, Beijing, which comprises an HS image with the spatial size of 2900×180 and 137 spectral bands ranging from 0.4–2.5 μm and an accompanied PAN image with the spatial size of 8700×540 . Parts of the HS data with the sizes of 180×180 , 1000×180 and 200×180 , and the connected PAN data, are selected for test, training, and validate, respectively. Totally, 9600 training samples and 960 validation samples with sizes of 20×20 are generated for our experiments.

In our experiments, two kinds of datasets are used: real datasets (including the Merced dataset and the Changping dataset) and simulated datasets (including the PaviaU dataset and the Houston dataset). A real dataset means it comprises both the observed LRHS data and the connected PAN data, whereas a simulated dataset means it contains only the observed LRHS data (i.e., no PAN data). Therefore, for simulated datasets, we have to use the observed LRHS data to artificially build the PAN data,

TABLE II
PROPERTIES OF THE DATASETS

Datasets	Data size	Spectral range	Bands channel	Instrument
Pavia University	HS 610×340	0.4-0.9 μm	102	ROSIS
Houston University	HS 340×1325	0.4-1.0 μm	144	CASI-1500
Merced	HS 2100×180	0.4-2.5 μm	134	Hyperion and ALI
Changping	HS 2900×180	0.4-2.5 μm	137	Hyperion and ALI

whose spatial information is the same as the observed LRHS data. Then, the observed LRHS images need to be degraded to a lower spatial resolution to collaborate the relatively high resolution PAN to perform reduced-resolution experiments. For real datasets, both the observed PAN data and observed LRHS data are available, where the PAN images are of the same spatial resolutions of the target pansharpened images. Then, the observed PAN images can naturally be used to pansharpen the observed LRHS images to the spatial resolution of the PANs, i.e., the full-resolution experiment. In addition, the observed PAN data and the observed LRHS data in real datasets can be degraded to a lower spatial resolution by involving blurring and down-sampling, and then also be used to conduct a reduced-resolution experiment, which is similar to that for the simulated datasets. Therefore, in our experiments, simulated datasets are usually only used to perform reduced-resolution experiments, while real datasets can be used to fulfill not only reduced-resolution experiments but also full-resolution experiments.

In reduced-resolution experiments, the observed LRHS images are used as the ground-truth images, which can be then utilized for calculating evaluation metrics CC, SAM, RMSE, ERGAS, and Q2n. However, in full-resolution experiments, the ground-truth images are absent. Therefore, evaluation metrics CC, SAM, RMSE, ERGAS, and Q2n cannot be obtained. In this situation, the evaluation metric QNR, which does not require the involvement of the ground-truth images, is utilized for evaluating the results of full-resolution experiments.

In our experiments, we utilize Wald's protocol [37] to arrange the simulated data of Pavia University and Houston University, where LRHS images are generated by blurring and down-sampling and PAN images are generated by averaging the visible spectral bands [38], and we also employ Wald's protocol in the Merced and Changping data to generate the associated LRHS images.

Several representative methods are used for comparison in our experiments, including GFPCA [39], Bayesian Sparse [40], MTF-GLP [17], GSA [41], SFIM [15], PNN, DRPNN [27], HyperPNN1 [34], HSpeNet1 [35], APNN [25], and RHDNet [42]. To facilitate a comprehensive comparison of the pansharpening methods, we employed five universal evaluation metrics: cross correlation (CC), spectral angle mapper (SAM) [43], root-mean-squared error (rmse) [6], erreur relative globale adimensionnelle de synthse (ERGAS) [44], and the extended universal image quality index (Q2n) [45].

B. Parameter Setting

We conducted all experiments using an NVIDIA GeForce RTX 2060 SUPER GPU. The convolutional layer employed the RELU activation function and rectified linear unit. The

TABLE III
QUALITY METRICS OF DIFFERENT HYPERSPECTRAL PANSHARPENING METHODS ON PAVIA UNIVERSITY DATASET

	CC(\uparrow)	SAM(\downarrow)	RMSE(\downarrow)	ERGAS(\downarrow)	Q2 ⁿ (\uparrow)
MTF-GLP	0.9393	5.2156	0.0353	3.1944	0.9048
GFPCA	0.8372	7.5954	0.0563	6.3472	0.6470
BayesNaive	0.9359	5.4216	0.0353	3.2396	0.9082
GSA	0.9273	5.8870	0.0381	4.0537	0.8738
SFIM	0.9365	5.2783	0.0361	3.5849	0.8813
PNN	0.9507	4.7235	0.0313	3.0095	0.9154
DRPNN	0.9617	4.1549	0.0278	2.6045	0.9381
HyperPNN1	0.9594	4.2986	0.0286	2.7068	0.9369
HSpeNet1	0.9604	4.1776	0.0281	2.7732	0.9361
APNN	0.9588	4.4997	0.0288	2.7149	0.9290
RHDNet	0.9629	3.9766	0.0275	2.5690	0.9402
Tree-SNet	0.9685	3.7226	0.0252	2.3981	0.9491
SpaE	0.9643	4.0802	0.0269	2.6655	0.94373
SpecE	0.9630	4.1506	0.0273	2.7109	0.9425

The bold numbers in the table denote the best performances under certain quantitative metrics.

loss function used in the experiment was $l1$ loss between the pansharpened data and the ground-truth. The training phase consisted of 100 000 iterations with a batch size of 64. The initial learning rate was set at 0.0001, decaying exponentially every 2000 iterations, and we utilized the Adam optimizer to minimize the loss function. Additionally, the tree structure of our Tree-SNet was set to be three layers empirically, considering the relatively low computational burden.

C. Simulated Image Experiments

These experiments are conducted on the Pavia University and Houston University datasets. PAN and LRHS images were generated using the Wald's protocol, as previously mentioned.

1) *Pavia University Dataset*: The spatial size of the degraded HS image is 122×68 pixels and the simulated PAN image is 610×340 pixels, while their ratio is 5. Table III shows the quantitative results of all these methods on the Pavia University dataset. As seen from the table, our proposed Tree-SNet achieves the best quality metrics among all these methods, with great improvements on all quantitative assessment results (i.e., CC, SAM, RMSE, ERGAS, and Q2ⁿ).

To identify artifacts and distortions in the pansharpened results. Fig. 4(a)–(o) shows the pseudocolor images achieved by different methods. Compared with other methods, the pansharpened image of our proposed method exhibits the best visual performance, with no noticeable artifacts and distortions in the pansharpened image, as shown in the enlarged area. The image of GFPCA shows the worst in the spatial preservation with much blurring. The image of SFIM has a large color difference compared with the ground-truth image, which means it performs the worst in the spectral restoration.

