HS-McHF: Hypersharpening With Multicomponent-Based Hierarchical Features Fusion Network

Zeinab Dehghan, *Student Member, IEEE*, Jingxiang Yang, *Member, IEEE*, Milad Taleby Ahvanooey, *Senior Member, IEEE*, Abdolraheem Khader, *Member, IEEE*, and Liang Xiao, *Senior Member, IEEE*

***Abstract*—Hypersharpening is one of the fusion-based superresolution approaches in remote sensing that improves the spatial and spectral resolution of a hyperspectral image (HSI) with a low spatial resolution. This requirement is achieved by fusing the HSI with a panchromatic image that has high spatial resolution to generate a newly combined variant, which has high spatial quality and high spectral resolution. While several studies in the literature applied neural networks for hypersharpening, there exist unsolved issues such as how to deeply discover the spatial–spectral correlation and inject geometric details without distortion. To address these issues, we propose a hypersharpening technique by applying a multicomponent-based hierarchical fusion network called (HS-McHF), which hierarchically learns the low and high-frequency spatial–spectral features. We then suggest an optimization model to discover the correlation between low-resolution HSI and high-resolution panchromatic images and solve it by stochastic gradient descent through a neural network. Moreover, we decrease the band overlapping in the initial HSI by combining a deconvolution model to prevent spectral distortion and reduce the noise in the panchromatic geometric details injection by deploying an encoder–decoder network. Our extensive experiments demonstrate that the HS-McHF provides superior efficiency compared to state-of-the-art fusion based superresolution approaches.**

***Index Terms*—Deep-learning, hyperspectral image (HSI), image fusion, pansharpening, superresolution.**

Manuscript received 6 March 2024; revised 8 May 2024; accepted 29 May 2024. Date of publication 4 June 2024; date of current version 14 June 2024. The work of Milad Taleby Ahvanooey was supported in part by the Ulam Research Fellowship under Grant BPN/ULM/2022/1/00069, and in part by the Polish National Agency for Academic Exchange (Narodowa Agencja Wymiany Akademickiej (NAWA)). This work was supported in part by the Natural Science Foundation of Jiangsu Province under Grant BK20200465, in part by the Jiangsu Provincial Social Developing Project under Grant BE2018727, in part by the Jiangsu Geological Bureau Research Project under Grant 2023KY11, in part by the Fundamental Research Funds for the Central Universities under Grant JSGP202204 and Grant 30920021134. *(Corresponding authors: Liang Xiao; Jingxiang Yang.)*

Zeinab Dehghan, Jingxiang Yang, Abdolraheem Khader, and Liang Xiao are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: dehghanzeinab@njust.edu.cn; yang123jx@njust.edu.cn; abdolraheem@njust.edu.cn; xiaoliang@njust.edu.cn).

Milad Taleby Ahvanooey is with the Faculty of Electronics and Information Technology, Warsaw University of Technology (WUT), 00–665 Warszawa, Poland, and also with the School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore 639798 (e-mail: m.taleby@ieee.org).

Digital Object Identifier 10.1109/JSTARS.2024.3408806

I. Introduction

SPECTRAL imaging refers to a set of analytical image processing approaches that integrates spectroscopy and conventional imaging data to gain spectral and spatial information from a visual object. Hyperspectral image (HSI) is generated by a spectral imaging approach that contain thousands of bands, which are captured by cameras with high spectral resolution in much narrower spectral bands (10–20 nm) [1]. Due to current software and sensor constraints, combining spectral, spatial, and temporal information through a single sensor is a challenging task. HSI with a high spatial–spectral resolution is obtained by integrating the complementary information from low spatial resolution HSI and high spatial resolution image data. This technique is called the superresolution mechanism, which addresses this challenging task by providing high-frequency information from low-resolution input data through two models: single hyperspectral superresolution (ShSR) and fusion-based hyperspectral superresolution (FhSR) with assistive information from a panchromatic (PAN) image, RGB image, or multispectral image (MSI). In the ShSR approaches such as [2], [3], and [4], authors have developed several models, such as filtering models as linear bicubic to approximate the pixels of the neighborhood, sparsity to explore spectral redundancy and deep learning algorithms, which solve the ShSR problem by applying convolutional neural network (CNN).

Fusion-based mechanisms can combine the spatial imaging data as guided images in the high-resolution with the spectral information in the low-resolution HSI (LR-HSI) to produce a high spatial–spectral resolution HSI. Hypersharpening as an efficient FhSR mechanism can generate output with high spatial–spectral resolution from low spatial resolution HSI and high spatial resolution PAN images. In the literature, researchers have suggested several models to address FhSR, such as component substitution (CS) [5], multiresolution analysis (MRA) [6], optimization algorithms [7], and CNN [8], [9]. Over the last decade, because of the advancements in computing systems (e.g., CPU and GPU), neural network applications have demonstrated more satisfactory results and better performance in several areas (e.g., superresolution in remote sensing). Another advantage of the deep neural network is the ability to overcome the limitations of traditional methods, such as CS and MRA, which require

prior image conditions, such as low rank and sparsity. To avoid the constraints and better explore geometric details and spectral information, we employ the multicomponent-based hierarchical features fusion network to enhance the superresolution of HSIs. The key point of our proposed method contains three modules. We detect the spatial feature mapping from the HSI with low resolution to the output with high resolution using a gradient descent optimization approach. Neural networks typically use the concatenation of HSI and PAN as their input. However, the HSI downsampling process leads to the loss of a significant amount of spectral information. This results in incomplete features that are inadequate to inject spatial–spectral details. To address this issue, the second module extracts the spectral features by learning the spectral response matrix to convert the assistive image to an HSI and combine it with the geometric feature maps. While neural-based networks have numerous advantages, they often lack sufficient spatial details because they cannot extract multiscale information. A solution to this problem is multiscale modeling, which combines low and high-resolution information to learn discriminative features. Therefore, we provide hierarchical blocks in the third module. These blocks generate a multiscale feature representation that can hierarchically combine two types of shallow and deep information. In addition, discovering the HSI channel's relevancy can effectively increase the model's efficiency and achieve decisive results, but it is ignored in many fusion-based networks in the literature. For this purpose, we investigate the potential of devising a spectral deconvolution (SD) method in our model for crafting bands overlapping reduction and having HSI with sharper spectral information.

In the following, we summarize the contributions of this study.

1) To inject the spatial information into spectral features and learn the spatial–spectral correlation, we propose a multicomponent-based hierarchical fusion network (HS-McHF) for FhSR, which consists of multiscale subnetworks to estimate the high-resolution HSI (HR-HSI) from the low-resolution version of HSI and PAN images.
2) To achieve the highest sharp spectral features, we address the bands overlapping issue by solving the optimization problem based on energy function adjustment between LR-HSI and PAN images. Then, we fuse the spatial and spectral features with high-pass information from the assistive image to craft HR-HSI with full features.
3) To obtain the spatial, spectral, and structural features from the observed pair, we design the loss functions, which utilize the simultaneously preserving spatial, structural, and spectral information to enhance the efficiency of the HS-McHF.

II. Related Works

In this section, we briefly discuss the state-of-the-art fusion-based superresolution neural network-based models considering their highlights and limitations. In the literature, researchers have developed neural networks-based models to hypersharpen satellite images [10], [11], [12] with fixed prior knowledge (e.g., HSI and PAN) and different network designs in terms of layers and objective function. These fusion-based models can be divided into three categories: input-level, feature-level, and model-based [13]. At the input level, researchers expand HSI in the LR-HSI to a size of PAN to generate an upsampled version of LR-HSI and enter these variables as input to the network. For instance, Masi et al. [12] introduced a deep learning-based pansharpening model in which a CNN with three layers was developed to create high-resolution MSI by combining the PAN and upsampled version of the LR-HSI. Moreover, Luo et al. [9] have presented an approach based on CNN, which combines the LR-HSI and PAN to develop the model input. This study applied a CNN-based block with three layers and a multiscale block containing several convolutional layers to extract and merge the results of the two blocks to obtain the output. Furthermore, He et al. [14] proposed a residual neural network, which applies seven convolutional layers using skip connection and L1 norm in the loss function. In this work, the authors assign the features of low-resolution MSI as PAN to feed the inputs. Later, Zheng et al. [15] developed a deep residual network (ResNet) that employs a spatial and spectral attention mechanism for pansharpening tasks. In this study, researchers extrapolated the input images by performing the DHP algorithm and PAN, as well as applying the multiple channel spatial attention blocks to fuse the inputs and generate the output.

In the feature-level-based approaches, authors merge the spectral and spatial features through the LR-HSI and PAN for hypersharping the HR-HSI. For example, Shim et al. [16] proposed an end-to-end deformable convolution network that trained based on the similarity between low- and high-resolution pixels to merge downsampled PAN information with the features of the low-resolution image. Furthermore, Shuang et al. [17] designed a fusion-based network considering several gradient projection-based blocks that extract the PAN and LR-HSI features separately. Wang et al. [18] presented a fusion model that considers a dual-path deep residual-based neural network to extract spectral features and a high-pass block to craft spatial information. Uezato et al. [19] designed an unsupervised remote sensing image fusion. In this work, the authors utilized a deep encoder–decoder network based on a skip connection to extract features of inputs and generate a fused output. Later, Shuang et al. [20] proposed a sparse coding network, which divides the features of LR-HSI into two groups (irrelevant and relevant PAN feature maps) to produce side, unique, and correlated information. As a result, they reconstructed the fused image according to the merged attributes. In a recent study by Wu et al. [21], an approach using a long short-term memories network was proposed for pansharpening application. The researchers reconstructed the LR-HSI by utilizing deconvolution bidirection learning to upsample it. Moreover, they employed two separate branches to extract both spatial and spectral features, which were combined using elementwise addition. Furthermore, He et al. [22] introduced a dynamic pansharpening CNN-based approach. They generated spatially adaptive rules and spectral predictive to produce dynamic pansharpening results and reduce spectral distortions. Later, Fan et al. [23] presented a pansharpening network based on a transformer model with cross-attention.

They used the multiscale embedding sequence of LR-HSI and PAN images to reconstruct the HR-HSI.

In the third category, researchers utilize optimization algorithms combining mathematical methods to solve fusion-based superresolution problems. For example, Scarpa et al. [10] utilized a pretrained three-layer CNN with residual learning method, which combines the PAN, upsampled LR-HSI, and radiometric indices extracted from LR-HSI. In this research, the authors applied the stochastic gradient descent (SGD) algorithm during the optimization phase to fine-tune the network parameters to enhance efficiency. Another optimization-based model has been proposed by Yin et al. [24] that trains the kernel in the convolutional layer to exploit the input features for reconstructing the output image. Moreover, they employed an iterative soft thresholding activation function to find feature maps and optimize the model parameters.

III. Proposed Method

In this section, we explain the structure of our fusion-based superresolution model in detail. The HS-McHF consists of three units that generate HSI with high spatial resolution. In each unit after the image reconstruction, we use a block to enhance the quality of the output image. At the first unit, the output is approximated using HSI with low-resolution and guidance PAN, and then the overlapping bands are reduced to maintain more spectral information. In the next unit, we fuse spatial, spectral, and high-pass features and denoise them to preserve only the essential details. Finally, the last unit includes a multiscale subnetwork for processing and learning the spatial and spectral features. Each subnetwork in our model has a novel feature extraction module in self-attention form and a deep residual feature extraction block for generating the HR-HSI.

