Bayesian Pan-Sharpening With Multiorder Gradient-Based Deep Network Constraints

Penghao Guo, Peixian Zhuang, and Yecai Guo

***Abstract*—Pan-sharpening aims at acquiring a multispectral image with a high spatial resolution by fusing a low-resolution multispectral image and a panchromatic image. In order to improve spatial details and reduce spectral distortions, we develop a new pan-sharpening model based on the Bayesian theory, which involves three assumptions: 1) the low-resolution multispectral images are generally decimated from the high-resolution multispectral images by convolution with a blurring kernel; 2) different from most pan-sharpening methods that use linear manners to preserve spatial information, we attempt a nonlinear manner based on a convolutional neural network composed of the proposed multiscale recursive blocks, and we train our network parameters in multiorder gradient domains to preserve more spatial structures; and 3) we introduce an anisotropic total variation prior in multiorder gradient domains to reconstruct better image edges and details. We establish the posterior probability model based on the above assumptions and derive an efficient optimization scheme to address the proposed objective function. Final experimental results demonstrate that the proposed model can overcome the restriction of a linear model and achieve better spectral and spatial fusion, compared with several traditional and deep-learning-based pan-sharpening approaches. In addition, our model achieves more promising generalization across different satellites than other deep-learning-based methods.**

***Index Terms*—Bayesian theory, convolutional neural network (CNN), multiorder gradient, pan-sharpening.**

I. Introduction

With the development of remote sensing technology, remote sensing images have been widely applied in many practical fields, such as physiognomy monitoring, vegetation identification, and object classification. Due to technical constraints, satellite capture systems only acquire two sorts of images in the same scene at the same time. One is the high-spatial-resolution panchromatic (PAN) image, which contains little spectral information. Another is the low-spatial-resolution multispectral (LRMS) image, which covers a more precise spectral range for the spectral resolution. To obtain a multispectral image with high-spatial resolution multispectral (HRMS), researchers have made efforts to fuse PAN and LRMS images, and this fusion is named pan-sharpening.

A variety of pan-sharpening methods have been proposed in recent decades [1], and most pan-sharpening approaches can be divided into four main categories [2]: component substitution (CS)-based methods, multiresolution analysis (MRA)-based methods, variational optimization (VO)-based methods, and deep learning (DL)-based methods. The CS-based methods obtain the fused images by injecting high spatial structures extracted from the PAN image into the upsampled LRMS image. Gram–Schmidt (GS) [3], hyperspherical color sharpening [47], intensity–hue–saturation [4], [5], and principal component analysis (PCA) [6], [7] are the most popular CS-based methods. Their fusion tends to preserve spatial information but introduces some spectral distortions. Different from the CS-based methods, the MRA-based methods extract high spatial structures by a spatial filter or other spatial operators, and decimated or undecimated wavelet transform [8], [9], Á Trous wavelet transform [10], and Laplacian pyramid [11] are similar MRA-based methods. They can sharpen the LRMS image at the expense of spatial distortions. The VO-based methods are based on VO models, which include the spectral fidelity, the spatial enhancement, and the prior. Ballester *et al.* [12] proposed the first variational pan-sharpening model based on a linear combination, which can well preserve spectral components. However, this method could produce the blurring effects due to the reason that estimating an accurate blur kernel is challenging. To improve blurred edges, Fang *et al.* [13] introduced a guided filter-based fusion method to promote spatial structures while minimizing the spectral distortion. Wang *et al.* [14] presented a variational model based on the Bayesian theory assumption that the HRMS image is coincident with that contained in the PAN image; the HRMS image and the LRMS image should share the same spectral information. However, all aforementioned methods conduct the fusion in linear manners, which cannot achieve a better tradeoff between spectral and spatial quality. In recent years, DL methods that belong to nonlinear manners in computer vision have been promising approaches. A lot of research studies have begun exploring pan-sharpening methods based on a deep convolutional neural network (CNN). Masi *et al.* [15] adapted a three-layer CNN architecture for superresolution to pan-sharpening. Yuan *et al.* [16] introduced a multiscale multidepth CNN for the pan-sharpening of remote sensing imagery. These methods train network parameters in the image domain; however, they are

Manuscript received November 26, 2019; revised January 8, 2020 and January 30, 2020; accepted February 12, 2020. Date of publication March 2, 2020; date of current version March 17, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61701245, in part by the Startup Foundation for Introducing Talent of Nanjing University of Information Science and Technology (NUIST) under Grant 2243141701030, in part by College Students Practice Innovation Training Program of NUIST under Grant 201910300079Y, and in part by a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions. *(Penghao Guo and Peixian Zhuang contributed equally to this work.) (Corresponding author: Peixian Zhuang.)*

The authors are with the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: pajiju820@163.com; zhuangpeixian0624@163.com; guo-yecai@163.com).