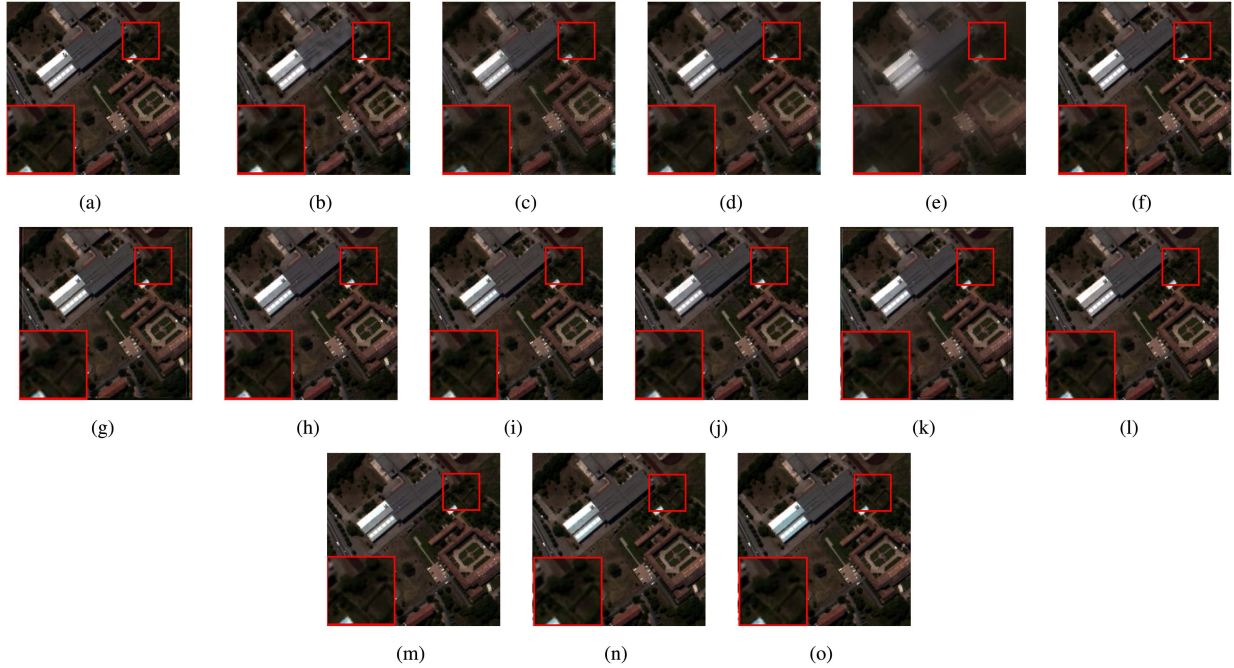


Fig. 4. HS pansharpening results for the Pavia University dataset (displayed in RGB). (a) Ground-truth. (b) GSA. (c) SFIM. (d) MTF-GLP. (e) GFPCA. (f) BayesSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HSpNet1. (k) APNN. (l) RHDNet. (m) Tree-SNet. (n) SpaE. (o) SpecE.

Fig. 5(a)–(o) shows the error images of different methods. The error image of our method shows smaller error compared with other error images in the high-frequency area, which means the pansharpening image of our method is more similar to the ground-truth image, with good spatial and spectral restoration.

To further demonstrate the effectiveness of the tree structure mechanism in our method, we graphically illustrate the feature maps produced by only spatial enhancement path and by only spectral enhancement path in Fig. 6. Each of the figures is a pseudocolor image composited by the three visible-light bands. As shown in the figure, the feature maps from different enhancement paths indicate different spatial and spectral characteristics. The feature map from the spatial enhancement path contains a large amount of details, e.g., sharp edges and clear structure contours, but yields bad color restoration compared to the ground-truth [as shown in Fig. 4(a)] which means obviously spectral distortion. In contrast, the feature map from the spectral enhancement path performs well in keeping spectral-fidelity; however, it is blurred, which contributes to the loss of spatial details. Therefore, it is verified that the spatial enhancement path tends to extract and improve spatial information, while the spectral enhancement path tends to refine spectral information; thus, the validity of our tree-structure method is further indicated.

2) *Houston University Dataset*: The spatial size of the degraded HS image is 68×265 pixels, and the simulated PAN image is 340×1325 pixels while their ratio is 5. The comparison of quality metrics for different methods with the Houston University dataset are shown in Table IV, from which our Tree-SNet performs the best with these scenes for all mentioned metrics. GFPCA and MTF-GLP obtain much worse performance than other methods, while BayesSparse and SFIM achieve better performance among other traditional methods.

TABLE IV
QUALITY METRICS OF DIFFERENT HYPERSPECTRAL PANSHARPENING METHODS ON HOUSTON UNIVERSITY DATASET

	CC(\uparrow)	SAM(\downarrow)	RMSE(\downarrow)	ERGAS(\downarrow)	Q2 m (\uparrow)
MTF-GLP	0.8382	6.4266	0.0337	3.6621	0.8131
GFPCA	0.9024	7.8804	0.0333	5.1751	0.6791
BayesSparse	0.8384	5.7959	0.0343	3.7304	0.8248
GSA	0.8620	5.6674	0.0320	3.5376	0.8287
SFIM	0.8381	6.8155	0.0346	3.8582	0.7849
PNN	0.8758	5.1532	0.0289	3.2751	0.8416
DRPNN	0.9071	4.1682	0.0258	2.8050	0.8884
HyperPNN1	0.9041	4.1984	0.0257	2.8387	0.8868
HSpNet1	0.9176	3.8101	0.0247	2.6847	0.9012
APNN	0.8977	4.4067	0.0265	2.9410	0.8723
RHDNet	0.9115	3.9467	0.0254	2.7476	0.8957
Tree-SNet	0.9239	3.6064	0.0235	2.5631	0.9097
SpaE	0.9168	3.9184	0.0243	2.6716	0.9024
SpecE	0.9152	3.9866	0.0245	2.6975	0.9004

The bold numbers in the table denote the best performances under certain quantitative metrics.

To identify the visual performance of different methods, Fig. 7(a)–(o) shows the pansharpening results of different methods. From these images, the result generated by our method exhibits the best performance with no blurring or color difference in the enlarged area. GFPCA performs worst in the spatial restoration while SFIM performs worst in the spectral restoration compared with other methods. In general, traditional methods exhibit worse performance in both spatial and spectral restoration compared with CNN methods.

In addition, Fig. 8(a)–(o) shows the error images of different methods. The error images show the difference between pansharpening images and ground-truth images. From Fig. 7, the pansharpening result of our proposed method obtains the

