





# A Two-Stream Multiscale Deep Learning Architecture for Pan-Sharpener

Jie Wei, Yang Xu , *Member, IEEE*, Wanting Cai, Zebin Wu , *Senior Member, IEEE*, Jocelyn Chanussot , *Fellow, IEEE*, and Zhihui Wei , *Member, IEEE*

**Abstract**—Pan-sharpening, which fuses the high-resolution panchromatic (PAN) image and the low-resolution multispectral image (MSI), is a hot topic in remote sensing. Recently, deep learning technology has been successfully applied in pan-sharpening. However, the existing methods ignore that the MSI and PAN image are at different resolutions and use the same networks to extract features of the two images. To address this problem, we propose a two-stream deep learning architecture, called coupled multiscale convolutional neural network, for pan-sharpening. The proposed network has three components, feature extraction subnetworks, fusion layer, and super-resolution subnetwork. In the feature extraction subnetworks, two subnetworks are used to extract the features of the MSI and PAN image separately. Different sizes of convolutional kernels are used in the first layers due to the different spatial resolutions. Thus, the source images are mapped to the similar scale. Then a multiscale asymmetric convolution factorization is used to extract features at different scales. In the fusion layer, the two feature extraction subnetworks are coupled. Features at the same scale are first summed, and then the features of all scales are concatenated as one feature map. At last, a super-resolution subnetwork is used to generate the high-resolution MSI. Experimental results on both synthetic and real data sets demonstrate that the proposed method outperforms the other state-of-the-art pan-sharpening methods.

**Index Terms**—Convolutional neural network (CNN), image fusion, multiscale, pan-sharpening.

Manuscript received June 28, 2020; revised August 16, 2020 and August 24, 2020; accepted August 26, 2020. Date of publication September 2, 2020; date of current version September 24, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61701238, Grant 61772274, Grant 61471199, Grant 61976117, Grant 11431015, Grant 61501241, Grant 61671243, and Grant 61802190, in part by the Jiangsu Provincial Natural Science Foundation of China under Grant BK20170858, Grant BK20180018, and Grant BK20191409, in part by the Fundamental Research Funds for the Central Universities under Grant 30919011234, Grant 30917015104, Grant 30919011103, and Grant 30919011402, in part by the China Postdoctoral Science Foundation under Grant 2017M611814 and Grant 2018T110502, in part by the Jiangsu Province Postdoctoral Science Foundation under Grant 1701148B, and in part by the Key Projects of University Natural Science Fund of Jiangsu Province, China under Grant 19KJA360001. (*Corresponding authors: Yang Xu; Zhihui Wei.*)

Jie Wei, Wanting Cai, and Zhihui Wei are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: sunnywj@163.com; dudu118273@163.com; gswei@njust.edu.cn).

Yang Xu and Zebin Wu are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China, and also with the GIPSA-lab, CNRS, Grenoble INP, Université Grenoble Alpes, 38000 Grenoble, France (e-mail: xuyangth90@njust.edu.cn; zebin.wu@gmail.com).

Jocelyn Chanussot is with Inria, CNRS, Grenoble INP, LJK, Université Grenoble Alpes, 38000 Grenoble, France (e-mail: jocelyn.chanussot@gipsa-lab.grenoble-inp.fr).

Digital Object Identifier 10.1109/JSTARS.2020.3021074

## I. INTRODUCTION

PAN-SHARPENING [1]–[3] refers to the fusion of a low spatial resolution multispectral image (LR-MSI) and a high spatial resolution panchromatic (PAN) image to obtain a high spatial resolution MSI (HR-MSI). The demand for pan-sharpening is increasing since most remote sensing applications [4]–[7] require images to be high in both spatial and spectral resolutions.

Over the last 30 years, many methods have been proposed for pan-sharpening and there are a number of ways to classify these methods. The latest way is to classify pan-sharpening methods into four categories:

- 1) component substitution (CS) based methods;
- 2) multiresolution analysis (MRA) based methods;
- 3) variational optimization (VO) based methods;
- 4) deep learning-based methods.

The CS-based methods and MRA-based methods are the first proposed families. In the CS-based methods, the MSI is projected into a new space using spectral transform. Then, one of the components is replaced by the PAN image. Finally, inverse projection is performed to obtain the HR-MSI. Representative methods include intensity hue saturation (IHS) [8], Gram–Schmidt (GS) [9], principal component analysis [10], etc. The MRA-based methods extract the spatial structures of PAN image and inject the extracted spatial information into the LR-MSI. Wavelet transform [11] and Laplacian pyramids [12] are usually used to decompose the image into multiscale channels and extract high frequency information. Filter estimation [13] first estimates the filter that models the blur between MSI and PAN image. Then, a high-pass modulation based model is used to inject the details. The VO-based methods [14]–[16] consider the fusion process as an ill-posed inverse problem which is composed of the LR-MSI fidelity term and the PAN fidelity term. Some additional priors, such as total variation [17] and nonlocal similarities [18], can be introduced in the objective function.

Deep learning [19] has achieved great success in computer vision, image processing, natural language processing, etc. More recently, deep learning-based inverse problems [20], [21] have been a frontier of image processing. Deep learning-based methods, especially convolutional neural networks (CNNs) [22], have become the most advanced approaches in pan-sharpening. In [23], the authors proposed a deep neural network (DNN) pan-sharpening method by stacking the sparse denoising autoencoder which can be considered as the first method applying deep learning in pan-sharpening. Then, pan-sharpening based on CNN

(PNN) [24] was proposed and it achieved promising results. In the PNN method, the LR-MSI is first upsampled and interpolated and then stacked with the PAN image to form the input of the neural networks. The architecture of the neural network is borrowed from the super-resolution CNN (SRCNN) [25] used in single image super-resolution. Later, Yuan *et al.* [26] proposed a multiscale and multidepth CNN (MSDCNN) architecture for pan-sharpening. In this method, two subnetworks are introduced and a multiscale feature extraction block is proposed in one of the subnetworks. The produced HR-MSI performs well in spatial structure preserving due to the introduced multiscale feature extraction.

Then, in [27], the authors proposed a two-stream fusion network (TFNet) that fused the PAN image and MSI in the feature level. Two CNNs are used to extract the features of PAN and LR-MSI separately and the extracted features are fused by concatenating in the fusion neural network. Yang *et al.* [28] proposed PanNet which trained the network in the high-pass domain instead of the image domain. So it is possible for the PanNet to generalize to other satellites by using the domain specific neural network. Shen *et al.* [29] proposed to combine deep learning and a variational model together for pan-sharpening. A deep residual CNN is used to learn the map from PAN image's gradient map and LR-MSI to the HR-MSI's gradient map. Then the learned gradient prior is imposed in the variational model. Further, the authors of PNN improved the PNN and proposed a target-adaptive CNN method [30]. In this method, a fine-tuning step is performed on the target data using the pretrained CNN and then the fine-tuned CNN is used as the final CNN for pan-sharpening. In [31], the authors proposed a remote sensing image fusion neural network (RSIFNN) which was also consisting of two feature extraction neural networks. In [32], a two-step MSI fusion method is proposed which estimates the spatial information and compensates the spectral information separately. In [33], a MSI fusion method based on convolutional autoencoder is proposed. The MSI is enhanced band by band with the trained neural network. Thus, the spatial details are well preserved.