A. Problem Formulation

In this section, we consider the HS-McHF as the superresolution problem to generate the HSI (denoted $\mathbf{X} \in \mathbb{R}^{W\times H\times C}$) with high spatial and spectral resolution from the observed low spatial resolution HSI and high spatial resolution guided image, where W and H are width and height, respectively, and C is the number of spectral bands. Let us assume that $\mathbf{Y} \in \mathbb{R}^{w\times h\times C}$ denotes the low spatial resolution HSI and $\mathbf{P} \in \mathbb{R}^{W\times H\times 1}$ is high spatial resolution PAN. Herein, $W > w$ and $H > h$ are the spatial width and height of the observed pairs, where $a = W/w = H/h$ is the upsampled factor. In this study, we assume that $\mathbf{Y}$ and $\mathbf{P}$ are spatial and spectral downsampled versions of the target HR-HSI, receptively, which can obtained by the following equations:

$$\mathbf{Y} = \mathbf{BX},\ \mathbf{P} = \mathbf{XR} \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{wh\times WH}$ is the spatial degradation matrix and $\mathbf{R} \in \mathbb{R}^{C\times 1}$ represents spectral degradation operator. The optimization problem can be formulated according to (1) to reconstruct the $\hat{\mathbf{X}}$ from the $\mathbf{Y}$ and $\mathbf{P}$

$$\underset{\mathbf{X}}{\operatorname{argmin}} f_1(\mathbf{Y}, \mathbf{BX}) + f_2(\mathbf{P}, \mathbf{XR}) + R(\mathbf{X}) \tag{2}$$

where $f_1(.)$ and $f_2(.)$ indicate the spectral and spatial cost function, respectively, and $R(.)$ is a regularization term, which can be replaced by the encoder–decoder neural network. Therefore, the optimization formula can be defined as follows:

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} ||\mathbf{Y} - \mathbf{BX}||_F^2 + ||\mathbf{P} - \mathbf{XR}||_F^2 + R(\mathbf{X}) \tag{3}$$

the first part of this formula reduces the spectral difference, and the second part injects the spatial information into the fused image. $R(\mathbf{X})$ regularizes the proposed network based on encoder–decoder network.

B. Network Architecture

To solve the superresolution fusion-based problem in (3), we propose the deep neural network, which consists of three units: initial estimation (IE), multicomponents fusion (MCF), and hierarchical features aggregation (HFA) unit. Fig. 1 shows the overall architecture of the HS-McHF model, which is described as follows.

1) IE Unit: In the IE unit, we first estimate the high-resolution version of $\mathbf{Y}$ (called $\hat{\mathbf{Y}}_1$) using $\mathbf{P}$ as guided image according to first part of (2), then correct the bands overlapping to preserve spectral features to enhance the HSI quality. To learn the spatial feature's mapping from $\mathbf{Y}$ to $\hat{\mathbf{Y}}_1$, we utilize the estimator proposed in the band-dependent spatial-detail (BDSD) approach [25]. For this purpose, we apply spatial degradation $\mathbf{B}$, which includes convolution layer with downsampling factor $\frac{1}{a}$ and its inverse operator $\mathbf{B}^{\mathrm{T}}$ to create $\mathbf{B}^{\mathrm{T}}\mathbf{BP}$ (downsampled and then upsampled version of $\mathbf{P}$). We believe that the geometric details transformation from $\mathbf{P}$ to $\mathbf{B}^{\mathrm{T}}\mathbf{BP}$ can be a factor of the feature mapping from $\mathbf{B}^{\mathrm{T}}\mathbf{Y}$ (upsampled version of $\mathbf{Y}$) to $\hat{\mathbf{Y}}_1$

$$\hat{\mathbf{Y}}_1 - \mathbf{B}^{\mathrm{T}}\mathbf{Y} = W \times \left(\mathbf{P} - \mathbf{B}^{\mathrm{T}}\mathbf{BP}\right)\mathbf{R}^{\mathrm{T}} \tag{4}$$

where $\hat{\mathbf{Y}}_1$ is the output of the BDSD block, and direct estimation of coefficients (W) is not possible. Hence, we utilize $\mathbf{X}$ as a reference to optimize and update the W weights by BDSD with least square [26] to inject the geometric features into $\mathbf{Y}$ until convergence. The optimization model can be written for each band $k = \{1, \ldots, C\}$ as follows:

$$W_k^* = \underset{W_k}{\operatorname{argmin}} \parallel \mathbf{X}_k - \left(\mathbf{B}_k^{\mathrm{T}}\mathbf{Y}_k + (W_k \times \mathbf{H}_k)\right) \parallel^2 \tag{5}$$

where $\mathbf{H} = (\mathbf{P} - \mathbf{B}^{\mathrm{T}}\mathbf{BP})\mathbf{R}^{\mathrm{T}}$ and $\mathbf{H}_k \in \mathbb{R}^{WH\times 1}$. $W_k^* \in \mathbb{R}^{WH\times WH}$ represents the optimal weights, which calculate with differentiating the (5) with respect to the coefficients W_k and setting the derivatives to the vector $\mathbf{0}$

$$\mathbf{H}_k^T\left((\mathbf{X}_k - \mathbf{B}_k^{\mathrm{T}}\mathbf{Y}_k) - (W_k \times \mathbf{H}_k)\right) = \mathbf{0}. \tag{6}$$

In case of least square-based BDSD, (6) can be solved by iterative solution according to the closed-form with the following formula:

$$W_k = [\mathbf{H}_k^{\mathrm{T}}\mathbf{H}_k]^{-1}\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{X}_k - \mathbf{B}_k^{\mathrm{T}}\mathbf{Y}_k\right) \tag{7}$$

whereas the closed-form solution is extremely expensive to compute, using an iterative method is more computationally efficient than the closed-form solution for the least squares

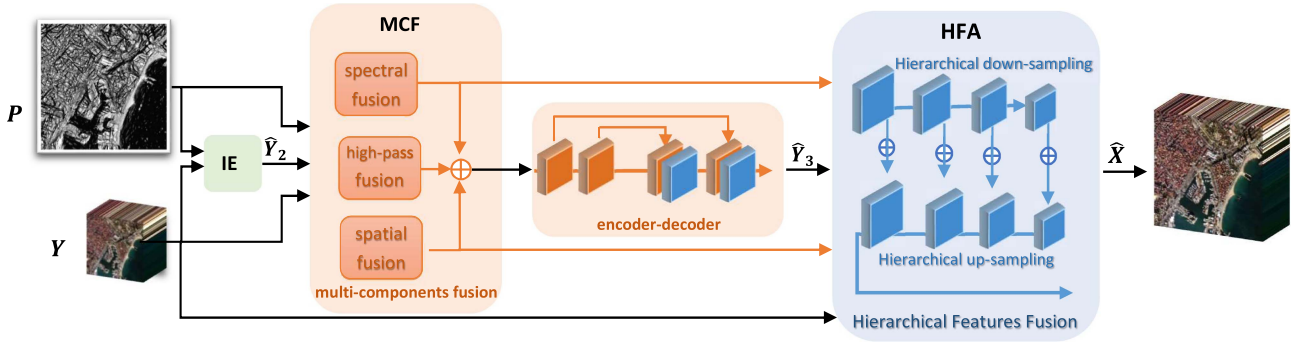


Fig. 1. Architecture of proposed model (HS-McHF) that contains three units: IE, MCF, and HFA units. After the image reconstruction, quality enhancement is applied to each unit. In IE, after adjusting the optimization coefficients via the BDSD approach and unfolding it by SGD, we deconvolve each band by addressing the bands overlapping problem. In MCF, we exploit and combine spatial, spectral, and high-pass properties to reach the HSI with full features. Then, we denoise them by deploying the encoder–decoder network. Eventually, HFA includes multiple subnetworks to prepare multiscale feature maps from high to low levels and then aggregate the extracted information hierarchically from low to high levels.

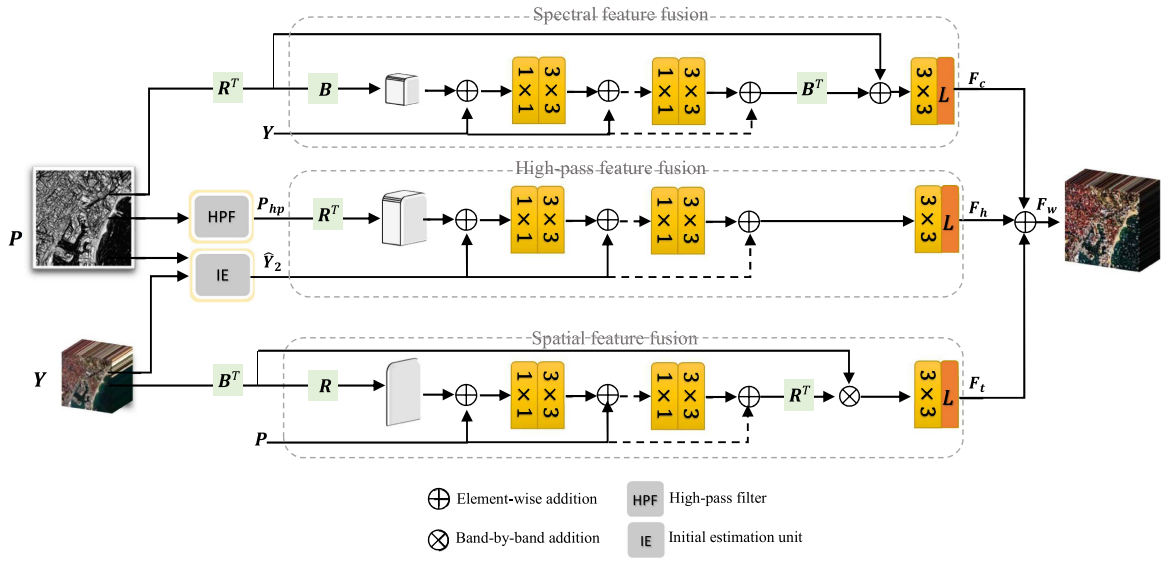


Fig. 2. MCF unit includes spatial feature fusion, high-pass feature fusion, and spectral feature fusion blocks, which extract spatial, high-pass, and spectral information in parallel and combine all details simultaneously.

problem. Therefore, we simulate the CNN with SGD steps to optimize the weights and approximate the $\hat{\mathbf{Y}}_1$.

a) Spectral Deconvolution: The signal-to-noise ratio (SNR) is a fundamental mechanism to calculate the instrument's performance. To achieve the highest SNR in each band of HSI, hyperspectral sensors are designed to have the most overlapping and maximum energy in spectral bands. While the higher bands overlapping can have an adverse effect on the unique spectral signature of the HSI channels [27]. Therefore, after optimizing the coefficients (W) and obtaining $\hat{\mathbf{Y}}_1$, in the second part of the IE unit, according to Schläpfer et al. [28], we apply SD method and correct the overlapping of each band of $\hat{\mathbf{Y}}_1$ by subtracting the weighted values of neighboring bands from the original band as the following:

$$\mathbf{L}_{i,\text{dec}} = (\mathbf{L}_i - \rho\mathbf{L}_{i+1} - \gamma\mathbf{L}_{i-1}) \tag{8}$$

where $\mathbf{L}_{i,\text{dec}}$ is deconvolved spectral band $\mathbf{L}_i$. $\mathbf{L}_{i+1}$ and $\mathbf{L}_{i-1}$ indicate right and left neighbors of the $\mathbf{L}_i$. ρ and γ are the weights learned by convolution layers. SD methods recover the original signal from the corrupted data and increase the signature of the sharp spectral features [27], [28]. Based on (8), we perform SD on each band of $\hat{\mathbf{Y}}_1$ to increase the sharp spectral features and decrease the bandwidth of each band of the IE unit output (called $\hat{\mathbf{Y}}_2$). Accordingly, by correcting the individual bands overlapping the HSI, we enhance the quality of the output image of the IE unit. This technique does not increase the spectral content of each voxel. However, it can avoid loss of the information of an HSI [28]. The simulation of the bands overlapping is shown in Fig. 3.