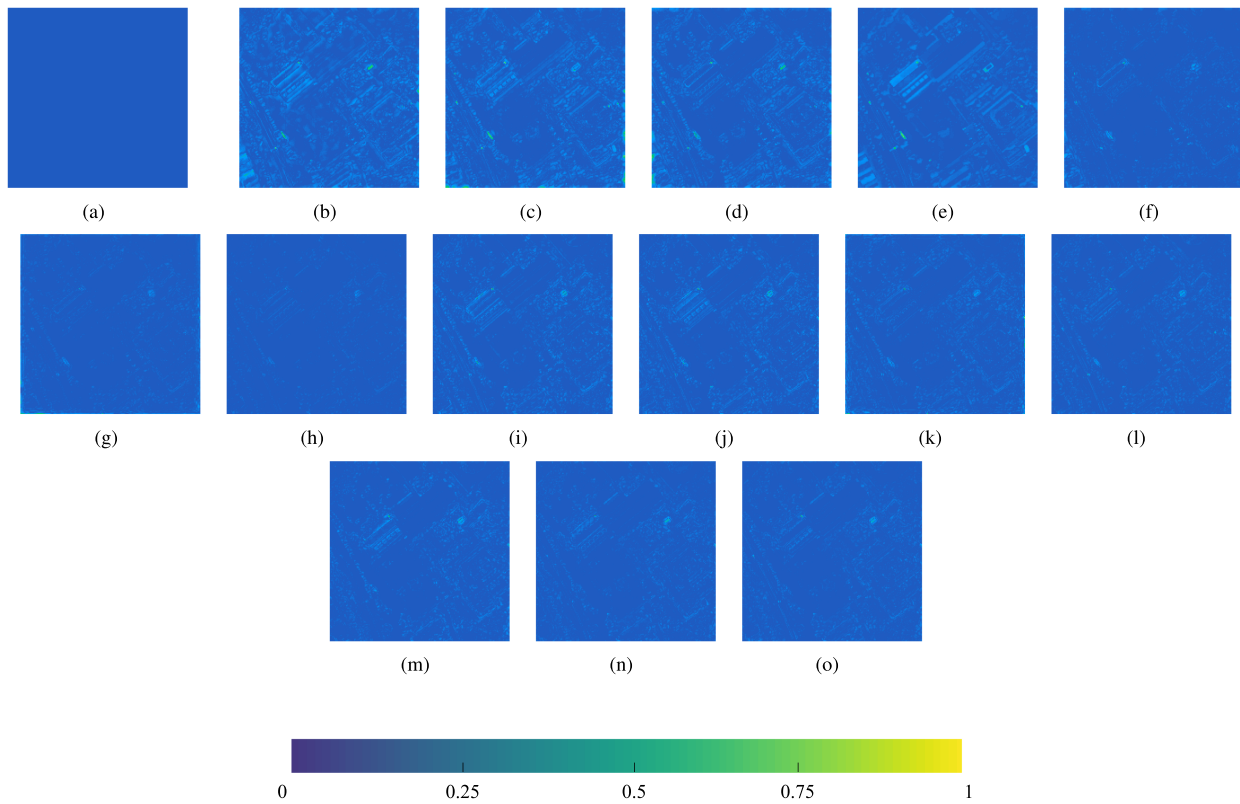


Fig. 5. Error images of different methods on the Pavia University dataset. (a) Ground-truth. (b) GSA. (c) SFIM. (d) MTF-GLP. (e) GFPCA. (f) BayesSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HSpNet1. (k) APNN. (l) RHDNet. (m) Tree-SNet. (n) SpaE. (o) SpecE.

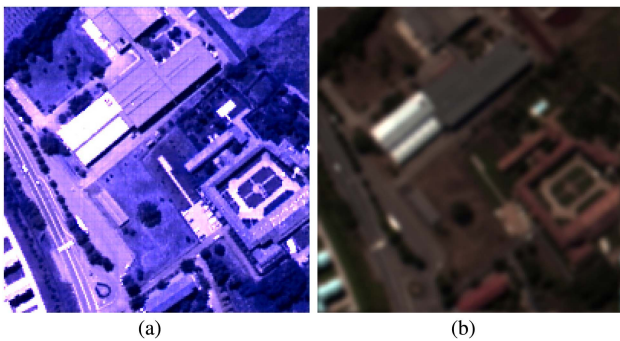


Fig. 6. Feature maps for the Pavia University dataset produced by (a) spatial enhancement path and by (b) spectral enhancement path.

best visual performance, which means it is the most similar to the ground-truth image among these mentioned methods. Generally, the results of CNN methods achieve much smaller error than traditional methods.

D. Real Image Experiments

The Merced dataset and the Changing dataset are utilized to perform evaluation in the real image experiments. Two kinds of experiments are carried out on the two datasets: the reduced-resolution experiment and the full-resolution experiment.

1) *Merced Dataset*: In the reduced-resolution experiment, the HS and PAN images are downsampled simultaneously and

then fused to yield HRHS images. Similar to the simulated-data experiment, the original HS images serve as reference images to evaluate the performance of different methods. The dimension of the test PAN image is 450×450 pixels, and the size of the experimental HS image is 150×150 pixels. Fig. 9(a)–(o) shows the visual results of all compared pansharpening methods for simulated Merced dataset, and Fig. 10(a)–(o) shows the error images compared to the ground-truth image. As observed from these images, our Tree-SNet obtains the best performance in both spatial and spectral restorations.

In the full-resolution experiment, the HS and PAN images are used as the inputs to generate a real HRHS image. The dimension of the test PAN image is 300×300 pixels, and the size of the experimental HS image is 100×100 pixels.

Table V gives the quantitative evaluation results for both the reduce-resolution experiments and the full-resolution experiments. CC, SAM, rmse, ERGAS, and $Q2^n$ are for the reduce-resolution experiments; while the quality with no reference (QNR) protocol [46] is used to evaluate the full-resolution pansharpening images. As shown in this table, our proposed Tree-SNet achieves the best performance with respect to the four quality metrics in the reduced-resolution experiment, while HSpNet1 yields better result under the SAM metric and the QNR metric. HSpNet1 is specifically designed for recovering spectral information, and the Merced dataset may tend to comprise relatively less spatial information but relatively more spectral information. Thus, HSpNet1 performs better