To prevent the spectral distortion, Liu *et al.* [34] combined a spectral discrimination-based detail injection model and a shallow-deep convolutional network. To address the short of ground truth issue, Ma *et al.* [35] proposed an unsupervised pan-sharpening method in which a spectral discriminator and a spatial discriminator are used to preserve the spectral and spatial information. Further, deep learning methods have been extended to hyperspectral pan-sharpening. Xie *et al.* [36] proposed the 3-D generative adversarial networks in hyperspectral pan-sharpening. A loss function is designed to comprehensively consider global constraint, spectral constraint, and spatial constraint. In addition, Zheng *et al.* [37] proposed an end-end deep learning-based network for hyperspectral image super-resolution through a multipath strategy. It can well keep spectral consistency and enhance the spatial texture information simultaneously.

From the existing deep learning based pan-sharpening methods, it can be concluded that deep learning-based pan-sharpening methods are consisting of three components: feature extraction

subnetwork, fusion layer, and super-resolution subnetwork. According to the position of the fusion layer, these methods can be classified into two categories. The first category is that the fusion layer is located in the input layer and feature extraction subnetwork is abandoned, such as PNN, MSDCNN, and PanNet. The second category conducts the fusion step in the hidden layer, such as TFNet and RSIFNN. In the second category, the networks before fusion step are considered as feature extraction subnetworks for LR-MSI and PAN image separately. Since feature extraction is contained in the network and the weights are updated by end-to-end training, the second category performs better. However, the feature extraction subnetworks of LR-MSI and PAN image are the same and the fusion layer usually uses concatenating operator to fuse different features. In fact, LR-MSI and PAN image reveal the spatial structures at different scales and similar feature extraction subnetworks also extract features at different scales. Therefore, the fusion layer is limited by the concatenation and cannot recover the spatial structure at all scales.

To overcome this problem, we propose a two-stream deep learning architecture, called coupled multiscale CNN (CMC) for pan-sharpening. In the proposed neural network, two separated feature extraction networks are used to extract the deep features from LR-MSI and PAN image. Different kernel sizes are used in the networks such that the extracted features are first mapped to the same scale. Besides, a novel multiscale asymmetric convolution factorization CNN is proposed to extract the multiscale features of the LR-MSI and PAN image separately. Then a novel multiscale fusion layer is designed to couple the two feature extraction subnetworks. It first fuses the features of the same scales and then concatenates the features in all scales. Thus, the features of LR-MSI and PAN image are fused at the same scale. At last, the super-resolution subnetwork is used to recover the HR-MSI. In the proposed method, the multiscale information of the observed images are accurately extracted and characterized by the asymmetric convolution factorization CNN. So the feature maps fed into the super-resolution layer contain more latent spectral-spatial information of the HR-MSI than the traditional SRCNN. Thus, an improved super-resolution result can be expected. Experiments on both synthetic and real data sets show that the proposed method outperforms the state-of-the-art pan-sharpening methods.

Overall, the highlights of the article can be concluded as follows.

First, a two-stream deep learning architecture is proposed for pan-sharpening. The MSI and PAN image are fed into the neural network separately and different convolutional kernels are used to map the input images into the same scale.

Second, a multiscale asymmetric convolution factorization CNN is proposed to extract the multiscale features of the input features. The proposed multiscale feature extraction method has less parameters and better performance.

Third, a novel feature fusion layer is proposed to make full use of the input images. The features are first fused at the same scale and then all the features at different scales are concatenated. In this way, the spatial structures are preserved without breaking the spectral structure.

The remainder of the article is organized as follows. The basic knowledge for pan-sharpening and deep learning are presented in Section II. Section III introduces the detailed architecture of the proposed CMC. Experimental results and comparisons are shown in Section IV. The conclusion is drawn in Section V.

## II. DEEP LEARNING FOR PAN-SHARPENING

In this section, we will formulate the deep learning-based pan-sharpening methods and present some basic knowledge of deep learning.

In the pan-sharpening problem, a pair of images  $\mathbf{Y} \in \mathbb{R}^{M \times N}$  and  $\mathbf{Z} \in \mathbb{R}^{m \times n \times B}$  are given. Here  $\mathbf{Y}$  represents the PAN image and  $\mathbf{Z}$  represents the LR-MSI which has been preinterpolated using a simple interpolating method. Our goal is to recover the HR-MSI  $\mathbf{X} \in \mathbb{R}^{M \times N \times B}$ . Note here that  $M > m, N > n$ ; so there are more parameters to be estimated and this is an ill-posed inverse problem.

The deep learning based methods learns a nonlinear mapping relationship from the inputs  $(\mathbf{Y}, \mathbf{Z})$  to the outputs  $\mathbf{X}$  by minimizing the objective function

$$\mathcal{L} = \|f(\mathbf{Y}, \mathbf{Z}) - \mathbf{X}\|_F^2 \quad (1)$$

where  $f(\cdot)$  represents the nonlinear mapping achieved by neural network. DNNs and CNNs are two commonly used neural networks for nonlinear mapping. In this article, we will use CNN as our basic architecture for pan-sharpening. So we will introduce some basic knowledge in the following.

CNN [38], [39] is the most popular neural network in deep learning and has achieved great success in computer vision and image processing. In DNN [40], fully connected layers are used to connect the input and the output. Thus, the parameters in DNN are numerous, which requires a large amount of training samples. CNN relieves this problem by local receptive fields and shared weights. In CNN, each neuron is connected to neurons in a small subset of its previous layer and the weights of the convolutional kernel remains the same when the same feature map is generated in the next layer. As a result, the number of parameters in the neural network decreases dramatically and the network is shift, scale, and distortion invariant. In addition, the correlations in the spatial dimensions are considered since the input images are not vectorized.

In CNN-based pan-sharpening methods, a super-resolution subnetwork is located between the output and the input. The input of the super-resolution subnetwork denoted by  $\mathbf{S}_0 \in \mathbb{R}^{M_s \times N_s \times D_s}$  can be the low-resolution image or the feature maps.  $\mathbf{S}_0^{(i)}$  denotes the  $i$ th feature map. Then the convolution over the input feature maps can be expressed as

$$\mathbf{S}_1^{(j)} = h(\mathbf{w}_j \odot \mathbf{S}_0 + b_j) \quad (2)$$

where  $\mathbf{w}_j \in \mathbb{R}^{k_1 \times k_2 \times D_s}$ ,  $j = 1, 2, \dots, s_1$  is a convolutional kernel.  $k_1$  and  $k_2$  represent the kernel size,  $s_1$  is the number of convolution kernels, and  $\odot$  denotes the 3-D convolution operator.  $b_j$  is a bias term.  $h(\cdot)$  is the nonlinear active function. A commonly used active function is the rectified linear unit (ReLU) function which can be expressed as

$$h(x) = \max\{0, x\}. \quad (3)$$

We can apply the convolution operator multiple times by stacking the convolutional layers with different kernel sizes and numbers. Usually, three layers are enough to achieve satisfying results, such as SRCNN, PNN, etc.