2) MCF Unit: In the second unit, we aim to extract spatial, spectral, and high-pass information in parallel based on $f_2(.)$ and $f_3(.)$ in (2) and then combine them to reach the HSI with full features. For that, we apply a series of convolutional layers

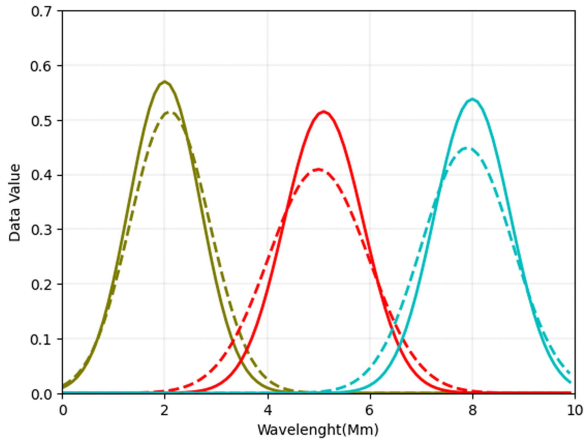


Fig. 3. Simulation of the effectiveness of bands overlapping reduction. We perform SD on each band of HSI to increase the sharp spectral effect and decrease the bandwidth of each band. Dash line and normal line represent their bands of HSI before bands deoverlapping and deconvolved spectral bands, respectively.

to explore nonlinear spatial–spectral features. After that, we combine features to generate richer spectral–spatial information with elementwise addition instead of nonlinear operation. This is done to avoid increasing computational complexity. In the spectral feature fusion block, we not only use $\mathbf{Y}$ to obtain the deep spectral features, but also we convert $\mathbf{P}$ into a cube with C layers to extract joint spatial–spectral information. To achieve this requirement, we learn the spectral response matrix $\mathbf{R}$ through the proposed network to convert $\mathbf{P}$ into HSI with C layers to reduce spectral distortion (see Fig. 2). Then spatial degradation $\mathbf{B}$ is applied to $\mathbf{R}^{\mathrm{T}}\mathbf{P}$ to downsample it into the same size as $\mathbf{Y}$. Moreover, we apply t times 1×1 and 3×3 convolution layers on the output and add the $\mathbf{Y}$ in each step as a residual manner to gain spectral features. Eventually, we use the 3×3 convolution layer with the Leaky-ReLU activation function to get high performance and smooth results. This process can be formulated as

$$\begin{aligned}\mathbf{F_{c1}} &= \sum_{i=1}^{t} C_i^{3\times3}(C_i^{1\times1}(\mathbf{B}\mathbf{R}^{\mathrm{T}}\mathbf{P})) + \mathbf{Y} \\ \mathbf{F_c} &= L_{\mathrm{ReLU}}(C^{3\times3}(\mathbf{F_{c1}}\mathbf{B}^{\mathrm{T}} + \mathbf{R}^{\mathrm{T}}\mathbf{P}))\end{aligned} \tag{9}$$

where C_i represents ith convolution layer in each iteration and $\mathbf{F_c}$ is the result of the spectral feature fusion block. Similarly, according to (9), to extract the spatial features in the spatial feature fusion block, the $\mathbf{B}^{\mathrm{T}}\mathbf{Y}$ is converted to the PAN layer through spectral response matrix $\mathbf{R}$ and then t times 1×1 and 3×3 convolution layers are applied with skip connection of the geometric information of the $\mathbf{P}$ to the output of each step. After repeating t times the process and converting the result to the cube form by $\mathbf{R}^{\mathrm{T}}$, the $\mathbf{B}^{\mathrm{T}}\mathbf{Y}$ is added band-by-band to inject the low-level information as shown in Fig. 2. We also apply the 3×3 convolution layer with the Leaky-ReLU activation function to gain spatial features called $\mathbf{F_t}$.

Besides, to exploit more accurate boundaries and edges, we gain the high-pass features $\mathbf{P_{hp}}$ by processing $\mathbf{P}$ as a guided image and appending the high-pass information in cube form into the $\mathbf{F_c}$ and $\mathbf{F_t}$. $\mathbf{P_{hp}}$ is obtained by subtracting the filtered image (by low-pass filter) from the original image. We believe that such issues degrade the performance of the existing models and quality of $\hat{\mathbf{X}}$. Hence, in a high-pass fusion block, we extract the boundary information of $\mathbf{P}$ and then convert it into cube form through spectral matrix $\mathbf{R}^{\mathrm{T}}$.

According to (9), in the same manner in spatial and spectral fusion block, we apply t times 1×1 and 3×3 convolution layers on $\mathbf{R}^{\mathrm{T}}\mathbf{P_{hp}}$ and append the injected spatial information of $\mathbf{P}$ to the results in each steps. In addition, to enhance the output HSI in high-pass feature fusion block, we execute the result via the convolution layer with kernel size 3×3 and Leaky-ReLU activation function and call it $\mathbf{F_h}$ as seen in Fig. 2. Finally, we aggregate the results of three blocks as $\mathbf{F_w} = \mathbf{F_h} + \mathbf{F_c} + \mathbf{F_t}$ and denoise them using the encoder–decoder model. We utilize U-net [29] as the regularization function with 1×1 convolution layers in an encoder and 3×3 deconvolution layer in the decoder part with Leaky-ReLU activation function to reconstruct the output $\hat{\mathbf{Y}}_3$. In the decoder part, the encoder's result of each convolution is concatenated to the decoder's deconvolution layer as follows:

$$\begin{cases} \mathbf{f_e}^i = L_{\mathrm{ReLU}}(W_{\mathrm{en}}^i * \mathbf{F_w}) & i = 1:k \\ \mathbf{f_d}^1 = L_{\mathrm{ReLU}}(W_{\mathrm{de}}^1 * \mathbf{f_e}^3) & i = 1 \\ \mathbf{f}^i = L_{\mathrm{ReLU}}\left(W_{\mathrm{de}}^i * \mathrm{Concat}\left(\mathbf{f_e}^{k+1-i}, \mathbf{f_d}^{i-1}\right)\right) & i \geq 1 \end{cases} \tag{10}$$

$$\hat{\mathbf{Y}}_3 = L_{\mathrm{ReLU}}(C^{3\times3}(\mathbf{f}^i)) \tag{11}$$

where W_{en} and W_{dn} indicate convolution and deconvolution layers and $k = 3$ is the number of iteration in each part.

3) HFA Unit: Finally, the third unit is HFA, which minimizes the (3) through the SGD algorithm to obtain multiscale spatial–spectral feature maps hierarchically. In the literature, some fusion methods attempted to extract the multiscale features hierarchically. For instance, Bandara and Patel [30] utilized a hierarchical network from low to high levels to extract the cross-features of PAN and LR-HSI. In this study, we apply a dual hierarchical multiscale network from low to high levels and then from high to low levels in an HFA unit. HFA includes multiple subnetworks to prepare up-to-down and down-to-up hierarchical multiscale feature maps (see Fig. 4). Each subnetwork contains three blocks: two residual feature extraction (residual block) and one self-attention (attention block).

The attention blocks obtain feature maps hierarchically from high to low levels of spectral voxels using a unique and standard feature extraction module (UCFE) based on the self-attention mechanism

$$\mathbf{h_U} = H_{\mathbf{U}}\left(\mathbf{Y}, \hat{\mathbf{Y}}_3\right) \tag{12}$$

where $H_{\mathbf{U}}$ represents the attention block and $\mathbf{h_U}$ denotes the result of the each $H_{\mathbf{U}}$. In the attention block, we apply the UCFE module in the self-attention mechanism to extract the features of bands containing the spectral information. The UCFE consists of a series of sparse coding-based modules to extract standard features [31]. In the original UCFE block, the input is subtracted from its feature response, and the original

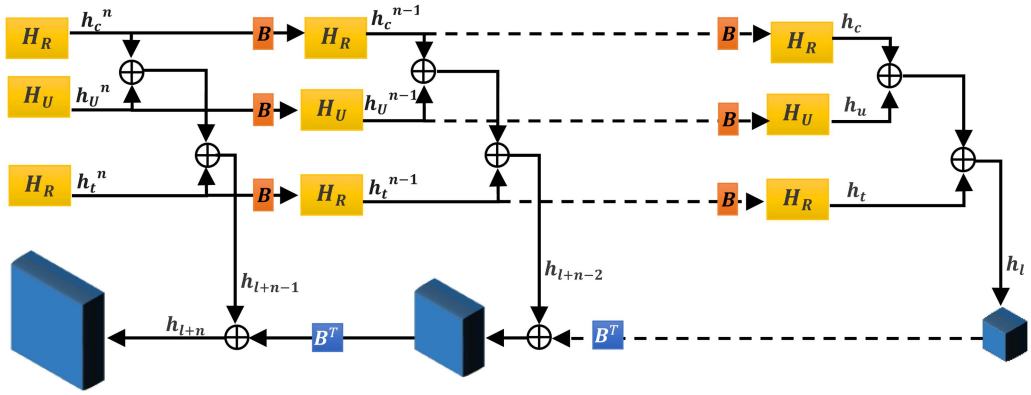


Fig. 4. Scheme of the HFA unit that creates multiscale feature maps from high to low levels via attention and residual blocks and then aggregates the extracted information hierarchically from low to high levels.

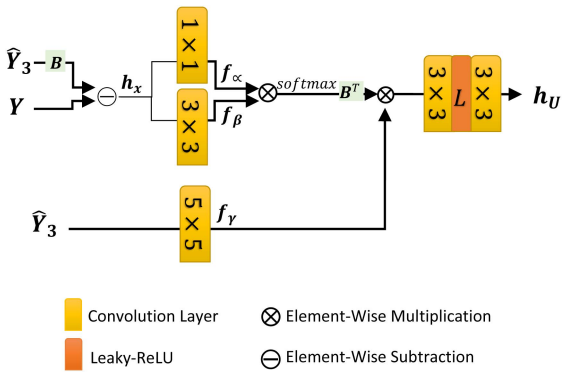


Fig. 5. Attention Block ($\mathbf{H_U}$) of subnetworks in HFA unit.

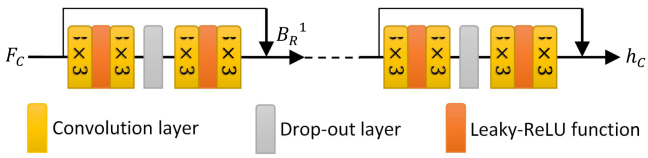


Fig. 6. Residual feature extraction block ($\mathbf{H_R}$) of subnetworks in HFA unit.

input is added to the result. On the other hand, self-attention extracts key feature maps from significant voxels of the hyperspectral bands. This mechanism [32] is one of the most influential parts of deep learning, which learns how to get over the encoder–decoder model's restriction by multiplying the encoder vectors with three matrices trained by the convolution layer. Therefore, we combine the UCFE with a self-attention mechanism to get standard and global spectral data of the input image (see Fig. 5). In the first sub-network, we find the information in $\mathbf{Y}$ which are not present in $\hat{\mathbf{Y}}_3$, $\mathbf{h_x} = \mathbf{Y} - \mathbf{B}\hat{\mathbf{Y}}_3$. Later on, we pass $\mathbf{h_x}$ and $\hat{\mathbf{Y}}_3$ through three convolution layers with different filter sizes to compute three feature maps $\mathbf{f}_\alpha = C_\alpha * \mathbf{h_x}$, $\mathbf{f}_\beta = C_\beta * \mathbf{h_x}$, and $\mathbf{f}_\gamma = C_\gamma * \mathbf{B}\hat{\mathbf{Y}}_3$. Then, the self-attention mechanism multiplies the $\mathbf{f}_\alpha$ by transposed $\mathbf{f}_\beta$ and performs the softmax operation on the result, then the upsampled result is multiplied by $\mathbf{f}_\gamma$. Afterward, we exploit features through the convolution filters with the Leaky-ReLU activation function and compute $\mathbf{h_U}$

$$\begin{cases} \mathbf{x_U} = \mathbf{B}^{\mathrm{T}}\mathrm{softmax}(\mathbf{h_x}^T C_\alpha C_\beta \mathbf{h_x})\mathbf{f}_\gamma \\ \mathbf{h_U} = C_1^{3\times3}(L_{\mathrm{Relu}}(C_0^{3\times3}(\mathbf{x_U}))) \end{cases} \tag{13}$$

C_α, C_β, and C_γ are $1\times1, 3\times3$, and 5×5 convolution layers to have different scale feature maps.