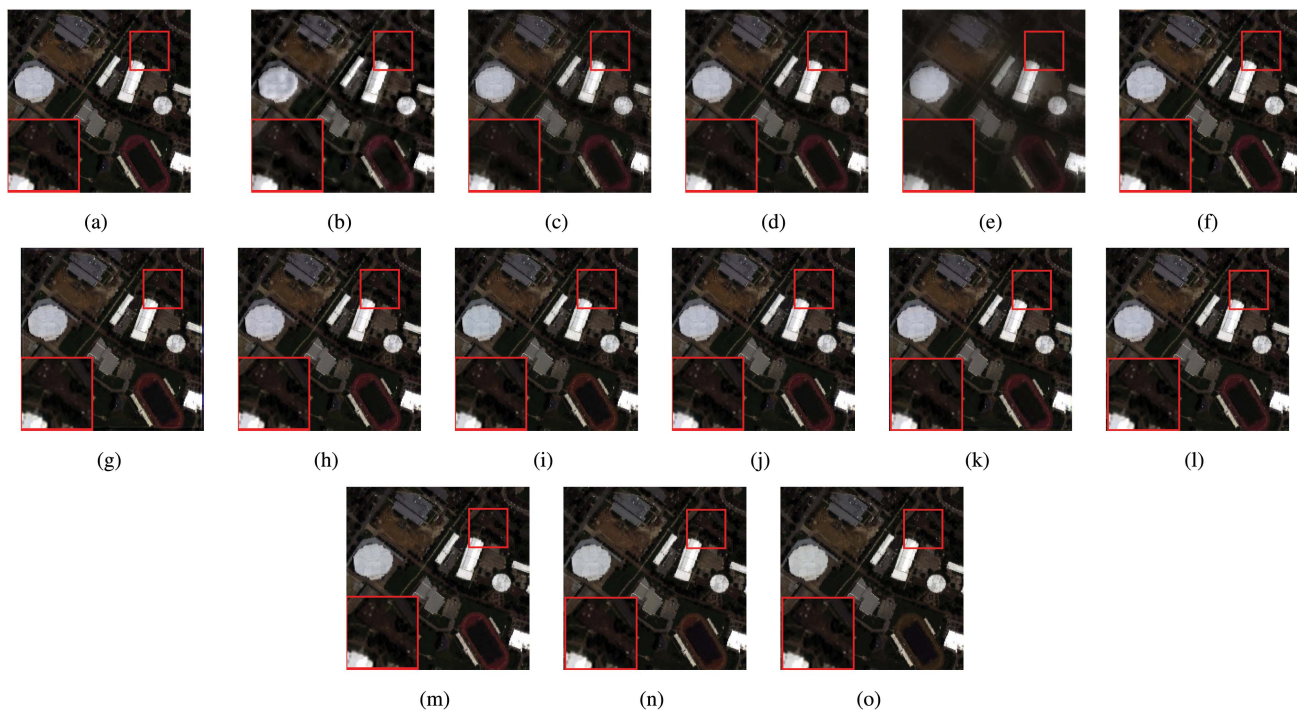


Fig. 7. HS pansharpening results for the Houston University dataset (displayed in RGB). (a) Ground-truth. (b) GSA. (c) SFIM. (d) MTF-GLP. (e) GFPCA. (f) BayesSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HSpeNet1. (k) APNN. (l) RHDNet. (m) Tree-SNet. (n) SpaE. (o) SpecE.

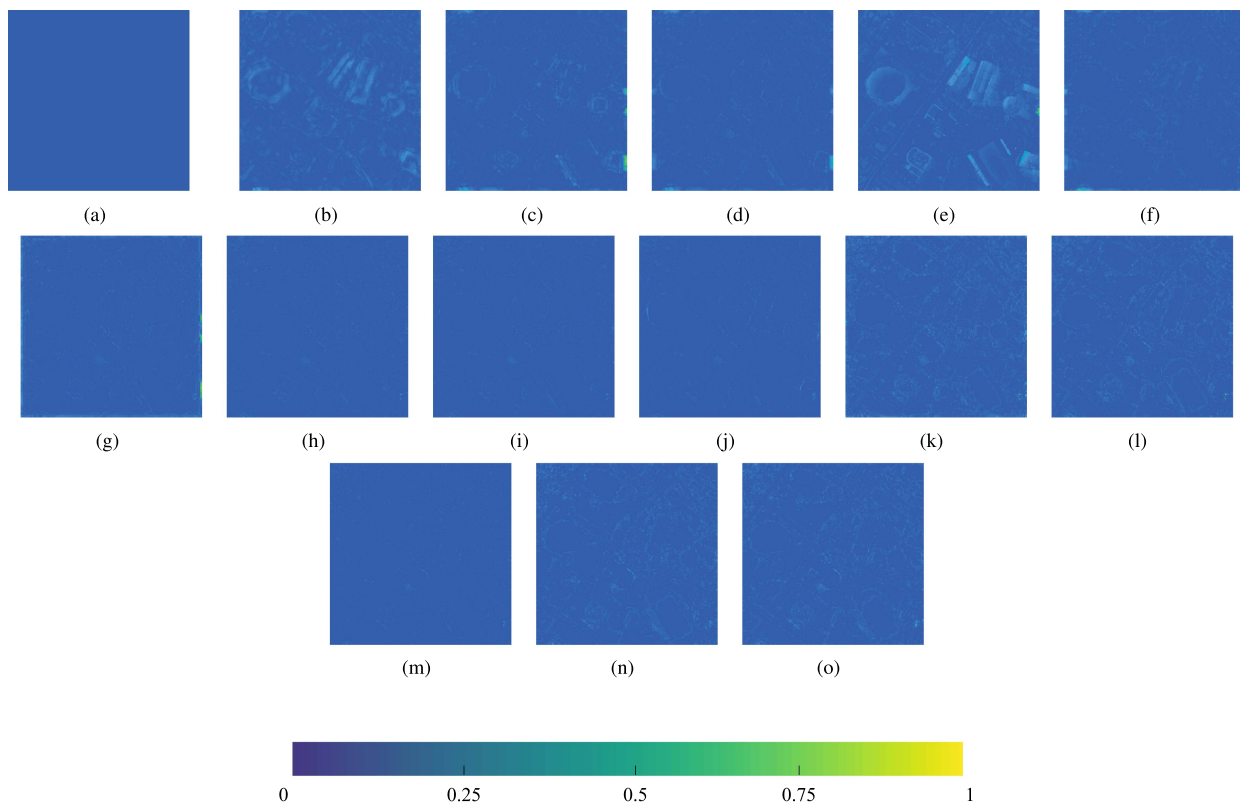


Fig. 8. Error images of different methods on the Houston University dataset. (a) Ground-truth. (b) GSA. (c) SFIM. (d) MTF-GLP. (e) GFPCA. (f) BayesSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HSpeNet1. (k) APNN. (l) RHDNet. (m) Tree-SNet. (n) SpaE. (o) SpecE.