## III. PROPOSED COUPLED MULTISCALE CNN ARCHITECTURE

### A. Motivation

In real applications, the LR-MSI and the PAN image are the only two source images fed into the neural network. However, the LR-MSI and the PAN image are obtained by different sensors and show information at different resolutions in the spatial domain. The existing CNN-based methods such as PNN and MSDCNN directly stack the PAN image and LR-MSI (usually enlarged to the desired size by simple interpolation methods) together for further processing. In these approaches, the features of different scales are treated equally, which will decrease the reconstruction quality. Later, RSIFNN and TFNet are proposed to solve this problem by introducing feature extraction subnetworks before the fusion of the two source images. In these methods, the subnetworks sharing the same architecture are used to extract the LR-MSI and PAN image's features. Improved results can be obtained. However, since the spatial scales of the source images are different, the extracted features produced by the same architecture are also at different scales. Directly fusing features at different scales is not suitable. To overcome this problem, different architectures should be used in the feature extraction subnetwork to align and extract the features at a similar scale.

In addition, the multiscale feature extraction can improve the reconstruction accuracy and robustness. Since we have two source images, the architecture of the fusion layer should be designed to take the difference between different scales into consideration. Features of the same scale in the two images should be merged. To this end, we propose to first fuse the features of the same scale in the fusion layer and then fuse the features of all scales. By this way, the spectral structure of LR-MSI will be preserved when fusing the spatial details of the PAN image.

### B. Overview of the Proposed CMC Architecture

The proposed CMC architecture can be divided into three components: feature extraction subnetworks, fusion layer, and super-resolution subnetwork. The feature extraction component contains two subnetworks: LR-MSI subnetwork and PAN subnetwork which process the LR-MSI and PAN image separately. The outputs of the feature extraction subnetworks are the multiscale features of the input LR-MSI and PAN image. Then, in the fusion layer, the two feature subnetworks are coupled and the extracted multiscale features are fused together as the input of the super-resolution subnetwork. The complete architecture of our proposed pan-sharpening architecture is shown in Fig. 1.

### C. Feature Extraction Subnetworks

In the proposed CMC architecture, the input is composed of a LR-MSI and a PAN image which will be fed into the LR-MSI



subnetwork and PAN subnetwork. In the first layer of the LR-MSI subnetwork, we use CNN with kernel size  $W_m \times W_m \times B$  followed by ReLu activation function to extract the features. Analogously, we add the convolutional layer in the first layer of PAN subnetwork, but the size of the convolutional kernel is  $W_p \times W_p \times 1$ . Since we want to fuse the LR-MSI and the PAN image in a multiscale fashion, the extracted features before multiscale feature extraction layers should show the spatial information at a similar scale. The PAN image and the LR-MSI have different spatial resolutions. If the same size kernels are used in the first layer, the extracted features are at different scales. Therefore, the kernel size in the first layer of the two feature extraction subnetworks should be different, which means  $W_m \neq W_p$ . In addition, the spatial resolution of the PAN image is higher than the LR-MSI and the LR-MSI is first enlarged by simple interpolation methods. So the local neighborhood in the PAN image contains more spatial information than the same local neighborhood of LR-MSI. In order to contain the same amount of spatial information, the kernel size in the first layer of the LR-MSI subnetwork should be larger than that of the PAN subnetwork. Thus, we have  $W_m > W_p$ . The number of convolutional kernels in the first layers of the two subnetworks are the same for further progress.

After the first convolutional layer, the feature maps in the two subnetworks are considered to show the features in the same scale. As demonstrated in [26], features in different scales correspond to different spatial structures. So the multiscale feature extraction can improve the reconstruction accuracy and robustness in different scenes. In [26], a multiscale feature extraction block is used where different kernel sizes are used in the same layer. However, it increases the number of parameters in the neural network. Inspired by the well-known Inception V3 architecture, we propose a multiscale asymmetric convolution factorization method to extract the multiscale features. Instead of using  $n \times n$  convolutional kernels, we use the  $1 \times n$  followed by  $n \times 1$  convolution. A graphic comparison of the tradition multiscale CNN and that the proposed multiscale asymmetric CNN is shown in Fig. 2. It can be observed that the number of layers of the network increases and that the nonlinear mapping is enhanced in the network. On the contrary, the number of parameters is dramatically reduced. Therefore, different from the method in [41], we use asymmetric convolution factorization for different convolutional kernel sizes in the same layer which can extract rich spatial information.

#### D. Fusion Layer

In both the LR-MSI subnetwork and the PAN subnetwork, the multiscale asymmetric convolution factorization is used to extract features at different scales. So the fusion of the extracted multiscale features is also an important factor. Thanks to the different kernel sizes used in the first convolutional layer, the feature maps fed into the multiscale asymmetric convolutional layer can be regarded as being in the same scale in both the LR-MSI subnetwork and the PAN subnetwork. Thus, it is reasonable to assume that the extracted features by the same kernel size are also corresponding to the same scale in the MSI and PAN image.

Since the PAN image contains high frequency information and the LR-MSI contains the spectral information, we first sum the features extracted by the same kernel size together as the features corresponding to this scale. The differences between the proposed method and the other multiscale methods are that we use two subnetworks to extract the multiscale features separately and fuse the multiscale features according to their scale. So the fused features contain rich spatial information without breaking the spectral structure. After obtaining the fused features in every scale, we concatenate these features along the spectral mode to form the output of the fusion layer. Thus, the two feature extraction subnetworks are coupled and the output of the fusion layer is the features maps containing rich multiscale information.

#### E. Super-Resolution Subnetwork

The super-resolution subnetwork used in the proposed method is inspired by PNN. PNN has been proven to be efficient for the super-resolution problem. In the proposed method, we have more convolutional kernels in the first layer of the network than PNN and SRCNN. This is because we have more features in the input layer. More convolutional kernels are necessary to reconstruct the HR-MSI.

#### F. Training Process

In the proposed CMC method, all the parameters are trained by minimizing the loss between the reconstructed HR-HSI and the corresponding ground truth images. The mean squared error is used to measure the loss function as follows:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n \|\mathbf{X}_i - f(\mathbf{Y}_i, \mathbf{Z}_i)\|_F^2. \quad (4)$$

Here  $\mathbf{X}_i \in \mathbb{R}^{P_1 \times P_2 \times B}$  represents the  $i$ th training sample of the ground truth images,  $\mathbf{Y}_i \in \mathbb{R}^{P_1 \times P_2}$  represents the  $i$ th training sample of the PAN images, and  $\mathbf{Z}_i \in \mathbb{R}^{P_1 \times P_2 \times B}$  is the  $i$ th training sample of the LR-MSIs.  $n$  is the batch size which represents the number of training samples randomly selected from the training set.  $\mathcal{L}$  is the loss function to be minimized in the training process. In real applications, due to the limited training samples, we extract patches from the images as training samples.  $P_1, P_2$  are the spatial sizes of the extracted patch.

### IV. EXPERIMENTAL STUDY

In this section, the performances of the proposed method and other state-of-the-art pan-sharpening methods are evaluated on both synthetic and real data sets.