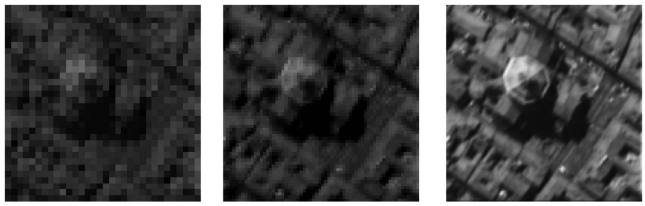
On the other hand, in residual blocks, we extract the spatial and spectral features of $\mathbf{F_c}$ and $\mathbf{F_t}$ hierarchically based on the residual mechanism in HFA unit

$$\begin{aligned} \mathbf{h_t} &= H_{\mathbf{R}}(\mathbf{F_t}) \\ \mathbf{h_c} &= H_{\mathbf{R}}(\mathbf{F_c}) \end{aligned} \tag{14}$$

where $H_{\mathbf{R}}$ indicates residual block, $\mathbf{h_c}$ is the output when $H_{\mathbf{R}}$ has $\mathbf{F_c}$ input and $\mathbf{h_t}$ represents the residual output when $\mathbf{F_t}$ is the input of residual block. To preserve more spatial–spectral information in a fused image, we apply a residual extraction block. We pass the $\mathbf{F_t}$ and $\mathbf{F_c}$ to the two residual blocks on the upper subnetwork separately to take advantage of the spatial features of $\mathbf{F_t}$ and spectral information of $\mathbf{F_c}$. The residual block includes multiple parts of convolution layers with the Leaky-ReLU as an activation function in the residual form to solve the fusion problem (see Fig. 6). ResNet accelerates learning with high performance in deep neural networks without eliminating gradient and degradation of accuracy [33]. In the residual block, we apply the skip connection to generate a residual stage and connect the input to the output. Also, each part of the residual block contains four convolution layers with a Leaky-ReLU activation function and a dropout layer, which can be depicted in Fig. 6

$$\begin{cases} \mathbf{E_R} = C_1^{3\times3}(L_{\mathrm{ReLU}}(C_0^{3\times3}(\mathbf{F_t} \text{ or } \mathbf{F_c}))) \\ \mathbf{B_R} = C_1^{3\times3}(L_{\mathrm{ReLU}}(C_0^{3\times3}(D_{\mathrm{out}}(\mathbf{E_R})))) \\ \mathbf{h_t} \text{ or } \mathbf{h_c} = \sum_{i=1}^{t} \mathbf{B_R}_i + \mathbf{B_R}_i \end{cases} \tag{15}$$

where $\mathbf{B_R}$ denotes the output of one part of the residual block, which is repeated t times. As mentioned, the HFA unit has two residual blocks, which in the first block, $\mathbf{B_{R_0}}$ is $\mathbf{F_t}$ and for second residual block, we set $\mathbf{B_{R_0}} = \mathbf{F_c}$. Essentially, we apply



(a) (b) (c)

Fig. 7. Visualization of the hierarchical feature maps from low to high levels by the HFA unit during the training phase. (a) $\mathbf{h_l}$. (b) $\mathbf{h_{l+1}}$. (c) $\mathbf{h_{l+n}}$.

the dropout function D_{out} through the residual phase between two convolutional blocks to enhance the robustness of the HS-McHF. We consider the probability r of the selected nodes to be ignored $r = 0.2$, and the rest of the nodes are preserved in our model by assuming $(1 - r) = 0.8$.

After computing the outputs of attention and residual blocks in the first stage, we transfer the downsampled outputs ($\mathbf{Bh_c}^n$, $\mathbf{Bh_U}^n$, and $\mathbf{Bh_t}^n$) to the next subnetwork. This process is iterated until the outputs are the same size as $\mathbf{Y}$. Then, in the final subnetwork, we gather all features from the three blocks ($\mathbf{h_c}, \mathbf{h_U}$, and $\mathbf{h_t}$) and call it $\mathbf{h_l}$. After that, we provide $\mathbf{B}^{\mathrm{T}}\mathbf{h_l}$ and append it to the summation of previous subnetwork outputs. This procedure is repeated from down to large scale at each level until the HSI result is the same size as $\mathbf{P}$ and obtain $\mathbf{h_{l+n}}$

$$\begin{cases} \mathbf{h_l} = \mathbf{h_c} + \mathbf{h_U} + \mathbf{h_t} \\ \mathbf{h_{l+n}} = \mathbf{h_c}^n + \mathbf{h_U}^n + \mathbf{h_t}^n + \uparrow \mathbf{h_{l+n-1}}. \end{cases} \tag{16}$$

Therefore, $\mathbf{h_l}$ is a feature-maps extracted from the nth scale of the last level, and $\mathbf{h_{l+n}}$ denotes the features at the first subnetworks. Fig. 7 demonstrates $\mathbf{h_l}$, $\mathbf{h_{l+1}}$, and $\mathbf{h_{l+n-1}}$ output images that were generated from the Pavia-Center dataset by capturing them at the same training iteration.

Finally, we obtain $\hat{\mathbf{X}}$ by collecting the extracted features from $\mathbf{h_{l+n}}$ through the two convolution layers and $\mathbf{B}^{\mathrm{T}}\mathbf{Y}$

$$\hat{\mathbf{X}} = C_1^{3\times3}(L_{\text{Relu}}(C_0^{3\times3}(\mathbf{h_{l+n}}))) + \mathbf{B}^{\mathrm{T}}\mathbf{Y} \tag{17}$$

where L_{Relu} represents Leaky-ReLU activation function.

C. Loss Function

The motivation of HS-McHF is to minimize the loss function by reducing the difference between ground truth and the hypersharpened image. Since the HSI has geometric and spectral information, we employ the specific loss function in the designed network to obtain the spatial, spectral, and structural features from input images. The mean squared error (MSE) calculates the spatial difference as

$$\text{MSE}(\mathbf{X}, \hat{\mathbf{X}}) = ||\mathbf{X} - f_\Theta(\mathbf{Y}, \mathbf{P})||_F^2. \tag{18}$$

As the similarity increases, the MSE reduces and has lower values. Because we intend to minimize the loss function, the MSE measure is used to calculate the spatial dissimilarity between the output of the network and the desired image. Note that the MSE computes only a spatial constraint between two images. To improve the performance of the loss function, we append the structural criteria to our loss function. Structural similarity index measure (SSIM) is an efficient measure to calculate the resemblance rate between two images by processing significant data if there exists identical spatial information through the HSI. Here, SSIM can be defined and simplified as follows [34], [35]:

$$\text{SSIM}(\mathbf{X}, \hat{\mathbf{X}}) = \sum_{i=1}^{C} \frac{4\mu_{\mathbf{X}}^i \mu_{f_\Theta(\mathbf{Y},\mathbf{P})^i} \sigma_{\mathbf{X}^i f_\Theta(\mathbf{Y},\mathbf{P})^i}}{\left(\mu_i^2 + \mu_{f_\Theta(\mathbf{Y},\mathbf{P})^i}^2\right)\left(\sigma_{\mathbf{X}}^{i\,2} + \sigma_{f_\Theta(\mathbf{Y},\mathbf{P})}^{i\,2}\right)} \tag{19}$$

where $\mu_{\mathbf{X}}$ and $\mu_{f_\Theta(\mathbf{Y},\mathbf{P})}$ represent the mean and $\sigma_{\mathbf{X}}$ and $\sigma_{f_\Theta(\mathbf{Y},\mathbf{P})}$ indicate the variance between $\mathbf{X}$ and $f_\Theta(\mathbf{Y}, \mathbf{P})$, respectively. $\sigma_{\mathbf{X}f_\Theta(\mathbf{Y},\mathbf{P})}$ is covariance of $\mathbf{X}$ and $f_\Theta(\mathbf{Y}, \mathbf{P})$. To enhance the performance, we suppose the spectral index is a third-party equation used to append the loss formula. Thus, in addition to spatial and structural constraints, the loss function includes a spectral constraint. The spectral angle mapper (SAM) is a criterion for comparing two HSIs [36], which can be expressed as

$$\text{SAM}(\mathbf{X}, \hat{\mathbf{X}}) = \cos^{-1} \sum_{i=1}^{C} \frac{< \mathbf{X}^i . f_\Theta(\mathbf{Y}, \mathbf{P})^i >}{||\mathbf{X}^i||_2 \, ||f_\Theta(\mathbf{Y}, \mathbf{P})^i||_2} \tag{20}$$

where $< . >$ denotes the inner product. If (20) is equal to "0," the best spectral quality can be achieved.

In our proposed network, the loss function can be denoted as $l(\Theta)$ between $f_\Theta(\mathbf{Y}, \mathbf{P})$ and $\mathbf{X}$

$$l(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \frac{\text{MSE}(\mathbf{X}^i, f_\Theta(\mathbf{Y}^i, \mathbf{P}^i)) \times \text{SAM}(\mathbf{X}^i, f_\Theta(\mathbf{Y}^i, \mathbf{P}^i))}{\text{SSIM}(\mathbf{X}^i, f_\Theta(\mathbf{Y}^i, \mathbf{P}^i))} \tag{21}$$

where N represents the number of images in training data and Θ denotes the parameters of the HS-McHF. Since the primary purpose behind the application of $l(\Theta)$ (21) is to preserve the spatial, spectral, and structural features, we propose $l_{\text{spectral}}(\Theta)$ and apply interpolation function B to $f_\Theta(\mathbf{Y}, \mathbf{P})$ to minimize the spectral error between $\mathbf{Y}$ and downsampled output of the network. For analyzing the geometric similarity, the spectral response matrix $\mathbf{R}$ is trained through the proposed network to convert the $\hat{\mathbf{X}}$ to the PAN layer to generate the PAN version of the output and compare the spatial differences through the l_{spatial}. Finally, the loss function $l_{\text{final}}(\Theta)$ can be computed as the sum of three proposed losses, which is defined as follows:

$$l_{\text{final}}(\Theta) = l(\Theta) + l_{\text{spectral}}(\Theta) + l_{\text{spatial}}(\Theta). \tag{22}$$

IV. Experimental Analysis

In this section, we evaluate the HS-McHF by performing it on three well-known hyperspectral datasets, such as the Chikusei [37], Salient [38], and Pavia-Center [39]. Then, we compare experimental results with six existing pansharpening models. To verify the proposed model, we implemented it on a computer based on Windows 10 with Intel(R) CPU E5-2620, NVIDIA Quadro 2000 GPU, and 32 GB RAM. Then, we train the HS-McHF using the Pytorch framework in Python 3.7 via SGD

TABLE I
DETAILED INFORMATION OF THE PAVIA-CENTER, CHIKUSEI, AND SALIENT DATASETS

Dataset	No. of images	No. of channels	Range of spectrum (nm)	No. of training	No. of validation	No. of test
Pavia-Center	1	102	430×860	20	5	15
Chikusei	1	128	363×1018	40	5	15
Salient	60	81	380×720	40	5	15

Fig. 8. First row: RGB images from the Pavia-Center dataset at bands 24, 44, and 60. Second row: RGB images from the Chikusei dataset at bands 30, 60, and 80. Third row: RGB images from the Salient dataset at bands 14, 40, and 60.

optimizer with a learning rate of 0.001. The training phase with validation is performed, and the PSNR function in 700 epochs is shown in Fig. 12. Table I explains the details of datasets in our experiments.