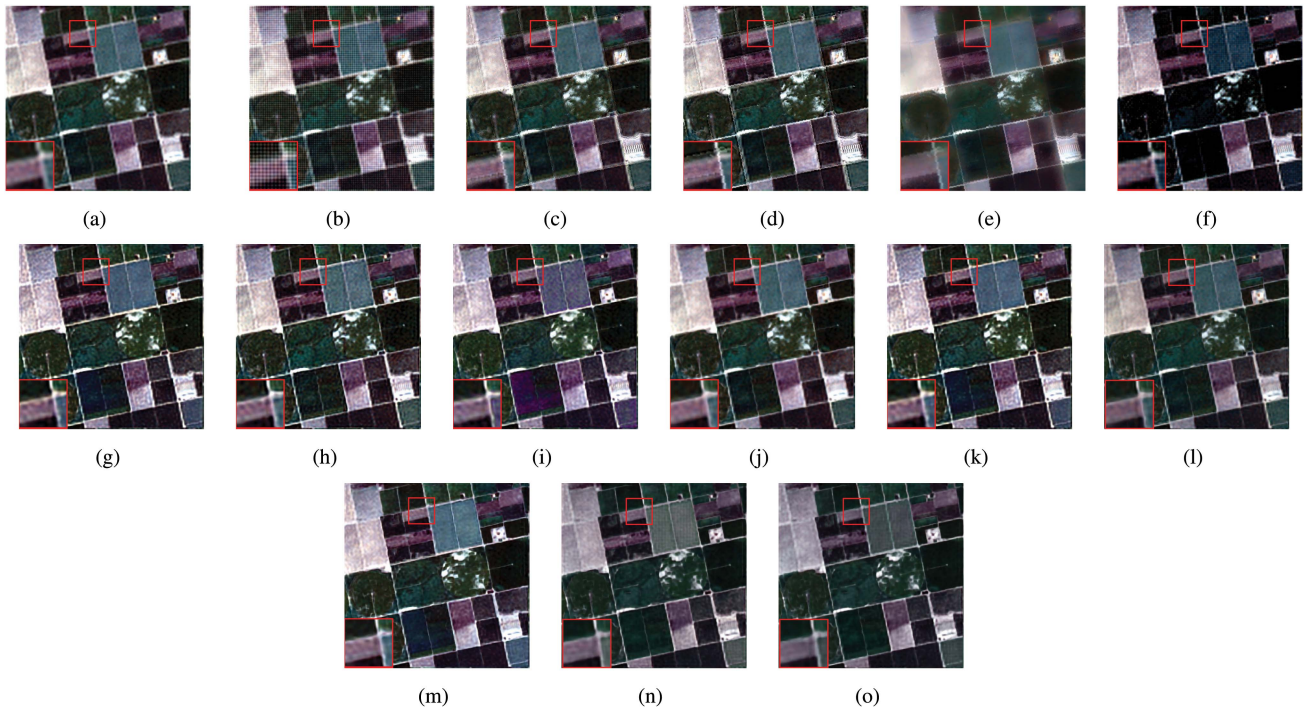


Fig. 9. HS pansharpening results for the reduced-resolution Merced dataset (displayed in RGB). (a) Ground-truth. (b) GSA. (c) SFIM. (d) MTF-GLP. (e) GFPCA. (f) BayesSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HSpeNet1. (k) APNN. (l) RHDNet. (m) Tree-SNet. (n) SpaE. (o) SpecE.

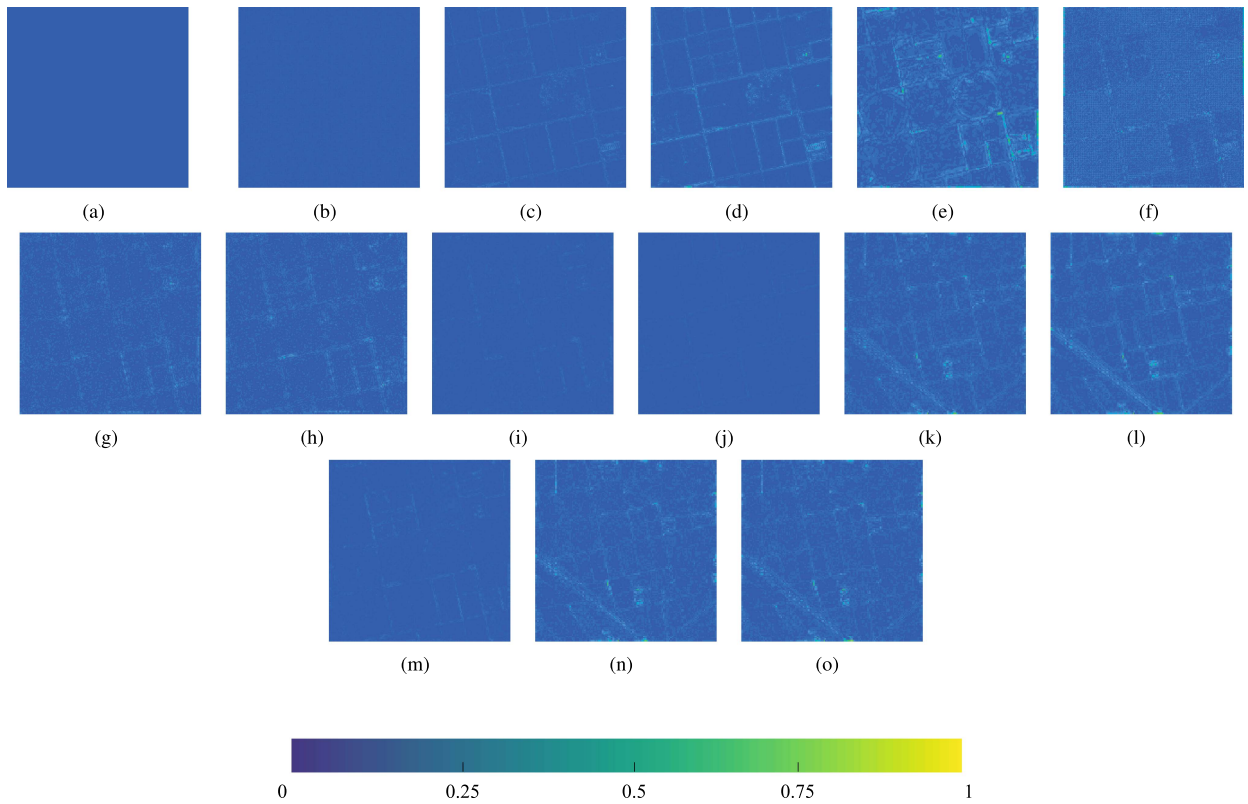


Fig. 10. Error images of different methods on the reduced-resolution Merced dataset. (a) Ground-truth. (b) GSA. (c) SFIM. (d) MTF-GLP. (e) GFPCA. (f) BayesSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HSpeNet1. (k) APNN. (l) RHDNet. (m) Tree-SNet. (n) SpaE. (o) SpecE.

TABLE V
QUALITY METRICS OF DIFFERENT HYPERSPECTRAL PANSHARPENING METHODS ON MERCED DATASET

	CC(\uparrow)	SAM(\downarrow)	RMSE(\downarrow)	ERGAS(\downarrow)	Q2 ⁿ (\uparrow)	QNR(\uparrow)
MTF-GLP	0.9189	3.9391	0.0299	6.9360	0.8314	0.8068
GFPCA	0.8143	7.4520	0.0450	10.7745	0.5477	0.81958
BayesSparse	0.9520	4.6994	0.0270	6.0273	0.8996	0.93832
GSA	0.9570	3.4558	0.0215	5.1404	0.9032	0.90785
SFIM	0.9349	3.7041	0.0265	6.5352	0.8514	0.88546
PNN	0.96733	3.083	0.018322	4.3788	0.93333	0.94768
DRPNN	0.96956	2.9103	0.017698	4.217	0.93698	0.95542
HyperPNN1	0.96886	3.1315	0.018102	4.2647	0.93782	0.94705
HSpeNet1	0.96504	2.7407	0.019425	4.3424	0.93062	0.96259
APNN	0.8977	2.9986	0.0175	4.2948	0.0.9363	0.9481
RHDNet	0.9693	2.7958	0.0179	4.1648	0.9388	0.9610
Tree-SNet	0.9733	2.809	0.016724	3.9202	0.94783	0.94217
SpaE	0.9691	3.2817	0.0183	4.1484	0.9419	0.9351
SpecE	0.9696	3.2358	0.0181	4.1780	0.9404	0.9381

The bold numbers in the table denote the best performances under certain quantitative metrics.