Digital Object Identifier 10.1109/JSTARS.2020.2975000

weak in spatial preservation. To address this problem, Yang *et al.* [17] proposed a deep network architecture called PanNet, which trains their network parameters in the high-pass filtering domain to preserve spatial structures. Unfortunately, these DL-based algorithms are limited in their generalization across different satellites.

In order to address the above issues, a novel pan-sharpening model based on the Bayesian theory is established. Our model is formulated into three probability terms. The first term can protect spectral information from LRMS images, the second term preserves the spatial structure of HRMS images, and the third term is added to ensure structural reconstruction of HRMS images. To sum up, the contributions of our article are presented as follows.

1) We develop a Bayesian pan-sharpening model based on the following assumptions.
   a) The LRMS images are universally generated from the HRMS images by convolution with a blurring kernel.
   b) A multiorder gradient CNN, composed of the proposed multiscale recursive blocks, is designed to preserve better spatial structures in a nonlinear manner.
   c) An anisotropic total variation prior in multiorder gradient domains is used to better reconstruct image edges and details.

   Our model has a better tradeoff between spectral and spatial fusion.
2) We attempt to use a multiorder gradient-based CNN (MCNN) to enforce the spatial preservation based on a nonlinear mapping of PAN and LRMS images. And our CNN output is incorporated with a variational-based optimization framework for pan-sharpening. Experimental results demonstrate that our method outperforms other pan-sharpening methods in structural fusion and spectral preservation. To the best of our knowledge, it is the first work that combines multiorder gradients with a CNN. Moreover, we perform the proposed CNN in multiorder gradient domains, which allows for bypassing the gap between different satellites by using network parameters from trained on one satellite dataset, and our model is more robust when generalized to new satellites.

II. Bayesian Pan-Sharpening Model

We assume that y_p represents the PAN image, y denotes the LRMS image, and x denotes the HRMS image. According to the Bayesian theorem, our pan-sharpening model is established as a posterior probability

$$p(x|y,f) \propto p(y|x)p(f|x)p(x) \tag{1}$$

where $p(y|x)$ and $p(f|x)$ are the likelihood of spectral and spatial preservation, respectively. $p(x)$ denotes the prior of the HRMS image x. f is a nonlinear manner formulated as

$$f = f_{\text{MCNN}}(y, y_p) \tag{2}$$

where f_{MCNN} is an MCNN to better preserve spatial structures.

Spectral preservation likelihood $p(y|x)$: The HRMS image and the LRMS image should contain the same spectral information. Followed our assumption that the LRMS image is generally decimated from the HRMS image by convolution with a blurring kernel. The relationship of y and x is modeled as

$$y = k * x + \varepsilon_1 \tag{3}$$

where ε_1 is the additive noise, which is a Gaussian distribution with zero mean and variance σ_1^2. We denote k as a blurring kernel. An averaging kernel [19], [29], [48] is widely used for k, since its experimental results are comparable with other results using estimated kernels [20], [21]. The likelihood $p(y|x)$ can be defined as

$$p(y|x) = p(\varepsilon_1) = N(\varepsilon_1|0,\sigma_1^2) = N(y|x,\sigma_1^2). \tag{4}$$

Spatial preservation likelihood $p(f|x)$: The PAN image contains more spatial information, and previous pan-sharpening methods assume that the spatial information of the HRMS image is acquired from the PAN image. The first P+XS method [12] assumes that a linear combination of all bands of the HRMS image should be closed to the PAN image. Different from the P+XS method that preserves the spatial information in the image domain, some approaches enforce the structure similarity in the gradient domain to avoid intensity differences of the HRMS image and the PAN image. Recently, an effective local linear regression model [22] has been proposed to constrain the gradient difference of the PAN image and the HRMS image. We break the linear limitation of these methods, which preserve the spatial information in linear model; our model enforces the spatial fusion in the multigradient domains for the PAN image and the LRMS image in a nonlinear manner. And the spatial preservation term is based on the following assumption:

$$\nabla^* f = \nabla^* x + \nabla^* \varepsilon_2 \tag{5}$$

where $\nabla^* f$ denotes an MCNN, which trains network parameters in multiorder gradient domains. ∇^* contains the first-order gradient $\nabla_1(\nabla_{1h} = [1,-1], \nabla_{1v} = [1;-1])$ and the second-order gradient ∇_2; meanwhile, we employ a Laplacian operator $[1,0,1;0,-4,0;1,0,1]$ [23] to represent the second-order gradient. The function (5) can be detailed as follows:

$$\begin{aligned} \nabla_1 f &= \nabla_1 x + \nabla_1 \varepsilon_{21} \\ \nabla_2 f &= \nabla_2 x + \nabla_2 \varepsilon_{22}. \end{aligned} \tag{6}$$

Since $\nabla^* \varepsilon_2$ is a random sequence following Gaussian distribution, $\nabla_1 \varepsilon_{21}$ and $\nabla_2 \varepsilon_{22}$ follow the Gaussian distributions $N(0,\sigma_{21}^2)$ and $N(0,\sigma_{22}^2)$, respectively.