A. Datasets Definition

In this section, we conduct extensive experiments with the HS-McHF on three publicly available hyperspectral datasets: Chikusei,[1] Salient,[2] and Pavia-Center[3] remote sensing datasets. As listed in Table I, the Salient dataset includes 60 photos from HS with a spatial size of 768 × 1024, each with 81 spectral channels in the visible spectrum 380 nm × 720 nm. We utilize the first 40 pictures for training HS-McHF and the last 20 images for validating and testing. We suppose the validation set is a separate part from the training images and apply this procedure to validate the HS-McHF and prevent the overfitting problem during the training. The test set is also different from both the training and validation set. Pavia-Center is a 1096 × 1096 pixels HSI with 102 spectral bands achieved over Pavia, Northern Italy. The spatial resolution is 1.3 m, and the spectral range is between 430 and 860 nm. The HSI is cropped without overlapping into 40 patches to train the network with 25 images and test the model with 15 HSIs. For convenience, we excluded channels with low SNR and focused on 80 frequency bands. The Chikusei is the airborne hyperspectral dataset acquired from Headwall imaging sensor over agricultural and urban areas in Chikusei, Japan, that contains HS images with a spatial size of 2517 × 2335 and 128 spectral bands in the range of 363–1018 nm. The HSI is cropped without overlapping into 40 patches 128 × 128 to train and validate the network with ten images and ten images to test the McHF-H. We omitted channels that had a low SNR and concentrated on analyzing 80 frequency bands. We simulate the training data according to Wald protocol. For three datasets, to simplify this process, we assume the spatial size of the PAN and the target image for the three datasets of 128 × 128. Then, we exploit a scaling factor of four to create the LR-HSI version with the size of 32 × 32 by the downsampling operator. Fig. 8 depicts the details of the RGB images in the Pavia-Center, Chikusei, and the Salient datasets.

B. Performance Evaluation

In this section, we evaluate the efficiency of HS-McHF, considering the spatial, structural, and spectral metrics. To analyze spatial equality, we apply five spatial metrics, such as MSE ERGAS, PSNR, and UQI, which are common in the literature. Moreover, we utilize the SAM to evaluate the performance of the proposed model and other state-of-the-art approaches in terms of spectral fidelity. In addition, we consider the SSIM as a structural and spatial quality measurement. Besides, we calculate the MSE rate to measure the features of ground-truth and fused images using (18), in which the best value of MSE is close to 0. Next, erreur relative globale adimensionnelle de synthese (ERGAS) [40] denotes the overall quality of the fused image [41], i.e., the optimal value for ERGSA is 0

$$\mathrm{ERGAS}(\mathbf{F},\mathbf{G}) = \frac{100}{d^2}\sqrt{\frac{1}{C}\sum\nolimits_{i=1}^{C}\frac{\mathrm{MSE}(\mathbf{F}^i,\mathbf{G}^i)}{\mu_{\mathbf{F}}^i}} \tag{23}$$

where $\mathbf{F}$ and $\mathbf{G}$ represent hypersharpened and grand truth images, respectively, and d is the spatial ratio between the PAN and the HSI. In addition, the peak signal-to-noise ratio (PSNR) defines the spatial symmetry between the fused and target pictures. The higher the value of PSNR implies a more significant similarity rate

$$\mathrm{PSNR}(\mathbf{F},\mathbf{G}) = \frac{1}{C}\sum_{i=1}^{C} 20\log_{10}\frac{\mathrm{MAX}_{\mathbf{F}}^i}{\mathrm{MSE}(\mathbf{F}^i,\mathbf{G}^i)} \tag{24}$$

where $\mathrm{MAX}_{\mathbf{F}}$ is the maximum value in $\mathbf{F}$ in the ith band of $\mathbf{F}$. Also, the universal quality image (UQI) index [42] computes the transformation rate of features from the hypersharpened

[1][Online]. Available: https://www.sal.t.u-tokyo.ac.jp/hyperdata.
[2][Online]. Available: https://github.com/gistairc/HS-SOD.
[3][Online]. Available: https://www.ehu.eus/ccwintco.

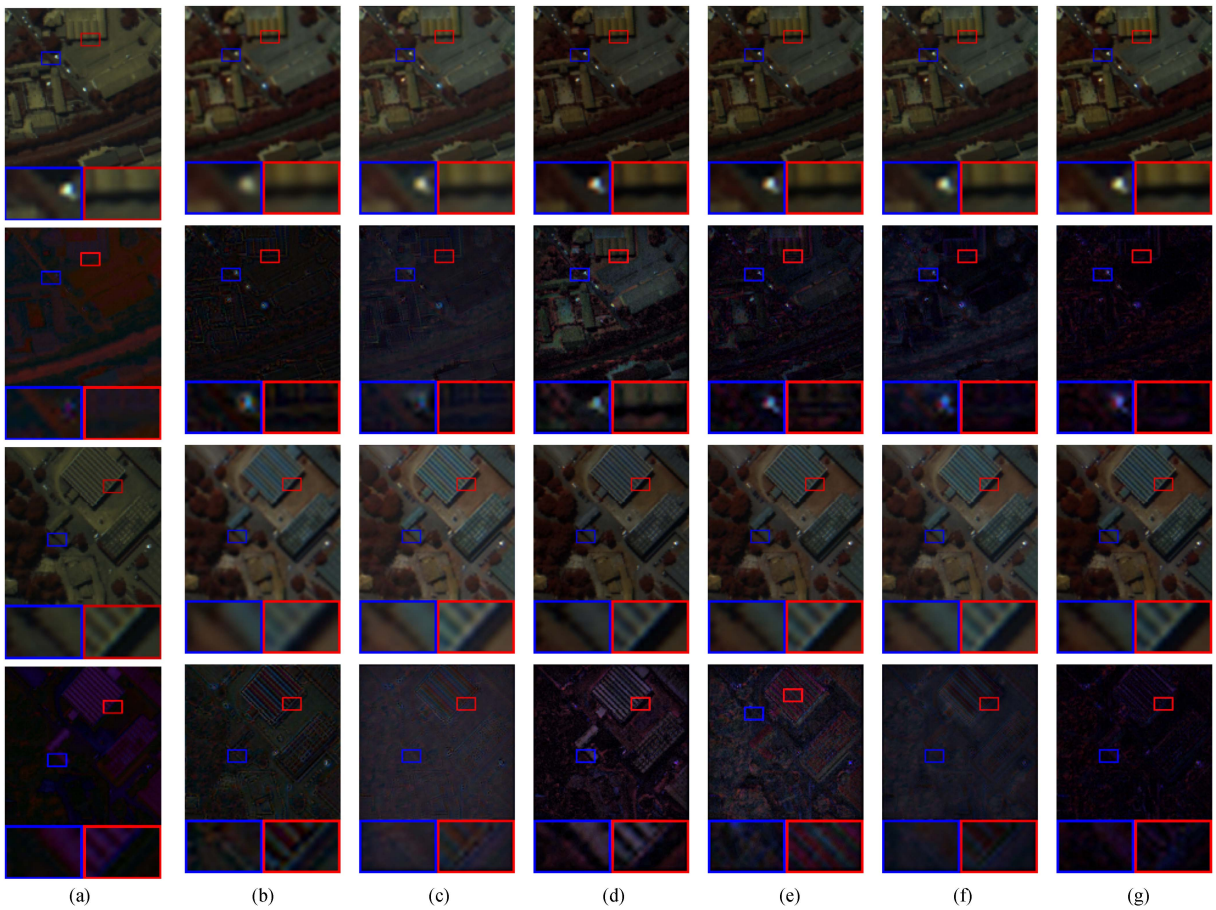


Fig. 9. Qualitative results on 50th band on Pavia-Center dataset. First and third rows: two reconstructed images with two comparison areas were upscaled in four times for clarity. Second and fourth rows: corresponding error. (a) PGCU. (b) DRPNN. (c) GDD. (d) Gppnn. (e) HyperPNN. (f) DHP-DARN. (g) HS-McHF.

image to the source picture for all spectral bands, i.e., when UQI approaches 1, the network output is equivalent to the desired picture

$$\mathrm{UQI}(\mathbf{F},\mathbf{G}) = \frac{1}{C}\sum_{i=1}^{C}\frac{4\sigma_{\mathbf{F}^i\mathbf{G}^i}\mathbf{F}^i\mathbf{G}^i}{(\sigma^2_{\mathbf{F}^i}+\sigma^2_{\mathbf{G}^i})(\mathbf{F}^{i2}+\mathbf{G}^{i2})} \tag{25}$$

where σ represents the standard deviation function. For spectral comparison, we utilize the SAM as defined in (20), which computes the spectral fidelity between the output of our model and target picture (e.g., the best value of SAM is 0). Eventually, we obtain the SSIM by employing (19) to measure the structural symmetry between the reconstructed image and ground-truth.

C. Comparison With State-of-The-Art Models

We investigate the performance of the HS-McHF by comparing with six state-of-the-art fusion-based models: DRPNN[4] [43], GDD[5] [19], Gppnn[6] [44], DHP-DARN[7] [15], HyperPNN[8] [14], and PGCU[9] [45]. We have chosen these six models because they deployed deep learning for hypersharpening applications. We conduct a comparative analysis considering the average of obtained results from the test images by experimenting with them on the Chikusei dataset (see Table III), Salient database (see Table IV) and Pavia-Center dataset (see Table II). For a fair comparison, we retrain all methods on our dataset and perform the three phases of training, validating, and testing similarly for each network.

D. Comparative Analysis

In this section, we verify the efficiency of the HS-McHF by comparing our results with six state-of-the-art FhSR models. As depicted in Tables II–IV, the HS-McHF provides superior performance compared to the other evaluated methods in most areas. Table II states that HS-McHF can preserve spatial, spectral, and structural information with a specific loss function after performing it on the Pavia-Center dataset. The reason for the satisfactory performance of our model is that we applied a dual deep optimization network with approximate feature maps transformation between the LR-HSI and PAN with a skip

[4][Online]. Available: https://github.com/matciotola.
[5][Online]. Available: https://github.com/tuezato/guided-deep-decoder.
[6][Online]. Available: https://github.com/shuangxu96.
[7][Online]. Available: https://github.com/yxzheng24.
[8][Online]. Available: https://github.com/wgcban/DIP-HyperKite/tree/main.
[9][Online] Available: https://github.com/Zeyu-Zhu/PGCU.

Fig. 10. Qualitative results on 24th band on Salient dataset. First and third rows: two reconstructed images with two comparison areas were upscaled four times for clarity. Second and fourth rows: corresponding error. (a) PGCU. (b) DRPNN. (c) GDD. (d) Gppnn. (e) HyperPNN. (f) DHP-DARN. (g) HS-McHF.