TABLE VI
QUALITY METRICS OF DIFFERENT HYPERSPECTRAL PANSHARPENING METHODS ON CHANGPING DATASET

	CC(\uparrow)	SAM(\downarrow)	RMSE(\downarrow)	ERGAS(\downarrow)	Q2 ⁿ (\uparrow)	QNR(\uparrow)
MTF-GLP	0.8577	3.5921	0.0210	5.4895	0.7330	0.6985
GFPCA	0.7797	5.8202	0.0261	6.8325	0.4469	0.8406
BayesSparse	0.9042	3.9398	0.0171	4.5201	0.8212	0.6659
GSA	0.8132	4.1288	0.0251	6.6639	0.6753	0.6594
SFIM	0.8770	3.5933	0.0193	5.0635	0.7553	0.7911
PNN	0.9315	3.8338	0.0144	3.8563	0.8686	0.9047
DRPNN	0.9361	3.4850	0.0139	3.7239	0.8756	0.9101
HyperPNN1	0.9341	3.7778	0.0141	3.7891	0.8759	0.8960
HSpeNet1	0.9312	3.3058	0.0144	3.8555	0.8653	0.8960
APNN	0.9327	3.7995	0.0142	3.7684	0.8772	0.9075
RHDNet	0.9408	3.3287	0.0134	3.5926	0.8868	0.8743
Tree-SNet	0.9469	3.5102	0.0127	3.4143	0.9016	0.9157
SpaE	0.9327	3.8721	0.0142	3.8374	0.8739	0.8862
SpecE	0.9355	3.8843	0.0140	3.7576	0.8801	0.8737

The bold numbers in the table denote the best performances under certain quantitative metrics.

under SAM, which is a spectral-preserving metric. As regards the QNR metric, its calculation involves the considerations of both spectral distortion and spatial distortion. As the Merced scene may contain relatively less spatial information but relatively more spectral information, the QNR metric on that dataset is especially vulnerable to spectral distortion. Therefore, HSpeNet1 tends to achieve high QNR performance on the Merced dataset. However, considering the experimental results of all the quantitative quality metrics illustrated in Table V, our Tree-SNet still achieves overall excellent performance. Fig. 11(a)–(n) graphically illustrates the pansharpening results of different methods for the real Merced dataset. The result of our Tree-SNet achieves better visual performance compared to the other methods, with less blurring and spectral distortion.

2) *Changping dataset*: In the reduced-resolution experiment, the size of the test PAN image is 180×180 , and the size of the test HS image is 60×60 . Fig. 12(a)–(o) shows the pseudocolor images of all compared pansharpening methods for reduced-resolution experiment s, and Fig. 13(a)–(o) shows

the error images. Our method exhibits the best performance in spatial restorations.

In the full-resolution experiment, the original HS and the connected PAN images are used to yield a real HRHS image. The size of the test PAN image is 360×360 , and the size of the connected HS image is 120×120 .

Table VI gives the quantitative evaluation results for both the reduce-resolution experiments and the full-resolution experiments. As shown in this table, our proposed Tree-SNet achieves the best performance with respect to the five quality metrics (i.e., CC, rmse, ERGAS, Q2ⁿ, and QNR). Similar to the experimental results on the Merced dataset, HSpeNet1 produces better result under SAM. This may be also due to that Changping dataset contains relatively less spatial information but relatively more spectral information. Fig. 14(a)–(n) gives the visual results of different methods for full-resolution Changping dataset. The result of our Tree-SNet exhibits the best visual performance, with least blurring and spatial distortion.

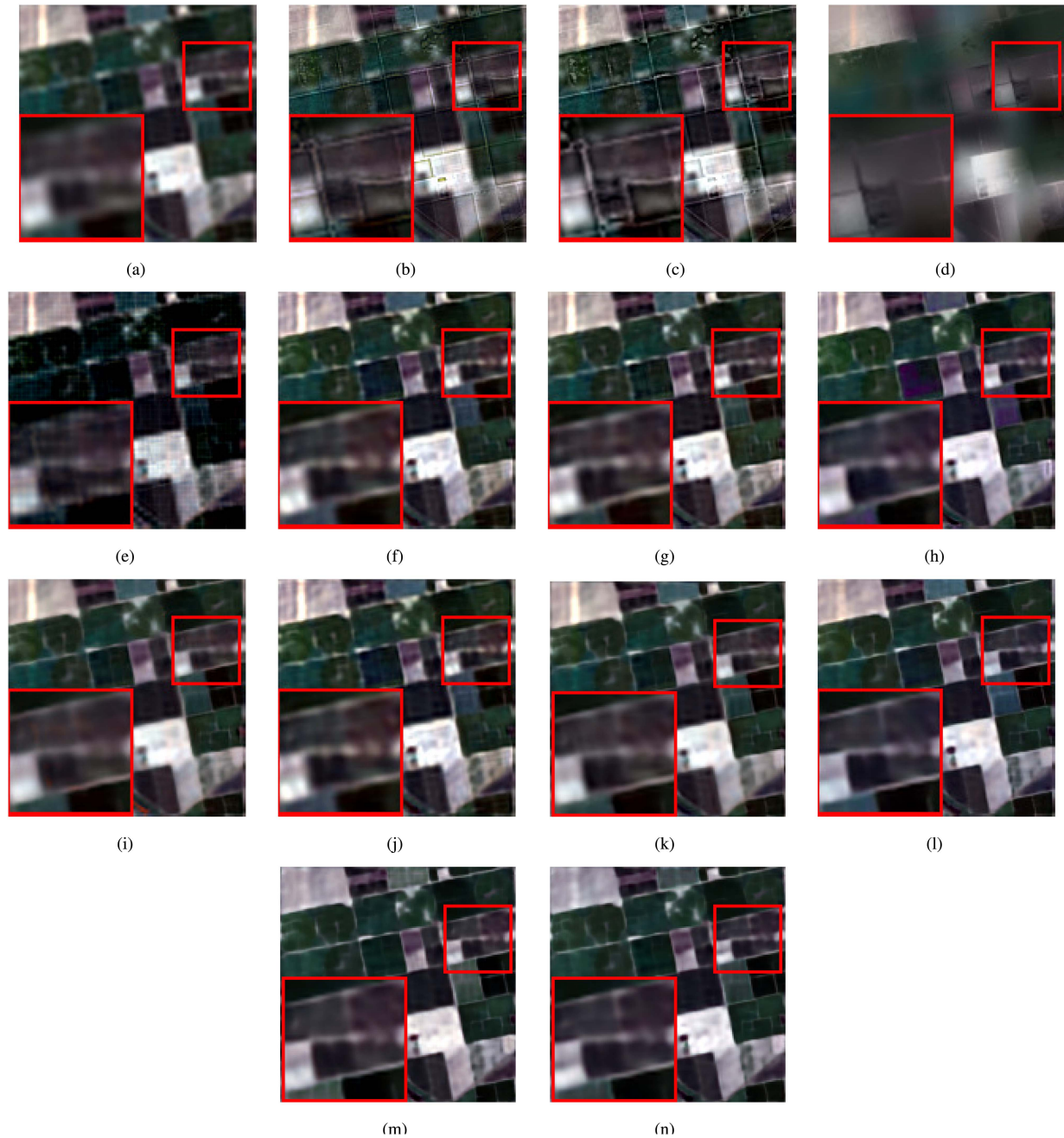


Fig. 11. HS pansharpener results for the full-resolution Merced dataset (displayed in RGB). (a) GSA. (b) SFIM. (c) MTF-GLP. (d) GFPCA. (e) BayesSparse. (f) PNN. (g) DRPNN. (h) HyperPNN1. (i) HSpNet1. (j) APNN. (k) RHDNet. (l) Tree-SNet. (m) SpaE. (n) SpecE.