The likelihood of spatial preservation $p(f|x)$ is formulated as

$$\begin{aligned} p(f|x) &= p(\nabla^* f|\nabla^* x) = p(\nabla_1 f|\nabla_1 x)p(\nabla_2 f|\nabla_2 x) \\ &= p(\nabla_1 \varepsilon_{21})p(\nabla_2 \varepsilon_{22}) \\ &= N(\nabla_1 f|\nabla_1 x, \sigma_{21}^2)N(\nabla_2 f|\nabla_2 x, \sigma_{22}^2). \end{aligned} \tag{7}$$

Prior $p(x)$: Natural image gradients contain edge-based structures and follow a heavy-tailed distribution learnt from generic image databases [24], [25]. We assume that image gradients are piecewise continuous, and the first-order gradient distribution of the HRMS image x is defined by a Laplacian distribution [21], [26] with location zero and scale s_1, and the second-order gradient distribution of x obeys a Laplacian distribution with

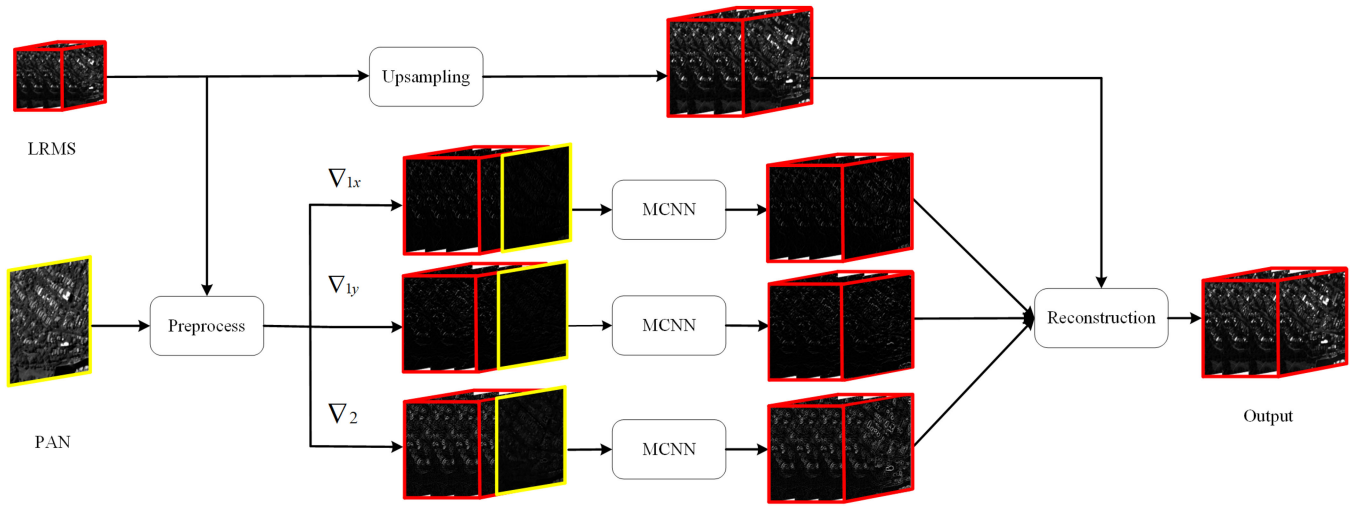

Fig. 1. Workflow of the proposed method.