#### A. Experimental Settings

The synthetic data sets are collected from three different sensors including: QuickBird, WorldView2, and WorldView3. Details about these data sets will be presented later. Since the HR-MSIs are not available in these data sets, we follow Wald's protocol to generate the LR-MSIs and PAN images. Then we extracted patches from the image pairs as the training set which will be fed into the network. The size of the patch is  $32 \times 32$

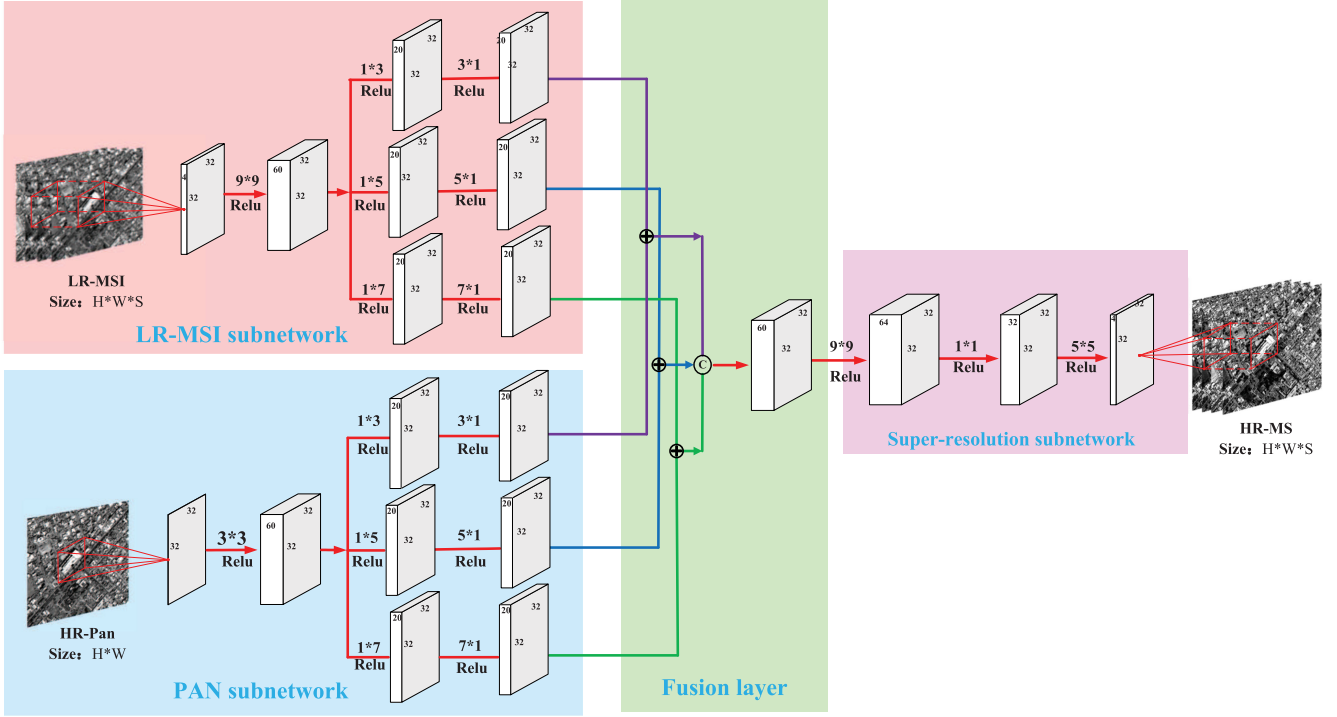


Fig. 1. Flowchart of the proposed method.

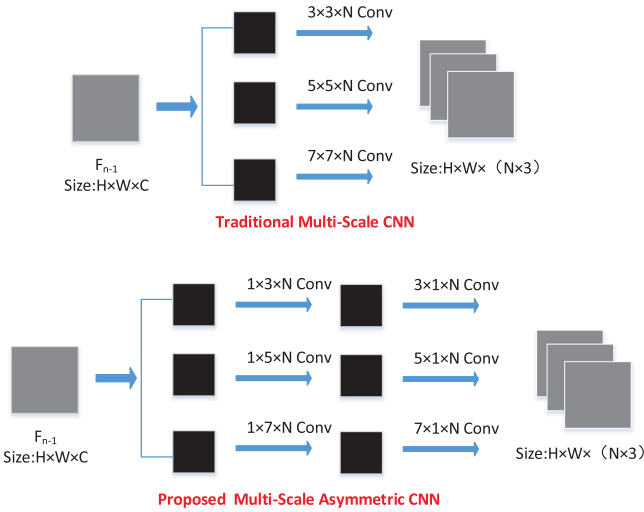


Fig. 2. Comparison of the traditional multiscale CNN and the proposed multiscale asymmetric CNN.

and 200 000 patches are extracted as the training set in all the synthetic experiments.

In the first layer of the feature extraction subnetwork, the kernel size is set to  $W_m = 9$  for the LR-MSI subnetwork and  $W_p = 3$  for the PAN subnetwork. The numbers of convolutional kernels are both set to 60 in the two layers.  $\{1 \times 3, 3 \times 1\}$ ,  $\{1 \times 5, 5 \times 1\}$ , and  $\{1 \times 7, 7 \times 1\}$  kernel pairs are used in the asymmetric layers. The number of convolutional kernels in the asymmetric convolutional layer is set to 20 for every scale. After the fusion layer, the feature maps has a size of  $32 \times 32 \times 60$ .

Finally, a four-channel HR-MSI is produced after the super-resolution subnetwork. The batch size in our experiment is set to 100 in every iteration and the total iteration number is 200 000. We use the AdamOptimizer to optimize the loss function and the learning rate is set to 0.0001 in every layer.

### B. Compared Methods and Quality Measures

We compare our proposed pan-sharpening method with some state-of-the-art pan-sharpening methods including the GS orthogonal method [42], atrous wavelet transform (ATWT) [43], adaptive IHS method (AIHS) [44], sparse fusion of images (SparseFI) [45], CNN-based pan-sharpening (PNN) [24], and MSDCNN [26]. The first four methods are traditional pan-sharpening methods and the last two are deep learning-based methods. All these methods are implemented using the parameters suggested by the authors and tested under the same conditions.

In our experiments, correlation coefficient (CC), root mean square error (RMSE), Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS), spectral angle mapper (SAM), and  $Q_4$  are used to evaluate the performances of the compared methods on the synthetic data sets. CC is used to characterize the geometric distortion. RMSE can be used to measure the spatial quality of the reconstructed image. ERGAS shows a global indication of the fusion quality. The SAM index can evaluate the quality of the spectral structure.  $Q_4$  measures the universal metric between the fused image and the reference image. All these quality measures require the reference HR-MSI which is available in the synthetic experiments. On the contrary, the reference HR-MSI is not available in real data experiments. So