TABLE II
AVERAGE QUANTITATIVE RESULTS OF THE PROPOSED METHOD AGAINST FUSION METHODS ON THE PAVIA-CENTER DATASET

Methods	SSIM	UQI	PSNR	SAM	MSE	ERGAS
Best value	1	1	$+\infty$	0	0	0
GDD	0.738	0.778	23.729	0.199	0.0065	8.633
DRPNN	0.803	0.849	23.701	0.156	0.0046	7.099
Gppnn	0.894	**0.963**	28.876	0.142	0.0021	5.690
DHP-DARN	0.877	0.913	29.060	0.149	0.0020	5.127
PGCU	0.896	0.864	29.767	0.160	0.0028	5.158
HyperPNN	0.897	0.905	29.047	0.149	0.0017	5.022
HS-McHF	**0.903**	0.953	**31.180**	**0.133**	**0.0010**	**4.514**

The bold values indicate the best performance.

TABLE III
AVERAGE QUANTITATIVE RESULTS OF THE PROPOSED METHOD AGAINST FUSION METHODS ON THE CHIKUSEI DATASET

Methods	SSIM	UQI	PSNR	SAM	MSE	ERGAS
Best value	1	1	$+\infty$	0	0	0
GDD	0.801	0.714	27.089	0.152	0.0039	8.491
DRPNN	0.888	0.762	28.721	0.139	0.0030	7.557
Gppnn	0.931	0.929	30.004	0.072	0.0016	7.050
DHP-DARN	**0.940**	0.911	31.409	0.068	0.0012	6.923
PGCU	0.938	0.895	30.981	0.081	0.0024	7.013
HyperPNN	0.922	0.867	30.614	0.092	0.0030	7.586
HS-McHF	0.930	**0.934**	**31.499**	**0.061**	**0.0011**	**6.589**

The bold values indicate the best performance.

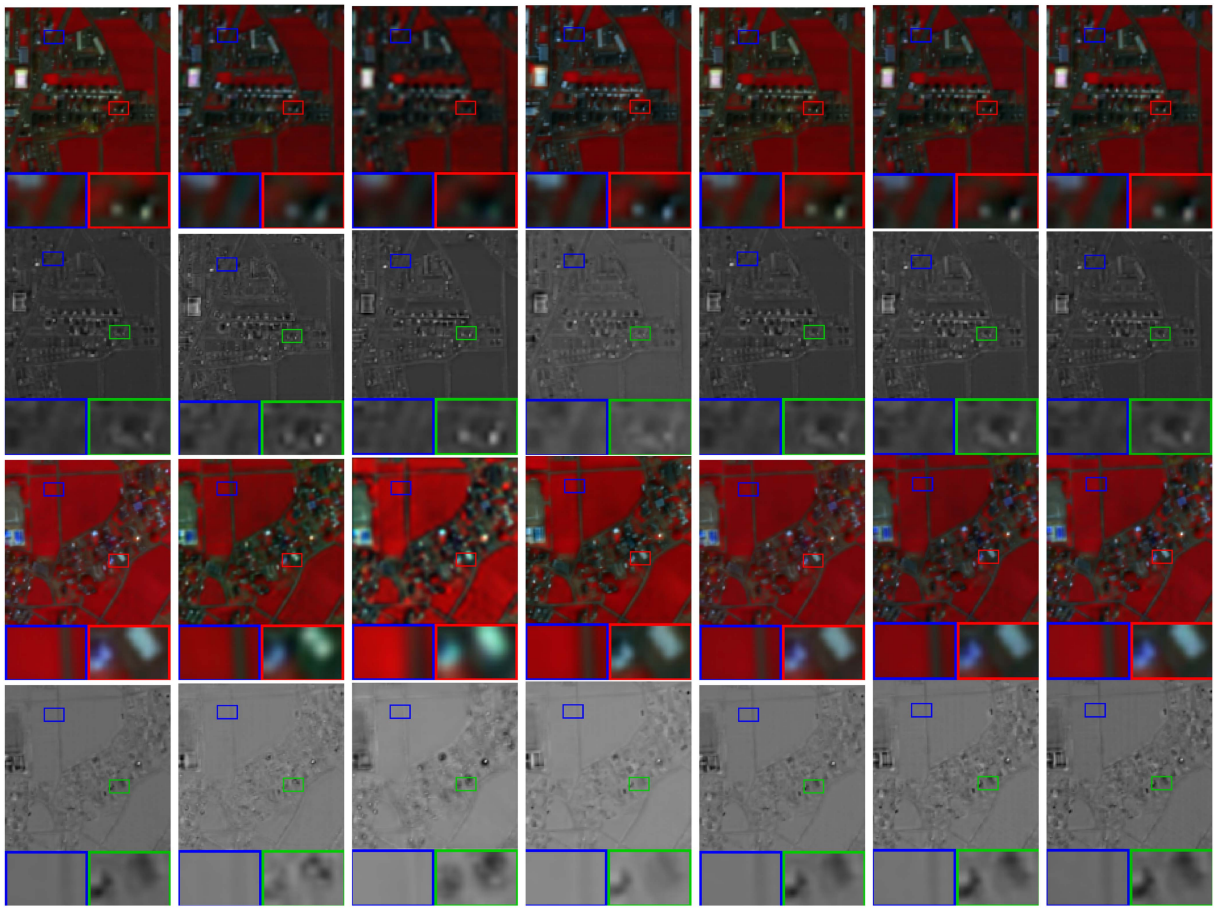


Fig. 11. Qualitative results on Chikusei dataset. First and third rows: two reconstructed images with two comparison areas were upscaled four times for clarity. Second and fourth rows: corresponding error. (a) PGCU. (b) DRPNN. (c) GDD. (d) Gppnn. (e) HyperPNN. (f) DHP-DARN. (g) HS-McHF.

TABLE IV
AVERAGE QUANTITATIVE RESULTS OF THE PROPOSED METHOD AGAINST FUSION METHODS ON THE SALIENT DATASET

Methods	SSIM	UQI	PSNR	SAM	MSE	ERGAS
Best value	1	1	$+\infty$	0	0	0
GDD	0.591	0.777	16.879	0.336	0.0233	10.543
DRPNN	0.701	0.768	20.756	0.182	0.0081	7.191
Gppnn	0.756	0.897	20.742	0.238	0.0071	7.245
DHP-DARN	0.540	0.781	18.258	0.256	0.0148	9.869
PGCU	0.796	0.877	22.809	0.159	0.0051	8.611
HyperPNN	0.762	0.796	21.392	0.198	0.0087	7.219
HS-McHF	**0.814**	**0.904**	**24.668**	**0.141**	**0.0044**	**6.024**

The bold values indicate the best performance.

connection that can overcome the vanishing gradient. Our model can achieve superior spatial performance in MSE, ERGAS and PSNR because of using the MCF unit to combine the high-pass information with the spatial–spectral features, and have less shift and details change because of applying a denoising module after combining features. The best value in SAM and SSIM indicators in the Pavia-Center dataset means less distance and spectral distortion between fused and target images in our model. Moreover, the new loss function of the HS-McHF could benefit from the geometric and spectral aspects simultaneously. Due to reducing the bands overlapping, the proposed HS-McHF has less spectral distortion among compression models and can avoid losing the information of a reconstructed HSI. Furthermore, due to extracting the feature maps via the self-attention mechanism, the quality of reconstructed images considerably improved. Eventually, combining the loss function and neural networks through our model offers a deeper learning process; hence, it affords efficient performance, particularly in terms of SSIM, PANR, and SAM indexes on the Pavia-Center database.

In our experiment on the Chikusei dataset, we evaluated the performance of our proposed method on remote sensed HSIs, which include both agricultural and urban areas. Table III

TABLE V
NUMBER OF NETWORK PARAMETERS IN SALIENT AND PAVIA-CENTTER DATASETS

	DRPNN	GDD	Gppnn	DHP-DARN	HyperPNN	PGCU	HS-McHF
Salient	2.1M	0.9M	2M	0.3M	0.1M	1.2M	0.1M
Pavia-Center	2.3M	1.4M	3M	0.4M	0.1M	1.2M	0.2M

TABLE VI
RUNNING TIME (IN SECONDS) OF A TRAINING EPOCH AND TEST PHASE IN NETWORKS

	Datasets	DRPNN	GDD	Gppnn	DHPDARN	Hyper-PNN	PGCU	HS-McHF
Train	Salient	3.94	2.93	4.07	2.38	0.78	1.62	3.07
	PaviaCenter	3.91	2.984	4.416	2.448	0.840	2.05	3.57
Test	Salient	0.069	0.0086	0.0094	0.0068	0.0028	1.60	0.0085
	PaviaCenter	0.098	0.0185	0.0101	0.0055	0.0032	2.01	0.0096

TABLE VII
AVERAGE QUANTITATIVE RESULTS OF THE PROPOSED METHOD ON THE PAVIA-CENTER DATASET

IE	MCF	HFA	SSIM↑	UQI↑	PSNR↑	SAM↓	MSE↓	ERGAS↓
√	×	×	0.492	0.619	21.223	0.209	0.0094	8.507
×	√	×	0.649	0.705	23.501	**0.134**	0.0069	7.001
×	×	√	0.767	0.798	23.774	0.151	0.0061	7.921
√	√	×	0.775	0.896	24.041	0.148	0.0065	7.042
√	×	√	0.792	0.875	23.269	0.152	0.0059	6.689
×	√	√	0.806	0.885	24.984	0.158	**0.0044**	6.178
√	√	√	**0.814**	**0.904**	**25.668**	0.141	**0.0044**	**6.024**

*Note that the symbols √ and × represent the evaluation of the proposed method with and without the inclusion of IE, MCF, and HFA units, respectively.
The bold values indicate the best performance.

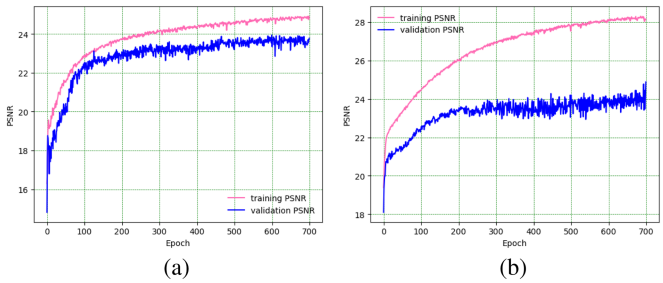


Fig. 12. Average PSNR after 700 epochs in the training and validation phase of the proposed method for (a) Salient dataset and (b) Pavia-Center dataset.

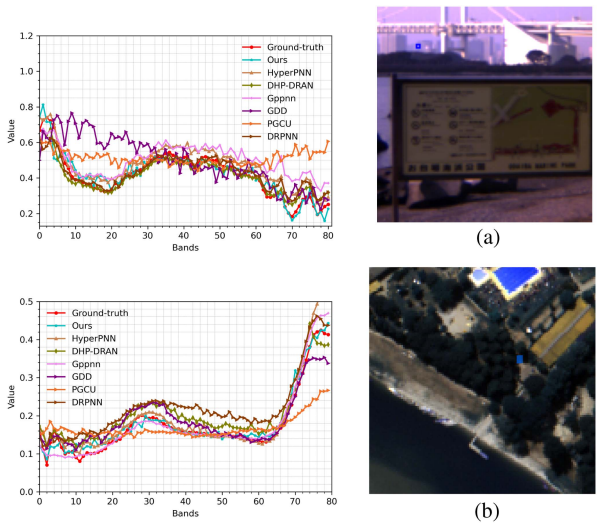


Fig. 13. Spectral signature of the specific pixel for the output image of the proposed and comparative methods compared with the target image in Salient and Pavia-Center datasets. (a) Salient dataset. (b) Pavia-Center dataset.

displays the average quantitative results of our HS-McHF on the test images compared to all other methods. It is evident that our HSRnet outperforms the other compared methods on most metrics. To analyze the efficiency of the HS-McHF on another dataset, we performed our model and the selected approaches on the Salient dataset (see Table IV). Our experimental results showed that the HS-McHF gives superior results considering all the evaluation metrics. As depicted in Tables VI and V, the training process of our network with the Salient HSI containing a high number of spectral layers increased the time complexity. Although our model has a deep network, it can eliminate overfitting by regularizing data in the training phase and using a dropout layer in the residual part. Network lightning is the reason for reduction the number of network parameters and improving the robustness and efficiency of the HS-McHF network. Fig. 12 shows PSNR values of the proposed model after 700 epochs on Salient, and Pavia-Center datasets. HyperPNN, because of the light network, is less involved in overfitting, and it can obtain better results on the Salient and Pavia-Center than other compared models.