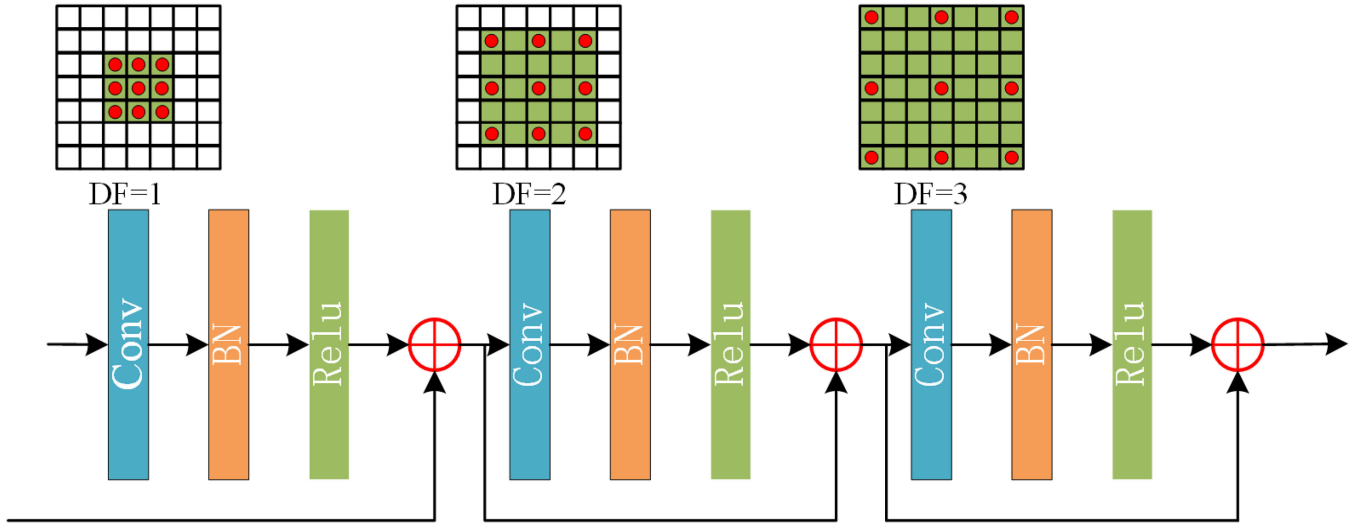

Fig. 2. MDRN block of the inference network.

location zero and scale s_2. The prior is written as

$$p(x) = L(\nabla_1 x|0, s_1)L(\nabla_2 x|0, s_2). \tag{8}$$

Taking all likelihood and prior definitions into (1), our probabilistic model can be derived as

$$p(x|y, f) = N(y|x, \sigma_1^2)N(\nabla_1 f|\nabla_1 x, \sigma_{21}^2)N(\nabla_2 f|\nabla_2 x, \sigma_{22}^2) \times L(\nabla_1 x|0, s_1)L(\nabla_2 x|0, s_2). \tag{9}$$

III. Optimization Algorithm

To obtain the HRMS image x, we first transform the maximum *a posteriori* problem into an energy minimization problem, namely, $E(x) = -\log(p(x|y, f))$, and our overall objective function is established as follows:

$$E(x) = \| y - k * x\|_2^2 + v_1\|\nabla_1 x - \nabla_1 f\|_2^2 + v_2\|\nabla_2 x - \nabla_2 f\|_2^2 + \xi\|\nabla_1 x\|_1 + \eta\|\nabla_2 x\|_1 \tag{10}$$

where $v_1 = \sigma_1^2/\sigma_{21}^2$, $v_2 = \sigma_1^2/\sigma_{22}^2$, $\xi = \sigma_1^2/s_1$, and $\eta = \sigma_1^2/s_2$. $\|y - k * x\|_2^2$ is the spectral preservation term following the assumption that the HRMS image x and the LRMS image y contain same spectral information. $\|\nabla_1 x - \nabla_1 f\|_2^2$ and $\|\nabla_2 x - \nabla_2 f\|_2^2$ denote the spatial preservation terms, which impose the l_2 norm to enforce the structural consistency between the PAN image y_p and the HRMS image x in multiorder gradient domains. $\|\nabla_1 x\|_1$ and $\|\nabla_2 x\|_1$ are the prior terms, which use the l_1 norm to enforce the multiorder gradient sparsity of the HRMS image x. The workflow of our model is shown in Fig. 1, and we will detail each component in turn.

A. Inference Network

Based on the architecture of pan-sharpening by CNNs (PNN) [15], we propose an efficient architecture of the inference network, which aims to extract the spatial information for pan-sharpening. The proposed dilated multilevel residual network