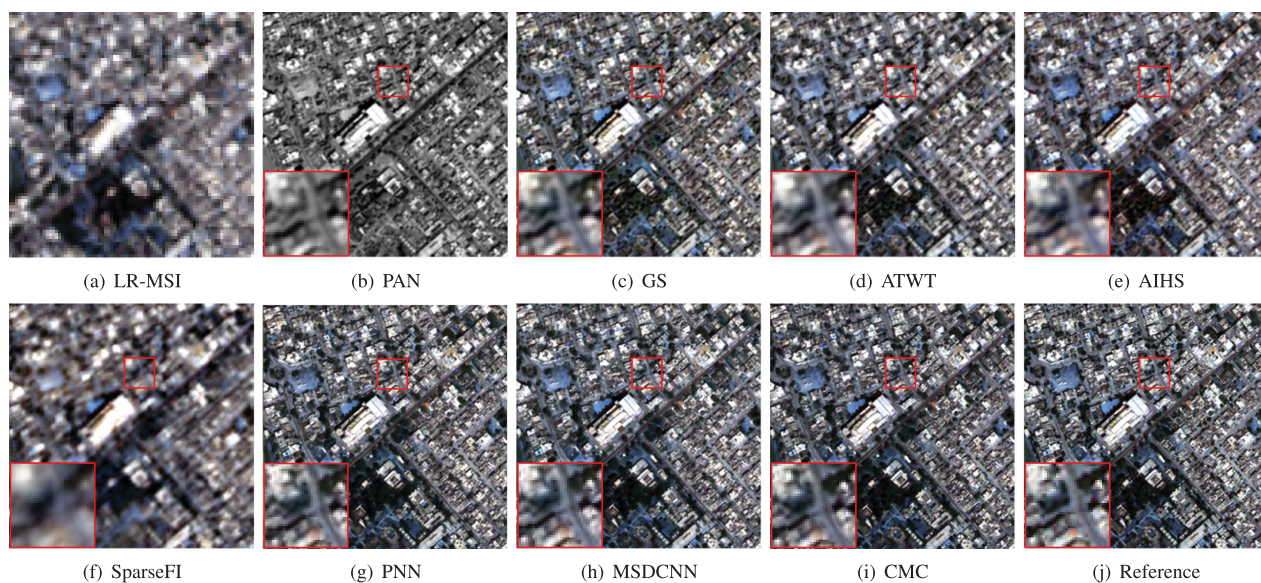


Fig. 3. Visual results of the QuickBird data set. (a) LR-MSI. (b) PAN image. (c)–(i) Pseudo color images of the compared results. (j) Pseudo color image of the reference image.

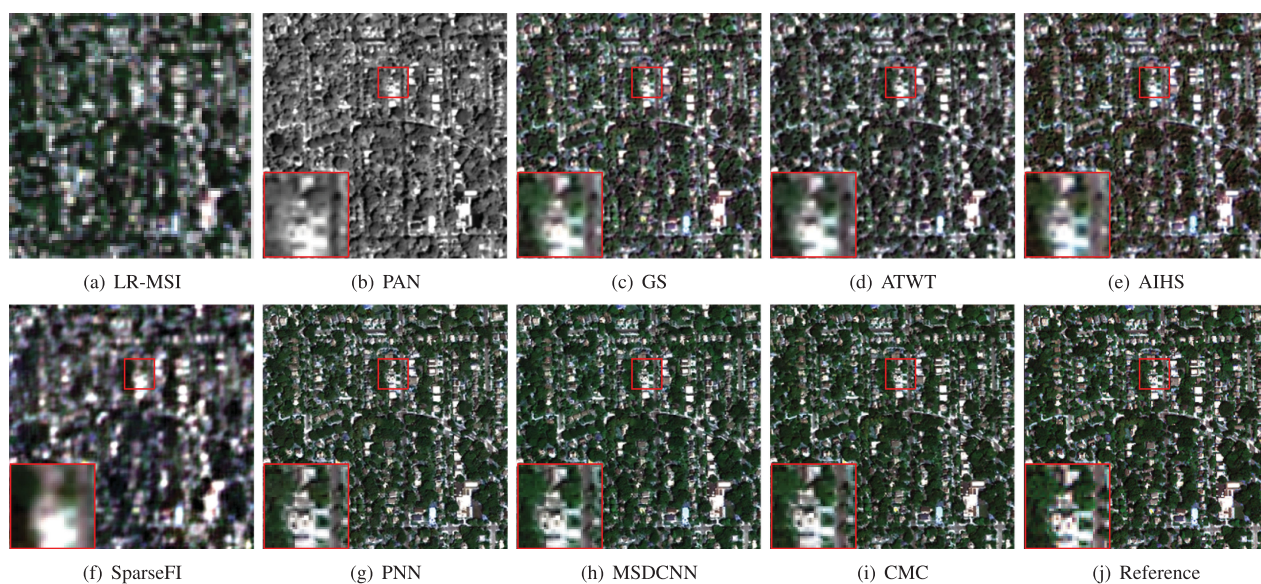


Fig. 4. Visual results of the WorldView-2 data set. (a) LR-MSI. (b) PAN image. (c)–(i) Pseudo color images of the compared results. (j) Pseudo color image of the reference image.

TABLE I  
QUALITY MEASURES ON QUICKBIRD DATA SET

Method	CC	RMSE	ERGAS	SAM	Q <sub>4</sub>
Reference	1	0	0	0	1
GS	0.8525	0.0066	3.3015	0.0690	0.7475
ATWT	0.8549	0.0057	3.0748	0.0634	0.8095
AIHS	0.8615	0.0062	3.2250	0.0654	0.7545
SparseFI	0.7329	0.0157	5.5916	0.1007	0.5618
PNN	0.8803	0.0048	2.8146	0.0548	0.8566
MSDCNN	0.8800	0.0048	2.8191	0.0545	0.8558
CMC	<b>0.8827</b>	<b>0.0047</b>	<b>2.8088</b>	<b>0.0529</b>	<b>0.8607</b>

The best results are showed in bold.

TABLE II  
QUALITY MEASURES ON WORLDVIEW-2 DATA SET

Method	CC	RMSE	ERGAS	SAM	Q <sub>4</sub>
Reference	1	0	0	0	1
GS	0.8199	0.0137	6.5159	0.1651	0.6819
ATWT	0.8223	0.0116	6.0518	0.1520	0.7722
AIHS	0.8282	0.0124	6.3180	0.1567	0.7182
SparseFI	0.6817	0.0316	8.5778	0.2735	0.4844
PNN	0.9104	0.0062	4.4379	0.1159	0.8959
MSDCNN	0.9165	0.0058	4.2697	0.1104	0.9022
CMC	<b>0.9182</b>	<b>0.0058</b>	<b>4.1888</b>	<b>0.1095</b>	<b>0.9046</b>

The best results are showed in bold.



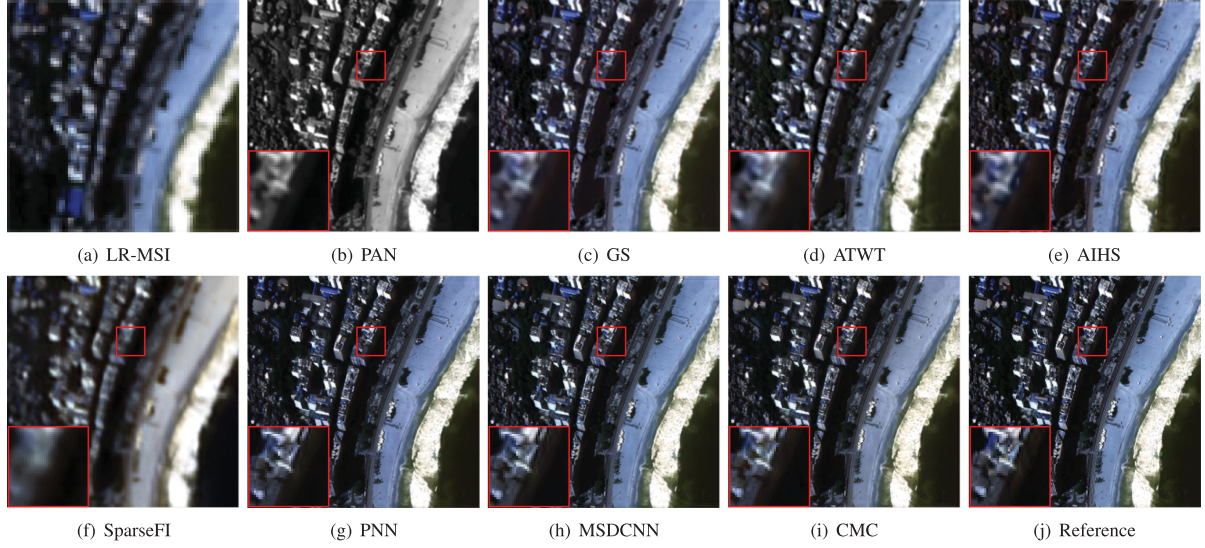


Fig. 5. Visual results of the WorldView-3 data set. (a) LR-MSI. (b) PAN image. (c)–(i) Pseudo color images of the compared results. (j) Pseudo color image of the reference image.