After performing the HS-McHF, Figs. 9–11 illustrate the experimental results, including two examples of hypersharpened images from each dataset separately versus six other compared approaches (see the first and third rows). The second and fourth rows show the absolute difference between the ground-truth

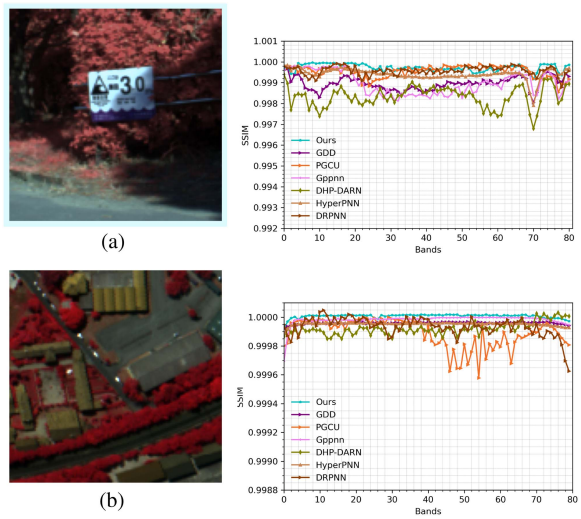


Fig. 14. Average SSIM curve of all bands for the comparative methods compared with proposed method for Salient, and Pavia-Center datasets. (a) Salient dataset. (b) Pavia-Center dataset.

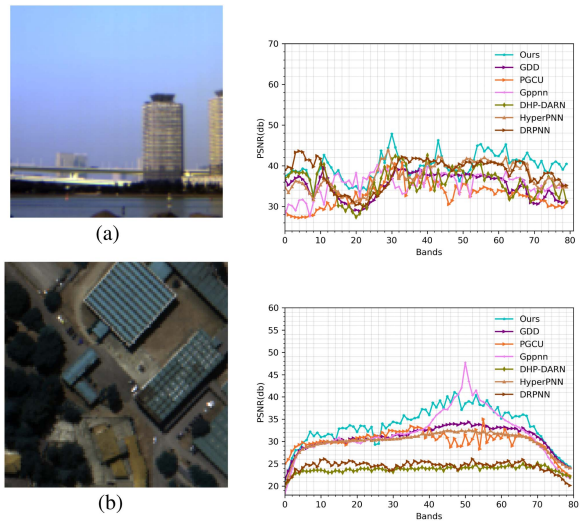


Fig. 15. Average PSNR curve for the fused HSI of all bands in Salient and Pavia-Center datasets. (a) Salient dataset. (b) Pavia-Center dataset.

and the reconstructed images. The output of our model on the Salient datasets is slightly brighter than the ground-truth, but the edges and boundaries are much closer to the desired image. The HS-McHF also achieves minimal absolute differences, which indicates less structural and spectral distortion. On the other hand, the output image of the Gppnn network is sharper than the ground-truth; hence, the high-pass features, such as edges, seem unnatural, especially on the Salient dataset.

Moreover, to analyze the spectral efficiency of the proposed method, the spectral signature of the fused HSI of the HS-McHF model compared with other methods is computed and shown in Fig. 13. We compare the spectral value of a pixel of all channels in the target and fused HSI in Salient, and Pavia-Center datasets. Our proposed model has the most similar conduct with the target HSI, which means less spectral distortion and better fusion performance. To further compare the spectral distortion and spatial preservation of the HS-McHF versus other fusion methods, we present the average PSNR and SSIM score in Figs. 15 and 14. The highest results of the PSNR and SSIM in most bands of the reconstructed HSI demonstrate the minimum shift and less spectral distortion between the fused and desired image of the proposed method.

As listed in Tables VI and V, we have evaluated the complexity of our model considering the number of neural network parameters and running time, respectively, i.e., the time required for an epoch of training duration and testing phase. Moreover, it can be seen from Table VI that our model requires more training and testing time because we applied deeper networks and additional convolutional layers, which leads to high performance and better results. Although the proposed model utilizes extra layers, it needs fewer parameters. This is because we employ the dropout technique to remove redundant nodes from the deep residual block. Hence, the HS-McHF is not as computationally complex as the Gppnn, GDD, and DHP-DARN.

V. Discussion

In order to prove the effectiveness of our proposed method, we conducted a number of experiments. These experiments were designed to validate our approach and ensure that it meets the standards of quality.

A. Analysis of the Efficiency of Three Units IE, MCF, and HFA

To assess the effectiveness of our model, we evaluated it with and without each of the three units on the Pavia-Center and Salient dataset. The outcomes are shown in Table VII, and they indicate that all three units, IE, MCF, and HFA, work together in a complementary manner. MCF and HFA outperform IE in all metrics, and in terms of UQI, MSE, and PSNR, we observed that unit HFA performs better than MCF. However, HS-McHF can achieve an improvement when all three components are present. As mentioned in Section III-B3, we demonstrate the three steps in the HFA unit by visualizing $\mathbf{h_l}$, $\mathbf{h_{l+1}}$, and $\mathbf{h_{l+n-1}}$. Fig. 7 shows that the final fusion result ($\mathbf{h_{l+n-1}}$) contains fewer artifacts and more structural information.

B. Analysis of Proposed Loss Function Effect

As mentioned, we suggested a new loss function in our model by combining the MSE, SSIM, and SAM criteria to obtain spatial, spectral, and structural information simultaneously from the fused image. However, the state-of-the-art models exploited the MSE as a performance evaluation metric while comparing the fused output and desired image [43], [46], [47], [48], [49]. Therefore, to demonstrate the effectiveness of our loss function, we implemented the HS-McHF using the MSE criteria, and the experimental results are shown in Table VIII. The results indicate that HS-McHF with the new loss function is more efficient and has an acceptable rate for most criteria. Hence, it can be remarked that the suggested loss function can impact the HS-McHF to preserve the geometric information and spectral features.

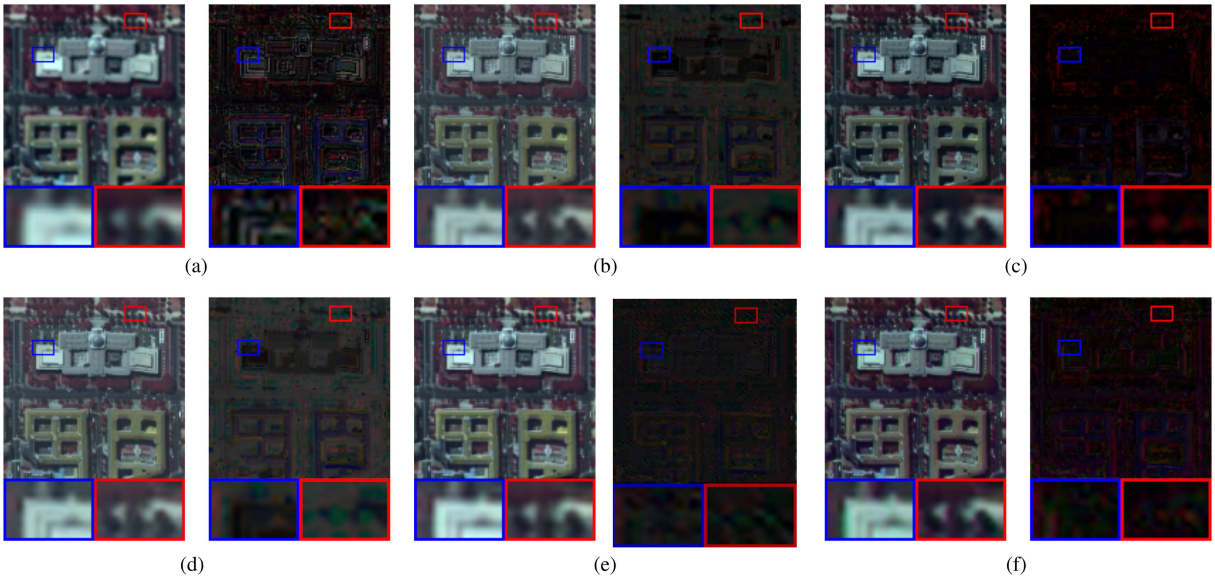


Fig. 16. Illustration of RGB results of our model compared to other networks considering three bands 5, 15, and 25 on the Washington DC mall dataset. The (a), (b), and (c) images demonstrate the reconstructed images, which contain pairwise comparison areas where up-scaled four times, and the (d), (e), and (f) images are corresponding errors. (a) DRPNN. (b) DHP-DARN. (c) Gppnn. (d) HyperPNN. (e) PGCU. (f) HS-McHF.

TABLE VIII
PERFORMANCE OF THE SUGGESTED LOSS FUNCTION VERSUS THE MSE CRITERIA USAGE IN PAVIA-CENTER AND SALIENT IN HS-MCHF

Dataset	Metrics	With spatial loss	With MSE loss
Pavia-Center	SSIM↑	0.903	0.857
	UQI↑	0.953	0.9041
	PSNR↑	31.180	29.257
	SAM↓	0.133	0.148
	MSE↓	0.0010	0.0024
	ERGAS↓	4.514	6.029
Salient	SSIM↑	0.872	0.730
	UQI↑	0.927	0.862
	PSNR↑	25.596	21.725
	SAM↓	0.120	0.239
	MSE↓	0.0036	0.0062
	ERGAS↓	5.744	7.760

TABLE IX
AVERAGE QUANTITATIVE RESULTS OF THE PROPOSED METHOD AGAINST FUSION METHODS ON REAL DATASET DC MALL

Methods	SSIM	UQI	PSNR	SAM	MSE	ERGAS
Best value	1	1	$+\infty$	0	0	0
DRPNN	0.731	0.884	19.854	0.120	0.0023	7.148
Gppnn	0.845	0.965	21.108	0.097	**0.0017**	5.647
DHP-DARN	**0.862**	0.930	19.692	**0.096**	0.0018	6.232
HyperPNN	0.828	0.871	18.499	0.118	**0.0017**	6.831
PGCU	0.814	0.902	20.198	0.127	0.0027	6.081
HS-McHF	0.856	**0.966**	**21.168**	0.099	**0.0017**	**5.616**

The bold values indicate the best performance.

C. Analysis of the Applicability of the Proposed Method on Real Data

We use the DC-mall[10] dataset to evaluate the efficiency of the proposed FhSR on real data. The HSI was provided over the Washington DC mall area on August 23, 1995, which contains a spatial size of 307 × 1280 and 191 spectral bands in the range of 400 nm × 2500 nm. the HSI is cropped without overlapping into 20 patches 128 × 128 to train and validate the network with 15 images and five images to test the HS-McHF. We considered the 128 × 128 HSIs as a reference image and created the LR-HSI version with the size of 32 × 32 by the downsampling operator with a factor of four. We select the first 40 bands for HSI and one at 10 for an assistive image. Fig. 16 shows the RGB images of the real dataset and the hypersharped results with corresponding errors for the HS-McHF and other fusion methods. It can be

[10][Online]. Available: https://engineering.purdue.edu.

seen the reconstructed HSI of our network is much closer to the desired image. In addition, we conducted experiments on the DC-mall dataset and analyzed the performance of HS-McHF versus the state-of-the-art models considering the standard evaluation metrics. As listed in the Table IX, our HS-McHF approach outperforms the evaluated methods.