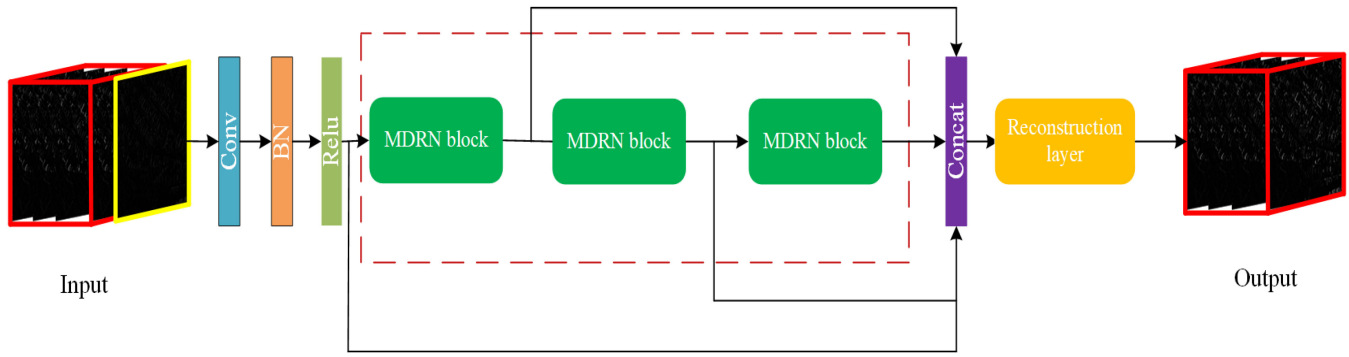

Fig. 3. Architecture of the proposed inference network.

(MDRN) block consists of dilated convolutional layers [27], [28], batch normalization (BN) [33] layers, rectified linear units (ReLU) [34], and residual learning [45]. Fig. 2 shows a block of the proposed network, and we introduce some components of DL that are used in our network.

1) Dilated Convolution: To efficiently capture image feature information, the CNN model should enlarge the receptive field during the training procedure. In the traditional convolutional layer, the receptive field can be enlarged by stacking more convolutional layers or increasing the filter size, but these approaches could generate some adverse effects such as computational burden and overfitting. To enlarge receptive field and reduce network complexity, we adopt dilated convolution as a substitute for the traditional convolutional layer. Different from the traditional convolutional layer, the dilated convolution enlarges the receptive field without introducing extra computational complexity. And a dilated rate parameter called dilated factor (DF) is mainly used to indicate the dilatation size. As shown in Fig. 2, there are three convolutional layers in a recursive block, which consists of one normal convolutional layer, one two-dilated convolutional layer, and one three-dilated convolutional layer. Fig. 2 also provides the visualization of the dilated filter with the DFs as 1, 2, and 3, respectively.

2) Batch Normalization: With the increase of convolutional layers, the computational burden is added, the network takes more runtime for converging, and BN aims to reduce convergence time for training. BN can reduce an internal covariate shift by the fixed means and variances of layer inputs to accelerate the training. BN not only reduces the training time significantly, but also improves the generalization ability of the network by reducing the dependence of gradients on the scale of the parameters or of their initial values.

3) Rectified Linear Units: The activation function increases the nonlinearity of the neural network model; ReLU are generally employed in the neural network model as an activation function.

4) Residual Learning: With the number of convolutional layers increased, CNNs can achieve more complicated nonlinear mappings and can result in poor training accuracy. To overcome this problem, residual learning [45] is proposed to introduce an equal fast connection to solve the problem of gradient disappearance. The residual learning has outstanding performance in image tasks [30]–[32]. Here, we adopt the basic residual block, which uses a shortcut during two contiguous convolutional layers as shown in Fig. 2.

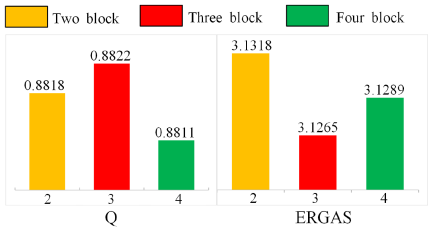

Fig. 4. Average ERGAS [39] and Q [42] of the proposed model with different repeated blocks of the inference network.

5) Network Architecture: We develop an MCNN framework consisting of repeated proposed blocks followed by the concatenation layer; the architecture of our inference network is shown in Fig. 3. Our network has three repeated blocks with 11 convolution layers and 32 filters per layer, and the size of each filter is 3×3. We show different numbers of repeated blocks in Fig. 4, where we find that deepening the network with more blocks does not mean that the extracted features are more favorable to pan-sharpening performance. And the number of blocks is set to 3 suitable for our framework.

B. Model-Based Optimization

It is challengeable to directly optimize the energy minimization problem (10), since $\nabla^* f$ is a nonlinear operation, which is generated by the proposed inference network, and the l_1 norm in the last term is hard to solve directly. In our optimization, two auxiliary variables q and s are introduced for l_1 norm approximations, and thus, our objective function (10) can be reformulated as follows:

$$\begin{aligned} E(x,q,s) = &\| y - k * x\|_2^2 \\ &+ v_1\|\nabla_1 x - f_1\|_2^2 + v_2\|\nabla_2 x - f_2\|_2^2 \\ &+ \xi\|\nabla_1 x - q\|_2^2 + \alpha\|q\|_1 + \eta\|\nabla_2 x - s\|_2^2 + \beta\|s\|_1 \end{aligned} \quad (11)$$

where $f_1 = f_{\text{MCNN}}(\nabla_1 y, \nabla_1 y_p)$, $f_2 = f_{\text{MCNN}}(\nabla_2 y, \nabla_2 y_p)$, and α and β are the parameters to balance each term in this model.