TABLE III  
QUALITY MEASURES ON WORLDVIEW-3 DATA SET

Method	CC	RMSE	ERGAS	SAM	$Q_4$
Reference	1	0	0	0	1
GS	0.9003	0.0118	7.4917	0.1083	0.7609
ATWT	0.9105	0.0107	7.1423	0.1112	0.7886
AIHS	0.8950	0.0122	7.5985	0.1065	0.7515
SparseFI	0.8562	0.0179	9.6064	0.2353	0.6918
PNN	0.9456	0.0065	5.5914	0.0953	0.8812
MSDCNN	0.9470	0.0064	5.5252	<b>0.0904</b>	0.8846
CMC	<b>0.9473</b>	<b>0.0063</b>	<b>5.4845</b>	0.0921	<b>0.8925</b>

The best results are showed in bold.

we use the spectral preservation  $D_\lambda$ , the spatial preservation  $D_S$ , and the quality with no reference (QNR) [46] to evaluate the performances of these methods in real data experiments.

### C. QuickBird Data Set Results

QuickBird satellite was launched on October 18, 2001 by DigitalGlobe. It showed a good performance and offered the highest resolution at the time it was launched. It has two sensors, one captures the PAN image at 0.6-m spatial resolution, and the other sensor captures a 4-band MSI at 2.4-m resolution. It covers the blue, green, red, and near-infrared band separately.

In this experiment, an MSI of size  $2000 \times 2700 \times 4$  and a PAN image of size  $8000 \times 10800$  are given. Following Wald's protocol, we downsample the two images by ratio 4. Then the downsampled images are used as the LR-MSI and the PAN image separately and the original MSI is considered as HR-MSI. Meanwhile, bicubic interpolation algorithm is used to upsample the LR-MSI to match the PAN image's size. We extracted patches from the LR-MSI, PAN image, and HR-MSI as the input and output of the network. The test LR-MSI has a pixel size of  $64 \times 64 \times 4$  and the test PAN image has a

pixel size of  $256 \times 256$ . Fig. 3(a) shows the LR-MSI of the test image in QuickBird data set and the PAN image is shown in Fig. 3(b). The captured scene contains a lot of buildings and seldom homogeneous regions exist. Fig. 3(c)–(i) shows the pseudo color images of the fused results by the compared methods and the reference pseudo color image is shown in Fig. 3(j). It can be observed that there are spatial distortions in the GS, ATWT, AIHS, and SparseFI methods. Deep learning based methods (PNN, MSDCNN, and CMC) can preserve the spatial structures well. Table I shows all the quality measures for the compared methods on the test image. It also demonstrates that the deep learning-based methods perform better than the traditional methods. Among the deep learning-based methods, PNN and MSDCNN achieve similar quality measures and the proposed CMC method outperforms the other methods. This indicates that the introduced CMC is efficient for pan-sharpening.

### D. WorldView-2 Data Set Results

The WorldView-2 data set was collected by the WorldView-2 satellite. The satellite was successfully launched in October 2009. It runs on a 770-km-high sun-synchronous orbit with a full-color channel and four multispectral channels. The multi-spectral channel has the same band as the four bands in the QuickBird satellite. The PAN image obtained by the full-color channel has a spatial resolution of 0.5 m and the MSI has a spatial resolution of 1.8 m.

In the training process, the MSI intercepted in the training data has a pixel size of  $1000 \times 2000 \times 4$ , and the PAN image has a pixel size of  $4000 \times 8000$ . The test LR-MSI has a pixel size of  $64 \times 64 \times 4$  and the test PAN image has a pixel size of  $256 \times 256$ . In this experiment, the captured scene contains the city of Washington, including buildings and green vegetation. Table II shows the quality measures of all the compared methods on the WorldView-2 data set. Our proposed method achieves the best results in all quality measures. In Fig. 4, spatial and



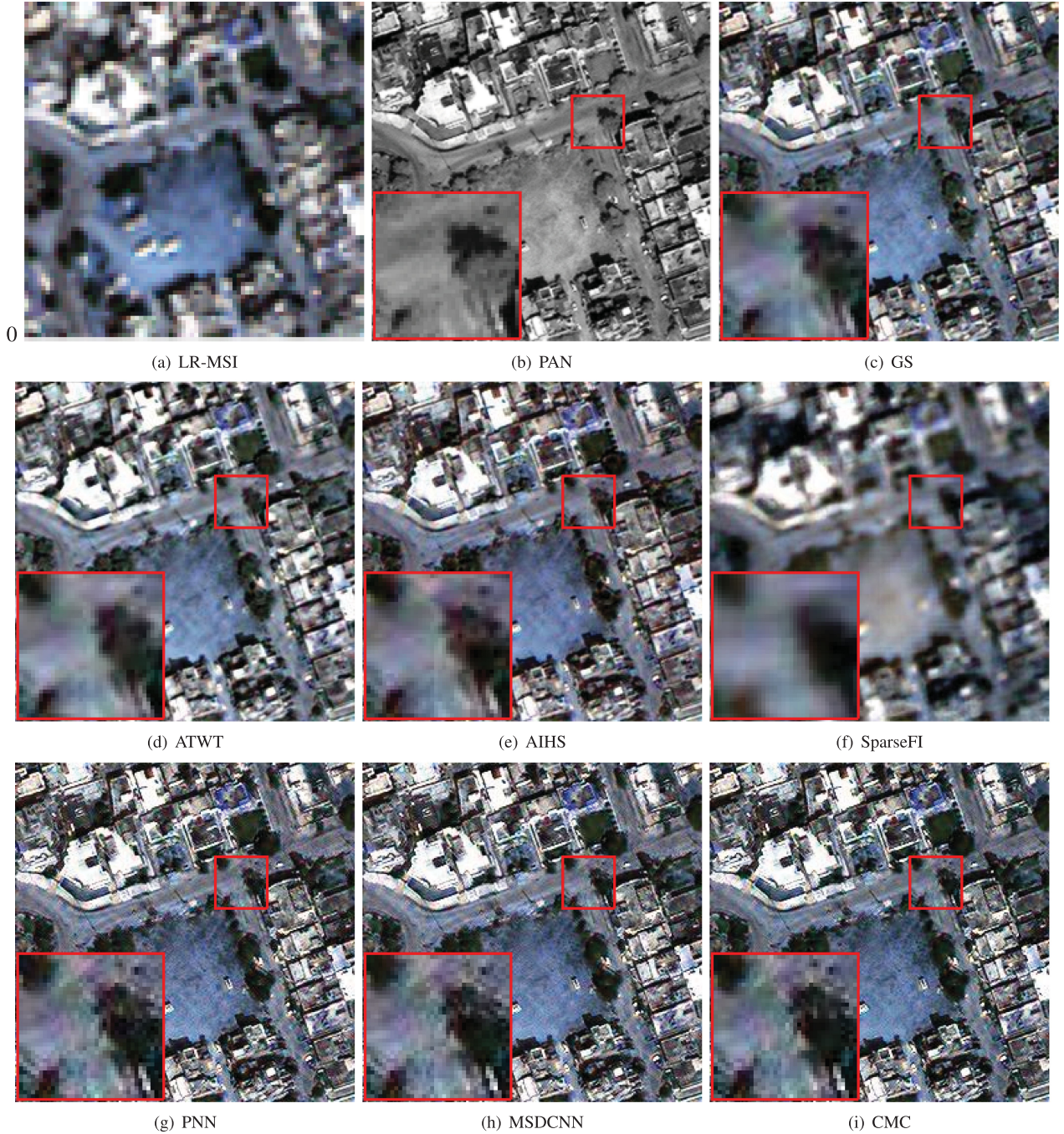


Fig. 6. Visual results of the real data set. (a) LR-MSI. (b) PAN image. (c)–(i) Pseudo color images of the compared results. (j) Pseudo color image of the reference image.