E. Ablation Study

In order to further examine the contributions of specific structures in the proposed Tree-SNet, we conduct ablation studies on the Pavia University dataset, the Houston University dataset, the Merced dataset and the Changping dataset. We mainly test the effectiveness of our Tree-SNet strategy comprising both spatial enhancement paths and spectral enhancement paths. In Tables III, IV, V, and VI, the result of removing spectral enhancement paths and the result of removing spatial enhancement paths are illustrated, where SpaE refers to the situation of removing spectral enhancement paths, while SpecE denotes the

situation of removing spatial enhancement paths. As observed from the tables, when the spectral enhancement paths or the spatial enhancement paths are removed from our Tree-SNet, its performance drops significantly.

With the removal of the spatial enhancement path or the spectral enhancement path, the routing nodes in the tree structure lose the function of dynamic tuning, i.e., the routing nodes are unable to distribute the data to the paths of different spectral and spatial processings with adaptive weights. Thus, the adaptively tuning of the pansharpener is unavailable. In this situation, individual spatial enhancement or individual spectral enhancement may occur, which leads to spectral/spatial distortion in

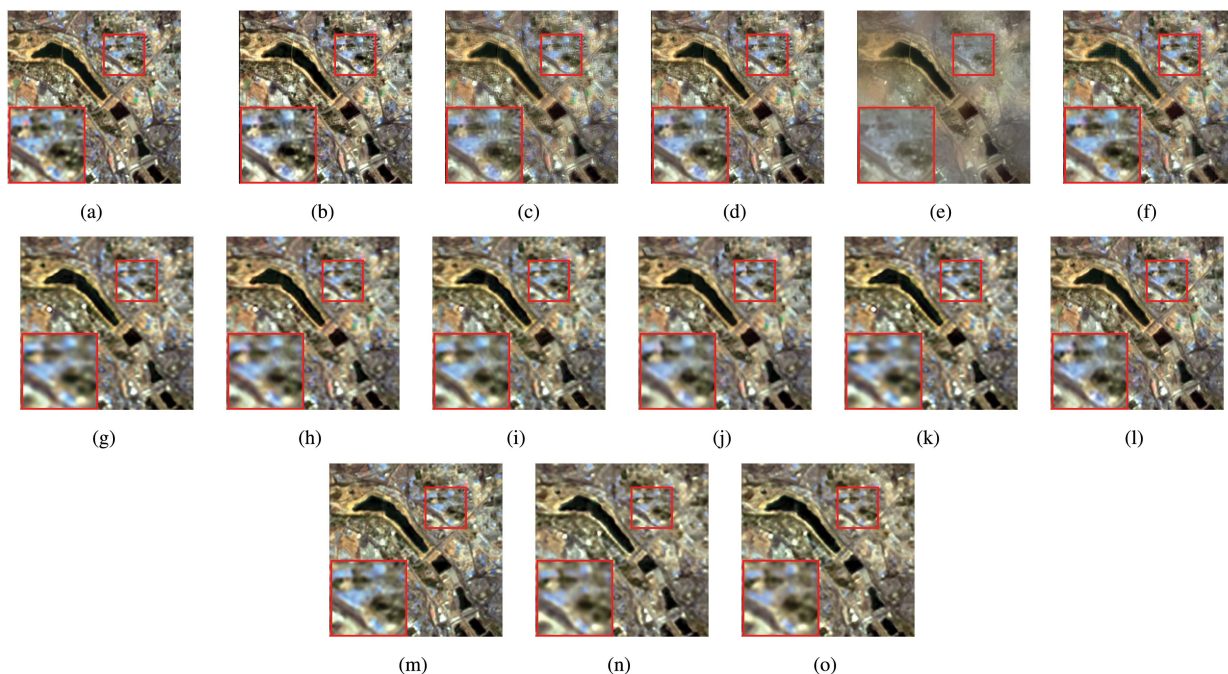


Fig. 12. HS pansharpening results for the reduced-resolution Changping dataset (displayed in RGB). (a) Ground-truth. (b) GSA. (c) SFIM. (d) MTF-GLP. (e) GFPCA. (f) BayesSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HSpeNet1. (k) APNN. (l) RHDNet. (m) Tree-SNet. (n) SpaE. (o) SpecE.

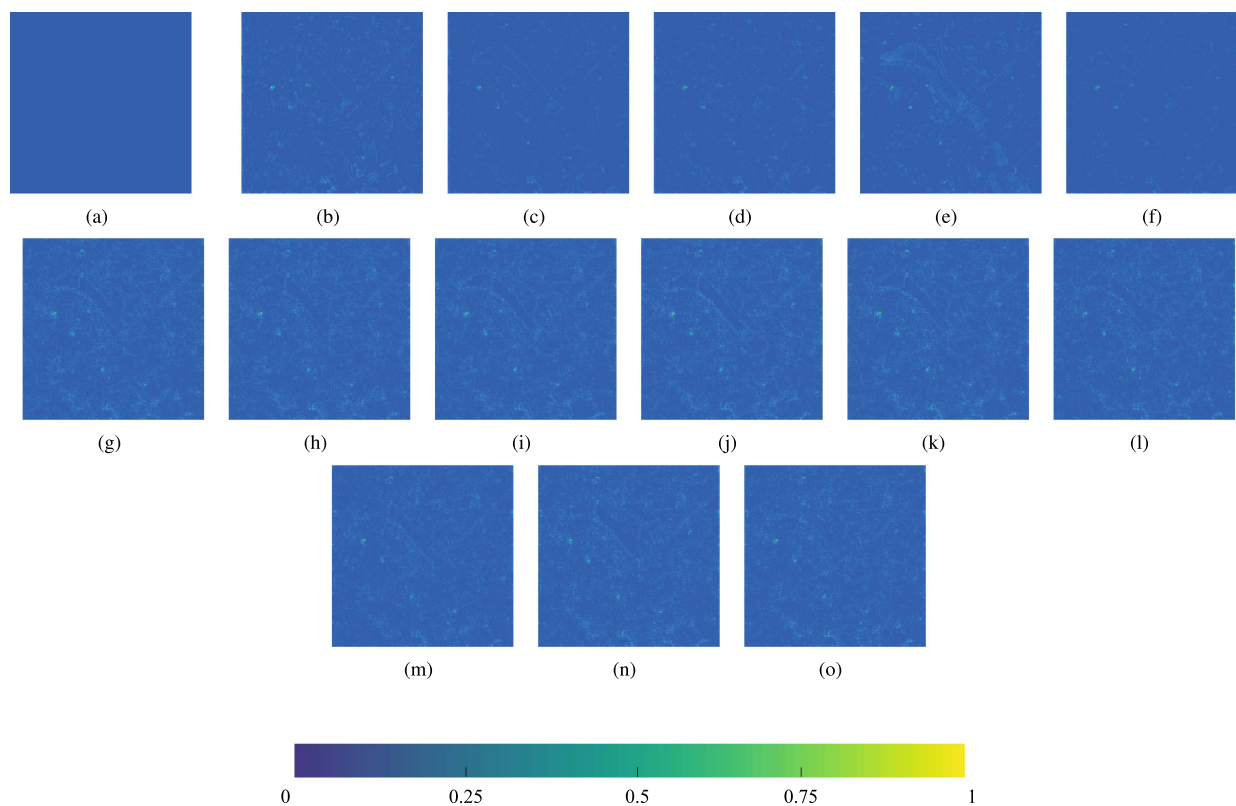


Fig. 13. Error images of different methods on the reduced-resolution Changping dataset. (a) Ground-truth. (b) GSA. (c) SFIM. (d) MTF-GLP. (e) GFPCA. (f) BayesSparse. (g) PNN. (h) DRPNN. (i) HyperPNN1. (j) HSpeNet1. (k) APNN. (l) RHDNet. (m) Tree-SNet. (n) SpaE. (o) SpecE.