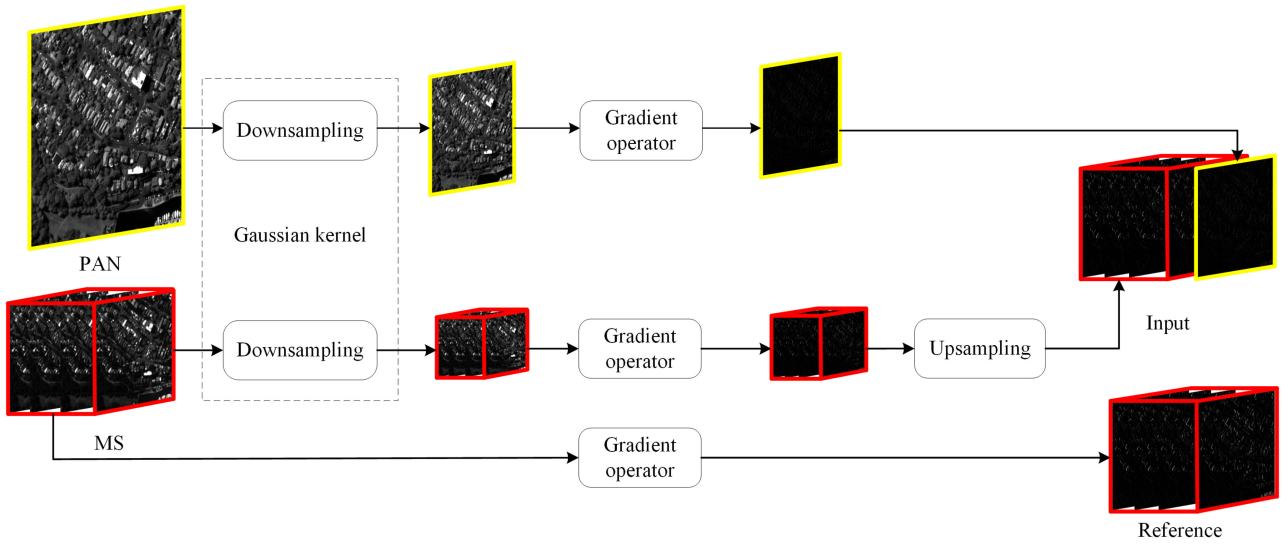

Fig. 5. Process of generating a training dataset through the Wald protocol.