spectral distortions can be observed in the traditional methods especially in SparseFI. On the contrary, deep-learning-based methods produce images that are very similar to the reference image. The image generated by the proposed CMC method has seldom differences with the reference image.

#### E. WorldView-3 Data Set Results

The WorldView-3 data set was collected by the WorldView-3 satellite which was launched by DigitalGlobe in the United

States in August 2014. It runs on a 617-km-high sun-synchronous orbit and is the world's highest resolution optical commercial satellite. It also contains a full-color channel and four multispectral channels. The multispectral channels are the same with the QuickBird and WorldView-2 data sets. The spatial resolution of the PAN image is 0.31 m and the spatial resolution of MSI is 1.24 m.

In this experiment, images of the Rio city are used and the size of the MSI used in the training data has a pixel size of  $1200 \times 1200 \times 4$ . The PAN image has a pixel size of  $4800 \times 4800$ . The captured scene is composed of buildings,



TABLE IV  
QUALITY MEASURES ON REAL DATA SET

	$D_\lambda$	$D_S$	QNR
Reference	0	0	1
GS	0.0345	0.1674	0.8039
ATWT	0.0939	0.1247	0.7931
AIHS	<b>0.0148</b>	0.1396	0.8477
SparseFI	0.1054	0.0862	0.8175
PNN	0.0594	<b>0.0805</b>	0.8648
MSDCNN	0.0655	0.0859	0.8543
CMC	0.0539	0.0810	<b>0.8695</b>

The best results are showed in bold.

roads, and grass. Fig. 5 shows the fused results of all the compared methods. The numerical results of all the compared methods and the proposed method are shown in Table III. Our proposed CMC method performs the best in all the quality measures except for SAM. MSDCNN is 0.0017 lower than our method in SAM. Although our method is not the best in spectral preservation, the proposed method gives a robust improvement in spatial structural preservation in the WorldView-3 data set.

#### F. Real Data Set Results

In the real data set experiment, we extract a LR-MSI of size  $64 \times 64 \times 4$  from the QuickBird data set. Then the corresponding PAN image has a pixel size of  $256 \times 256$ . The fused HR-MSI is  $256 \times 256 \times 4$ . The QNR,  $D_\lambda$ , and  $D_S$  of all the compared methods on the real data set are shown in Table IV. Note that these quality measurements only measure the similarities between the fused image and the low-resolution observation, not the real fidelity at the real resolution. So these quality measurements cannot be the only criteria.

From Table IV, we can see that the proposed CMC method is the best in QNR measurement and the second best in spatial preservation  $D_S$ . In spectral preservation  $D_\lambda$ , AIHS and GS are better than the proposed method. Fig. 6 shows the fused results of the compared methods. We can see that the results of GS, ATWT, and SparseFI are blurry. The deep learning-based methods have sharp edges and the proposed method contains less noise than the other deep learning methods. This indicates that the proposed method can successfully recover the HR-MSI in real data pan-sharpening.

#### G. Parameter Analysis

Multiscale feature extraction is the key step in the proposed method. Thus, the kernel sizes  $W_m$  and  $W_p$  are very important and we analyze the impact of the two parameters. In the proposed method,  $W_m$  should be larger than  $W_p$ . Since we want to map the PAN image and MSI into the same spatial scale, the ratio between  $W_m$  and  $W_p$  is a key factor to our method. Therefore, we fix  $W_p$  to the commonly used kernel size 3 and change  $W_m$  among  $\{7, 9, 11\}$ . To compare the effect of different  $W_m$ , we plot the training loss along the iteration number in different  $W_m$ . Fig. 7 shows the comparison results. It can be seen that the loss decreases the fastest when  $W_m = 9$  and the loss values of curves

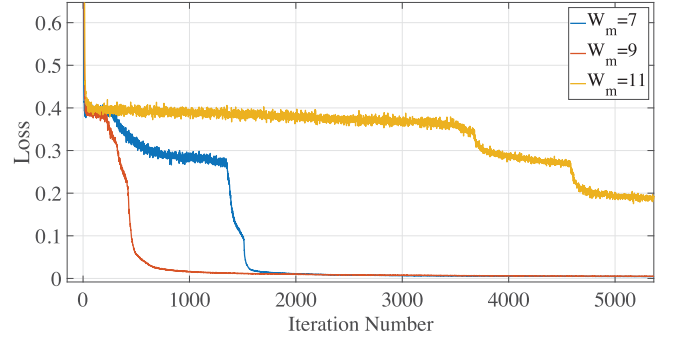


Fig. 7. Training loss versus iteration number with different  $W_m$  on the QuickBird data set.

TABLE V  
TRAINING TIME FOR DEEP LEARNING METHODS

Training Time(min)	QuickBird	WorldView-2	WorldView-3
PNN	381	340	350
MSDCNN	650	612	502
CMC	531	518	355

$W_m = 7, 9$  are the lowest when iteration number equals 5000. Thus, we set  $W_m = 9$ ,  $W_p = 3$  in our experiments.

#### H. Training Time Analysis

In this section, we will analyze the training time of the proposed method. We compare our method with other deep learning based pan-sharpening methods. In the test process, it costs less than 5 s to fuse the input LR-MSI and PAN image and little difference can be observed between these deep learning-based methods. So we only show the training time for PNN, MSDCNN, and the proposed method in Table V. It can be observed that PNN is the fastest due to its simple net architecture. The proposed CMC method is faster than MSDCNN. With the introduced multiscale asymmetric convolution factorization method, the number of parameters in the network is smaller; so the training time is decreased. Meanwhile, the performances of the proposed method are improved. This means the CMC is effective and efficient for pan-sharpening.

#### V. CONCLUSION

In this article, a two-stream deep learning architecture is proposed for pan-sharpening. The proposed network is composed of feature extraction subnetworks, fusion layer, and super-resolution subnetwork. In the feature extraction subnetworks, two separated subnetworks are used to extract features of the MSI and the PAN image separately. Based on the observation that the MSI and the PAN image have different spatial resolutions, different sizes of convolutional kernels are used to match the scale of the two source images in the first layer. A multiscale asymmetric convolution factorization method is proposed to extract multiscale features. Then, the two feature extraction subnetworks are coupled in the fusion layer. Features of the same scale are first summed and then features of all scales are concatenated as one feature map. Finally, the super-resolution

subnetwork enlarges the feature to the desired resolution and channels. Experimental results on both synthetic and real data sets show that our method performs better than other state-of-the-art pan-sharpening methods.