Fig. 14. HS pansharpening results for the full-resolution Changping dataset (displayed in RGB). (a) GSA. (b) SFIM. (c) MTF-GLP. (d) GFPCA. (e) BayesSparse. (f) PNN. (g) DRPNN. (h) HyperPNN1. (i) HSpNet1. (j) APNN. (k) RHDNet. (l) Tree-SNet. (m) SpaE. (n) SpecE.

the pansharpening result. Therefore, this experiment indicates the validity of the structure of Tree-SNet which is composed of both spatial enhancement paths and spectral enhancement paths.

IV. CONCLUSION

In this article, we develop an innovative Tree-SNet method to deal with the nonstationary spatial-spectral coupling in HS pansharpening. Our design introduces a hierarchical tree structure to guide the pansharpening CNN to fulfill the spatial-spectral adaptiveness. In the tree structure-guided neural network, routing nodes are employed to automatically tune the data delivered

to tree paths according to the local data characteristic; whereas SpatE modules and SpecE modules are dynamically run in tree paths to further reinforce the adaptive spatial-spectral processing. Our method achieves remarkable performance in tackling the HS pansharpening task.

REFERENCES

- [1] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 279–317, Dec. 2019.
- [2] L. Loncan et al., "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.

- [3] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.
- [4] Z. Wu et al., "Scheduling-guided automatic processing of massive hyperspectral image classification on cloud computing architectures," *IEEE Trans. Cybern.*, vol. 51, no. 7, pp. 3588–3601, Jul. 2021.
- [5] P. Du, S. Liu, J. Xia, and Y. Zhao, "Information fusion techniques for change detection from multi-temporal remote sensing images," *Inf. Fusion*, vol. 14, no. 1, pp. 19–27, Jan. 2013.
- [6] A. Mohammadzadeh, A. Tavakoli, and M. J. V. Zojj, "Road extraction based on fuzzy logic and mathematical morphology from pan-sharpened IKONOS images," *Photogrammetric Rec.*, vol. 21, no. 113, pp. 44–60, Mar. 2006.
- [7] X. Wei and M. Yuan, "Adversarial pan-sharpening attacks for object detection in remote sensing," *Pattern Recognit.*, vol. 139, Jul. 2023, Art. no. 109466.
- [8] G. Vivone et al., "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.
- [9] F. D. Javan et al., "A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 171, pp. 101–117, Jan. 2021.
- [10] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.
- [11] A. Mookambiga and V. Gomathi, "Comprehensive review on fusion techniques for spatial information enhancement in hyperspectral imagery," *Multidimensional Syst. Signal Process.*, vol. 27, no. 4, pp. 863–889, Apr. 2016.
- [12] G. Vivone, A. Garzelli, Y. Xu, W. Liao, and J. Chanussot, "Panchromatic and hyperspectral image fusion: Outcome of the 2022 WHISPERS hyperspectral pansharpening challenge," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 166–179, 2023.
- [13] Te-Ming Tu, P. S. Huang, Chung-Ling Hung, and Chien-Ping Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 309–312, Oct. 2004.
- [14] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008.
- [15] J. Liu, "Smoothing filter based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, May 2000.
- [16] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez, "Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005.
- [17] B. Aiuzzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and pan imagery," *Photogrammetric Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.
- [18] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 930–934, May 2014.
- [19] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rouge, "A variational model for P+ XS image fusion," *Int. J. Comput. Vis.*, vol. 69, no. 1, pp. 43–58, Apr. 2006.
- [20] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "A new pansharpening algorithm based on total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 318–322, Jan. 2014.
- [21] X. He, L. Condat, J. M. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4160–4174, Sep. 2014.
- [22] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.
- [23] M. Golipour, H. Ghassemian, and F. Mirzapour, "Integrating hierarchical segmentation maps with MRF prior for classification of hyperspectral images in a Bayesian framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 805–816, Feb. 2016.
- [24] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, Sep. 2004.
- [25] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.
- [26] L. He et al., "Pansharpening via detail injection based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1188–1204, Apr. 2019.
- [27] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multi-spectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.
- [28] Y. Rao, L. He, and J. Zhu, "A residual convolutional neural network for pan-sharpening," in *Proc. Int. Workshop Remote Sens. with Intell. Process.*, 2017, pp. 1–4.
- [29] G. Vivone et al., "A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods," *IEEE Geosci. Remote Sens. Magazine*, vol. 9, no. 1, pp. 53–81, Mar. 2021.
- [30] G. Masi, D. Cozzolino, and L. Verdoliva, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, Jul. 2016, Art. no. 594.
- [31] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multidepth convolutional neural network for remote sensing imagery pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 978–989, Mar. 2018.
- [32] Y. Zheng, J. Li, Y. Li, K. Cao, and K. Wang, "Deep residual learning for boosting the accuracy of hyperspectral pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1435–1439, Aug. 2020.
- [33] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1529–1543, May 2020.
- [34] L. He, J. Zhu, J. Li, A. Plaza, J. Chanussot, and B. Li, "HyperPNN: Hyperspectral pansharpening via spectrally predictive convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 3092–3100, Aug. 2019.
- [35] L. He, J. Zhu, J. Li, D. Meng, J. Chanussot, and A. Plaza, "Spectral-fidelity convolutional neural networks for hyperspectral pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5898–5914, 2020.
- [36] 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 Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7303–7317, Dec. 2022.
- [37] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetric Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, Jun. 1997.
- [38] B. Aiuzzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.
- [39] W. Liao et al., "Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2984–2996, Jun. 2015.
- [40] Q. Wei, N. Dobigeon, and J. -Y. Tourneret, "Bayesian fusion of multi-band images," *J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, Sep. 2015.
- [41] B. Aiuzzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS + Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.
- [42] J. Qu, T. Zhang, W. Dong, Y. Yang, and Y. Li, "Cross-guided pyramid attention-based residual hyperdense network for hyperspectral image pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.
- [43] R. Yuhas, A. Goetz, and J. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, vol. 1, pp. 147–149.
- [44] L. Wald, *Data fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*. Paris, France: Les Presses de l'École des Mines, 2002.
- [45] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 662–665, Oct. 2009.
- [46] L. Alparone, B. Aiuzzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogrammetric Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.