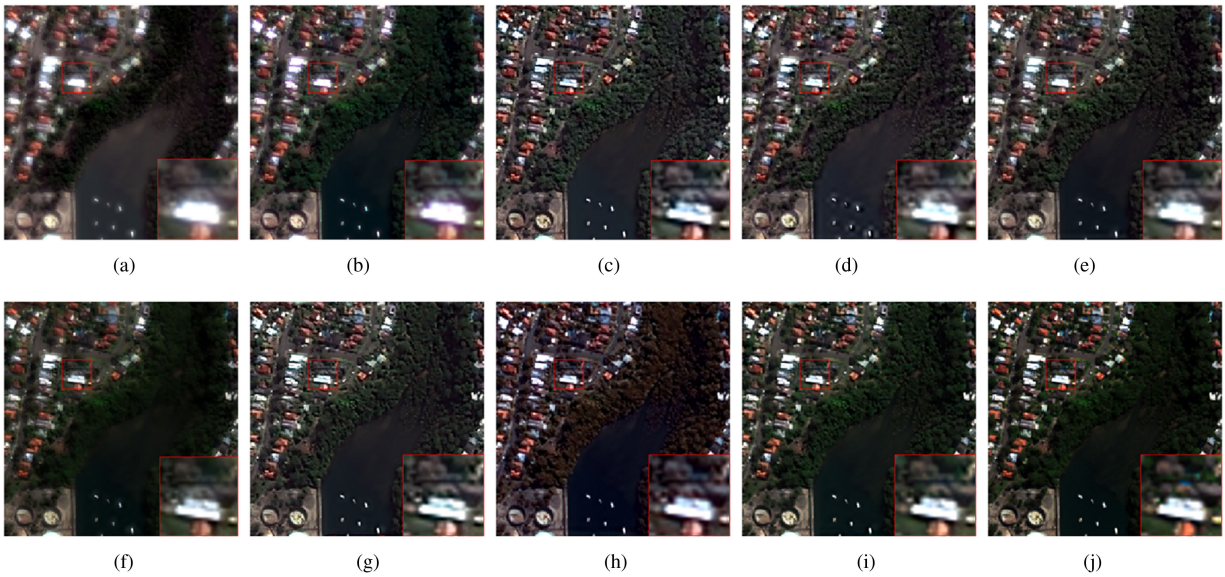

Fig. 6. Comparisons of reduce scale experiment on a WorldView-2 image (visualized using the color composite of RGB bands). (a) PCA. (b) PRACS. (c) BDSD. (d) Indusion. (e) MTF-GLP. (f) LGC. (g) GDN. (h) PNN. (i) Our. (j) Ground truth.

We adopt an alternative iteration optimization scheme for finding a local optimal solution to (11). The optimization procedure reduces by iterating between two subproblems that can be optimized individually, and their $(t+1)$th iterations are formulated as follows:

$$\begin{aligned} P1: \quad q^{t+1} &= \|\nabla_1 x^t - q\|_2^2 + \alpha\|q\|_1 \\ s^{t+1} &= \|\nabla_2 x^t - s\|_2^2 + \beta\|s\|_1 \end{aligned} \tag{12}$$

$$\begin{aligned} P2: x^{t+1} = &\|y - k * x\|_2^2 + v_1\|\nabla_1 x - f_1\|_2^2 \\ &+ v_2\|\nabla_2 x - f_2\|_2^2 + \xi\|\nabla_1 x - q^t\|_2^2 \\ &+ \eta\|\nabla_2 x - s^t\|_2^2. \end{aligned} \tag{13}$$

The update algorithms for the above two subproblems are detailed as follows.

Update for $P1$: We use a shrinkage operation to solve the update of the auxiliary variables q and s at the $(t+1)$th iteration:

$$\begin{aligned} q^{t+1} &= \text{shrink}(\nabla_1 x^t, \alpha) \\ s^{t+1} &= \text{shrink}(\nabla_2 x^t, \beta) \end{aligned} \tag{14}$$

where $\text{shrink}(x, \lambda) = \frac{x}{|x|} \cdot \max(|x| - \lambda, 0)$.

Update for $P2$: We reconstruct the ideal HRMS image x by addressing a least-squares problem $P2$ in the Fourier domain, and x has a closed-form solution that satisfies

$$\begin{aligned} &K^T(Kx - y) + v_1\nabla_1^T(\nabla_1 x - f_1) + v_2\nabla_2^T(\nabla_2 x - f_2) \\ &+ \xi\nabla_1^T(\nabla_1 x - q^t) + \eta\nabla_2^T(\nabla_2 x - s^t) = 0 \end{aligned} \tag{15}$$

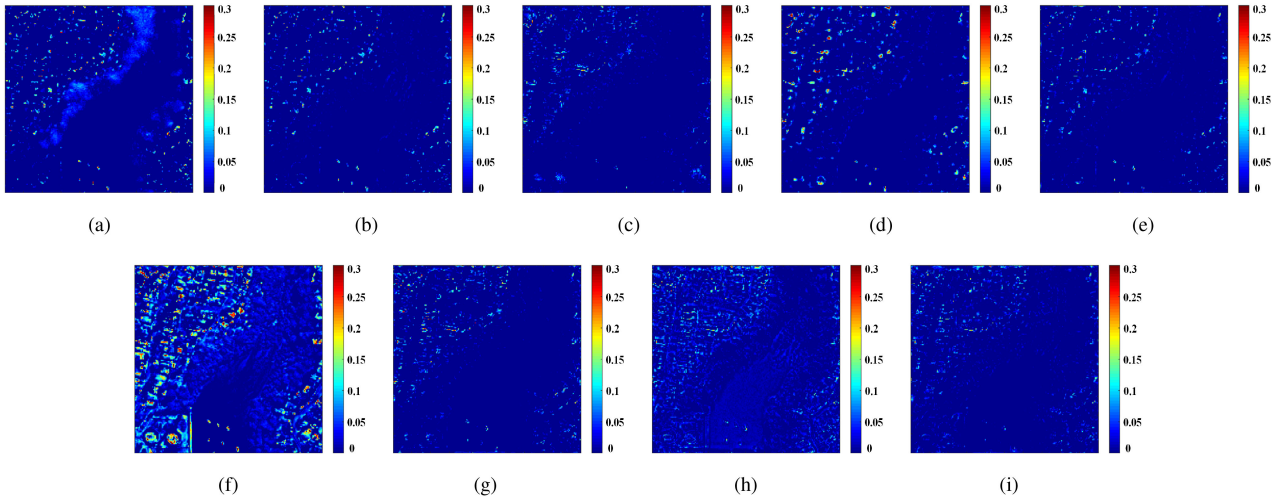

Fig. 7. Residuals between the HRMS image reconstructions and the ground truth from Fig. 6. (a) PCA. (b) PRACS. (c) BDSD. (d) Indusion. (e) MTF-GLP. (f) LGC. (g) GDN. (h) PNN. (i) Our.