VI. Conclusion

This article proposed a novel FhSR model called HS-McHF with three units and quality enhancement blocks, which utilizes spatial–spectral fusion design and deep multilayer features aggregation network for fusing LR-HSI and PAN images to produce an HR-HSI. In this study, to gain feature maps and deep details of significant voxels of all channels, we use a self-attention module in a new design with the high-pass geometric injection to combine spectral and spatial information. Instead of the traditional loss function, we combine spectral, spatial, and structural criteria to preserve geometric details and spectral information. Then, we compared the HS-McHF with the state-of-the-art methods. Our experimental results on three well-known hyperspectral datasets confirmed that HS-McHF offers superior efficiency considering the standard evaluation metrics.

References

[1] GISGeography, "Multispectral vs hyperspectral imagery explained," Mar. 2024. [Online]. Available: https://gisgeography.com/multispectral-vs-hyperspectral-imagery-explained

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

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

[4] H. Wang, C. Wang, and Y. Yuan, "Asymmetric dual-direction quasi-recursive network for single hyperspectral image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 11, pp. 6331–6346, Nov. 2023.

[5] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

[6] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[7] P. Liu, L. Xiao, and T. Li, "A variational pan-sharpening method based on spatial fractional-order geometry and spectral–spatial low-rank priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1788–1802, Mar. 2018.

[8] M. Kremezi et al., "Pansharpening PRISMA data for marine plastic litter detection using plastic indexes," *IEEE Access*, vol. 9, pp. 61955–61971, 2021.

[9] S. Luo, S. Zhou, Y. Feng, and J. Xie, "Pansharpening via unsupervised convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4295–4310, 2020.

[10] 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.

[11] L. Zhang, W. Li, L. Shen, and D. Lei, "Target-adaptive CNN-based pansharpening," *Int. J. Remote Sens.*, vol. 41, pp. 7201–7216, 2020.

[12] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, 2016, Art. no. 594.

[13] D. Shen, J. Liu, Z. Xiao, J. Yang, and L. Xiao, "A twice optimizing net with matrix decomposition for hyperspectral and multispectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4095–4110, 2020.

[14] 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.

[15] Y. Zheng, J. Li, Y. Li, J. Guo, X. Wu, and J. Chanussot, "Hyperspectral pansharpening using deep prior and dual attention residual network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8059–8076, Nov. 2020.

[16] G. Shim, J. Park, and I. S. Kweon, "Robust reference-based super-resolution with similarity-aware deformable convolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8422–8431.

[17] S. Xu, J. Zhang, Z. Zhao, K. Sun, J. Liu, and C. Zhang, "Deep gradient projection networks for pan-sharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 1366–1375.

[18] J. Wang, Z. Shao, X. Huang, T. Lu, and R. Zhang, "A dual-path fusion network for pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5403214.

[19] T. Uezato, D. Hong, N. Yokoya, and W. He, "Guided deep decoder: Unsupervised image pair fusion," in *Proc. Comput. Vis.–Lecture Notes Comput. Sci.*, 2020, pp. 87–102.

[20] X. Shuang et al., "Deep convolutional sparse coding network for pansharpening with guidance of side information," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2021, pp. 1–6, doi: 10.1109/ICME51207.2021.9428131.

[21] X. Wu, J. Feng, R. Shang, X. Zhang, and L. Jiao, "Multiobjective guided divide-and-conquer network for hyperspectral pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5525317.

[22] L. He, D. Xi, J. Li, H. Lai, A. Plaza, and J. Chanussot, "Dynamic hyperspectral pansharpening CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5504819.

[23] W. Fan, F. Liu, and J. Li, "Pansharpening via multiscale embedding and dual attention transformers," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 2705–2717, 2024.

[24] H. Yin, "PSCSC-net: A deep coupled convolutional sparse coding network for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5402016.

[25] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 228–236, Jan. 2008.

[26] G. Vivone, "Robust band-dependent spatial-detail approaches for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6421–6433, Sep. 2019.

[27] A. S. Kumar, T. Radhika, V. Keerthi, D. S. Jain, V. K. Dadhwal, and A. K. Kumar, "Spectral deconvolution and non-overlap bands sampling for ims-1 hyperspectral imager data," *J. Indian Soc. Remote Sens.*, vol. 42, no. 2, pp. 429–434, 2013.

[28] D. Schläpfer, A. Börner, and M. Schaepman, "The potential of spectral resampling techniques for the simulation of apex imagery based on aviris data," in *Proc. 8th JPL Airborne Earth Sci. Workshop*, 1999, pp. 377–384.

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Interv.–MICCAI Springer, Lecture Notes Comput. Sci.*, 2015, vol. 9351, pp. 234–241.

[30] W. G. C. Bandara and V. M. Patel, "Hypertransformer: A textural and spectral feature fusion transformer for pansharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1757–1767.

[31] X. Deng and P. L. Dragotti, "Deep convolutional neural network for multi-modal image restoration and fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3333–3348, Oct. 2021.

[32] M.-T. Luong, H. Pham, and C. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1412–1421.

[33] F. Rousseau, L. Drumetz, and R. Fablet, "Residual networks as flows of diffeomorphisms," *J. Math. Imag. Vis.*, vol. 62, pp. 365–375, 2020.

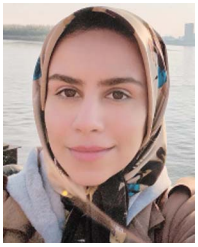
[34] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 14, pp. 600–612, Apr. 2004.

[35] P. Ndajah, H. Kikuchi, M. Yukawa, H. Watanabe, and S. Muramatsu, "SSIM image quality metric for denoised images," in *Proc. Int. Conf. Visual., Imag. Simul.*, 2010, pp. 53–57.

[36] F. Kruse et al., "The spectral image processing system (sips)–interactive visualization and analysis of imaging spectrometer data," in *Proc. Int. Conf. Visual., Imag. Simul.*, 1993, vol. 44, no. 2, pp. 145–163,.

[37] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over chikusei," Space Appl. Lab., Univ. Tokyo, Tokyo, Japan, Tech. Rep. SAL-2016-05-27, May 2016. [Online]. Available: http://park.itc.u-tokyo.ac.jp/sal/hyperdata/TechRepSAL20160527.pdf

[38] N. imamoğlu et al., "Hyperspectral image dataset for benchmarking on salient object detection," in *Proc. 10th Int. Conf. Qual. Multimedia Exp.*, 2018, pp. 1–3.
[39] F. Dell'Acqua, P. Gamba, A. Ferrari, J. Palmason, J. Benediktsson, and K. Arnason, "Exploiting spectral and spatial information in hyperspectral urban data with high resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 322–326, Oct. 2004.
[40] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?," in *Proc. 3rd Conf. Fusion Earth Data: Merging Point Meas. Raster Maps Remotely Sens. Images*, Jan. 2000, pp. 99–103. [Online]. Available: https://api.semanticscholar.org/CorpusID:17796189
[41] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, Nov. 2018.
[42] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[43] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.
[44] S. Xu, J. Zhang, Z. Zhao, K. Sun, J. Liu, and C. Zhang, "Deep gradient projection networks for pan-sharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 1366–1375.
[45] Z. Zhu, X. Cao, M. Zhou, J. Huang, and D. Meng, "Probability-based global cross-modal upsampling for pansharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 14039–14048.
[46] A. Azarang, H. E. Manoochehri, and N. Kehtarnavaz, "Convolutional autoencoder-based multispectral image fusion," *IEEE Access*, vol. 7, pp. 35673–35683, 2019.
[47] W. Huang, Y. Zhang, J. Zhang, and Y. Zheng, "Convolutional neural network for pansharpening with spatial structure enhancement operator," *Remote Sens.*, vol. 13, no. 20, 2021, Art. no. 4062. [Online]. Available: https://www.mdpi.com/2072-4292/13/20/4062
[48] W. Wang, Z. Zhou, H. Liu, and G. Xie, "MSDRN: Pansharpening of multispectral images via multi-scale deep residual network," *Remote Sens.*, vol. 13, no. 6, 2021, Art. no. 1200. [Online]. Available: https://www.mdpi.com/2072-4292/13/6/1200
[49] L. Zhang, W. Li, L. Shen, and D. Lei, "Multilevel dense neural network for pan-sharpening," *Int. J. Remote Sens.*, vol. 41, pp. 7201–7216, 2020.



Zeinab Dehghan (Student Member, IEEE) received the M.Sc. degree in artificial intelligence from Shiraz University, Shiraz, Iran, in 2017. She is currently working toward the Ph.D. degree with the Laboratory of Computer Vision and Image Analysis, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China.

Her research interests include hyperspectral image superresolution, image fusion, feature selection, and deep learning.

Jingxiang Yang (Member, IEEE) received the joint Ph.D. degree in control theory and control engineering and engineering science from Northwestern Polytechnical University, Xi'an, China, and Vrije Universiteit Brussel, Brussels, Belgium, in 2019.

He is currently a Lecturer with the Nanjing University of Science and Technology, Nanjing, China. His research interests include deep learning and its applications in hyperspectral image processing.

Milad Taleby Ahvanooey (Senior Member, IEEE) received the Ph.D. degree in computer engineering with honors from the Nanjing University of Science and Technology, Nanjing, China, in 2019.

He is currently an Assistant Professor and ULAM Scientist with the Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland. From 2022 to 2024, he was a Senior Research Fellow with Nanyang Technological University, Singapore. In addition, he was a Researcher with the Institute of Multimedia Information Processing, IM School, Nanjing University, Nanjing, China, where he attained his Postdoctoral fellowship from 2019 to 2022. His research interests include the application of AI in cybersecurity and digital trust areas, and he holds three patent applications.

Dr. Taleby Ahvanooey was a recipient of one of the **CSC Elite Outstanding award**s for Ph.D. students at NJUST in 2019 and the ULAM NAWA scholarship in 2022. He serves as an Editorial Board Member of Frontiers in Big Data (2023–present), Cyber Security Insights Magazine (2022– present), Cryptography Journal (2021–2023), as a reviewer board of ACM TOMM, Computers in Human Behavior (Elsevier) as well as a Technical Program Committee (TPC) Member of several International conferences, such as SecureComm EAI 2022, ACM ARES 2021, 2022, 2023, and 2024, IEEE ICME 2021, 2022, and 2023, MIC-Security 2021, IEEE MIC-Computing 2021, and IEEE ICDIS 2023 (43 d Series).

Abdolraheem Khader (Member, IEEE) received the B.S., M.Sc., and Ph.D. degrees in computer science from Karary University, Omdurman, Sudan; Sudan University of Science and Technology, Khartoum, Sudan; and Nanjing University of Science and Technology, Nanjing, China, in 2011, 2014, and 2023, respectively.

He is currently a Postdoctor in computer science and technology with the Nanjing University of Science and Technology. His research interests include the areas of deep learning and hyperspectral image superresolution.

Liang Xiao (Senior Member, IEEE) received the B.S. degree in applied mathematics and the Ph.D. degree in computer science from the Nanjing University of Science and Technology (NJUST), Nanjing, China, in 1999 and 2004, respectively.

From 2009 to 2010, he was a Postdoctoral Fellow with the Rensselaer Polytechnic Institute, Troy, NY, USA. Since 2014, he has been the Deputy Director with the Jiangsu Key Laboratory of Spectral Imaging Intelligent Perception, Nanjing, China. He was the Second Director of the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, NJUST, where he is currently a Professor with the School of Computer Science. He has authored or coauthored more than 100 international journal articles including IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. His research interests include inverse problems in image processing, computer vision, and image understanding, pattern recognition, and remote sensing.