In future, we will use the multiscale CNN to learn the deep priors and combine the learnt deep priors with the variational model.

## REFERENCES

- [1] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.
- [2] 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.
- [3] 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.
- [4] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptive Markov random field approach for classification of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 973–977, Sep. 2011.
- [5] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.
- [6] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LEMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 147, pp. 193–205, 2019.
- [7] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, Jul. 2019.
- [8] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inform. Fusion*, vol. 2, no. 3, pp. 177–186, 2001.
- [9] B. Aiazzi, 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.
- [10] 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.
- [11] H. Li, B. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graphical Models Image Process.*, vol. 57, no. 3, pp. 235–245, 1995.
- [12] B. Aiazzi, 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, 2006.
- [13] G. Vivone *et al.*, "Pansharpening based on semiblind deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1997–2010, Apr. 2015.
- [14] P. Liu, L. Xiao, J. Zhang, and B. Naz, "Spatial-hessian-feature-guided variational model for pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2235–2253, Apr. 2016.
- [15] M. Guo, H. Zhang, J. Li, L. Zhang, and H. Shen, "An online coupled dictionary learning approach for remote sensing image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1284–1294, Apr. 2014.
- [16] L.-J. Deng, G. Vivone, W. Guo, M. Dalla Mura, and J. Chanussot, "A variational pansharpening approach based on reproducible kernel hilbert space and heaviside function," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4330–4344, Sep. 2018.
- [17] 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.
- [18] J. Duran, A. Buades, B. Coll, and C. Sbert, "A nonlocal variational model for pansharpening image fusion," *SIAM J. Imag. Sci.*, vol. 7, no. 2, pp. 761–796, 2014.
- [19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, 2015, Art. no. 436.
- [20] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.
- [21] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [22] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2020.3015157](https://doi.org/10.1109/TGRS.2020.3015157).
- [23] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang, "A new pan-sharpening method with deep neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1037–1041, May 2015.
- [24] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, 2016, Art. no. 594.
- [25] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vision*, Springer, 2014, pp. 184–199.
- [26] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multipdepth convolutional neural network for remote sensing imagery pan-sharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 978–989, Mar. 2018.
- [27] X. Liu, Y. Wang, and Q. Liu, "Remote sensing image fusion based on two-stream fusion network," in *Proc. Int. Conf. Multimedia Model.*, Springer, 2018, pp. 428–439.
- [28] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 1753–1761.
- [29] H. Shen, M. Jiang, J. Li, Q. Yuan, Y. Wei, and L. Zhang, "Spatial-spectral fusion by combining deep learning and variation model," *arXiv:1809.00764*, 2018.
- [30] 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.
- [31] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018.
- [32] S. Eghbalian and H. Ghassemian, "Multi spectral image fusion by deep convolutional neural network and new spectral loss function," *Int. J. Remote Sens.*, vol. 39, no. 12, pp. 3983–4002, 2018.
- [33] A. Azarang, H. E. Manoochehri, and N. Kehtarnavaz, "Convolutional autoencoder-based multispectral image fusion," *IEEE Access*, vol. 7, pp. 35673–35683, 2019.
- [34] L. Liu *et al.*, "Shallow-deep convolutional network and spectral-discrimination-based detail injection for multispectral imagery pan-sharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1772–1783, 2020.
- [35] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "PAN-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Inform. Fusion*, vol. 62, pp. 110–120, 2020.
- [36] W. Xie, Y. Cui, Y. Li, J. Lei, Q. Du, and J. Li, "HPGAN: Hyperspectral pansharpening using 3-D generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, 2020, to be published, doi: [10.1109/TGRS.2020.2994238](https://doi.org/10.1109/TGRS.2020.2994238).
- [37] K. Zheng, L. Gao, Q. Ran, X. Cui, B. Zhang, W. Liao, and S. Jia, "Separable-spectral convolution and inception network for hyperspectral image super-resolution," *Int. J. Mach. Learning Cybern.*, vol. 10, no. 10, pp. 2593–2607, 2019.
- [38] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [40] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [41] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 2818–2826.
- [42] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6 011 875, Jan. 4, 2000.
- [43] J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.



- [44] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive IHS pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 746–750, Oct. 2010.
- [45] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, May 2013.
- [46] L. Alparone, B. Aiazzi, 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, 2008.



**Jie Wei** was born in Shanxi, China, in 1981. She received the B.Sc. and M.Sc. degrees in computer science and technology from the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2004 and 2006, respectively. She is currently working toward the Ph.D. degree in computer science at the Nanjing University of Science and Technology.

Her research interests include hyperspectral image classification, sparse representation, and deep learning.



**Yang Xu** (Member, IEEE) received the B.Sc. degree in applied mathematics and the Ph.D. degree in pattern recognition and intelligence systems from Nanjing University of Science and Technology (NUST), Nanjing, China, in 2011 and 2016, respectively.

He is currently an Associate Professor with the School of Computer Science and Engineering, NUST. His research interests include hyperspectral image classification, hyperspectral detection, image processing, and machine learning.



**Wanting Cai** received the B.Sc. degree in computer science and technology from School of Computer Science and Technology, Anhui University, Hefei, China, and the M.Sc. degree in computer science and technology from the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2016 and 2019, respectively.

Her research interests include hyperspectral image fusion and deep learning.



**Zebin Wu** (Senior Member, IEEE) received the B.Sc. and Ph.D. degrees in computer science and technology from Nanjing University of Science and Technology, Nanjing, China, in 2003 and 2007, respectively.

He is currently a Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology. He was a Visiting Scholar with the GIPSA-lab, Grenoble INP, Université Grenoble Alpes, Grenoble, France, from August 2018 to September 2018. He was a Visiting Scholar with the

Department of Mathematics, University of California Los Angeles, CA, USA, from August 2016 to September 2016 and from July 2017 to August 2017. He was a Visiting Scholar with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain, from June 2014 to June 2015. His research interests include hyperspectral image processing, parallel computing, and remotely sensed big data processing.



**Jocelyn Chanussot** (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the Savoie University, Annecy, France, in 1998.

Since 1999, he has been with the Grenoble Institute of Technology, Grenoble, France, where he is currently a Professor of Signal and Image Processing. Since 2013, he has been an Adjunct Professor with the University of Iceland, Reykjavik, Iceland, and since 2019, an Adjunct Professor with the Chinese Academy of Sciences, Beijing, China. His research interests include image analysis, data fusion, and arti-

ficial intelligence. He has authored or coauthored over 350 papers in international journals and conferences.

Dr. Chanussot is the recipient of multiple best paper awards and recognitions. He is a Highly Cited Researcher (Clarivate Analytics). He is the General Chair of the IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006–2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing (2009). He has been an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, since 2007 and for the IEEE TRANSACTIONS ON IMAGE PROCESSING, since 2017.



**Zhihui Wei** (Member, IEEE) was born in Jiangsu, China, in 1963. He received the B.Sc. and M.Sc. degrees in applied mathematics and Ph.D. degree in communication and information system from South East University, Nanjing, China, in 1983, 1986, and 2003, respectively.

He is currently a Professor and Doctoral Supervisor with Nanjing University of Science and Technology (NUST), Nanjing, China. His research interests include partial differential equations, mathematical image processing, multiscale analysis, sparse repre-

sentation, and compressive sensing.