Algorithm 1: Outline of the Proposed Model.

Input: LRMS image y, PAN image y_p, blur kernel k, parameters $v_1, v_2, \alpha, \beta, \eta, \xi$, multiorder gradient operations ∇^*, max iteration M.
Initialize: $x^0 = q^0 = s^0 = 0$
step1:
$f_1 = f_{\text{MCNN}}(\nabla_1 y, \nabla_1 y_p), f_2 = f_{\text{MCNN}}(\nabla_2 y, \nabla_2 y_p)$
step2:
for $t = 1$ to M
$q^{t+1} = \text{shrink}(\nabla_1 x, \alpha)$
$s^{t+1} = \text{shrink}(\nabla_2 x, \beta)$
$x^{t+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(K^T y + v_1 \nabla_1^T f_1 + v_2 \nabla_2^T f_2 + \xi \nabla_1^T q^t + \eta \nabla_2^T s^t)}{\bigwedge_1 + \bigwedge_2 (v_1 + \xi) + \bigwedge_3 (v_2 + \eta)}\right)$
end for
Output: HRMS image x.

where K respects the matrix form of the kernel k. By using the fast Fourier transform (FFT) to speed up the solving process, the solution x is derived as follows:

$$x^{t+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(K^T y + v_1 \nabla_1^T f_1 + v_2 \nabla_2^T f_2 + \xi \nabla_1^T q^t + \eta \nabla_2^T s^t)}{\bigwedge_1 + \bigwedge_2 (v_1 + \xi) + \bigwedge_3 (v_2 + \eta)}\right) \tag{16}$$

where $\mathcal{F}$ is the FFT operator, $\mathcal{F}^{-1}$ denotes the inverse operator of $\mathcal{F}$, and $\bigwedge_1$, $\bigwedge_2$, and $\bigwedge_3$ represent the eigenvalues of $K^T K$, $\nabla_1^T \nabla_1$, and $\nabla_2^T \nabla_2$, respectively. We summarize the main steps of the proposed method in Algorithm 1.

IV. Experiment Results

We will provide numerous experiments to demonstrate the effectiveness of the proposed method. All experiments are implemented using MATLAB 2018a on a desktop computer with Intel Core i7-8700, 16-GB RAM, and 64-bit Windows10 operating system. Simultaneously, the learning phase of our inference network is carried on GPU that is NVIDIA GTX1080Ti with CUDA 9.0 through the DL platform Caffe [35]. The datasets of our experiments are acquired from WorldView-2 and QuickBird. Our MCNN model cannot be trained directly when lacking the HRMS images at the original scale. To solve the problem, we follow the Wald protocol [36] for our network training and experiment simulation. In Fig. 5, we provide a pictorial workflow of generating the training datasets based on the Wald protocol. In the workflow, we smooth the MS and PAN images with a Gaussian smoothing kernel [46], and we downsample the smoothed component by a factor of 4; the original MS image is regarded to be the HRMS image. Then, we extract the multiorder gradient component by using the multigradient operator from the downsampled MS and PAN images, respectively. The multiorder gradient component of the downsampled MS is upsampled with the bicubic interpolation to obtain the LRMS; accordingly, the gradient of the HRMS is the corresponding ground truth. The size of training/validation patches is set to be 32×32, our inference network is trained for 2.5×10^5 iterations, the momentum is set to 0.9, and the base learning rate is set to 10^{-3} and is divided by 10 at 10^5 and 2×10^5 iterations. The training process of our network costs roughly 4 h. According to satisfactory pan-sharpening results, we empirically set the same parameters for all experiments to prove the stability of the proposed algorithm: $v_1 = 0.4, v_2 = 0.8, \zeta = \eta = 0.001,$ and $\alpha = \beta = 100$.

We compare our method with several pan-sharpening methods: three CS-based methods, i.e., PCA [6], band-dependent spatial detail with local parameter estimation (BDSD) [18], partial replacement adaptive component substitution (PRACS) [37]; three MRA-based methods, i.e., decimated wavelet transform using an additive injection model (Indusion) [44], comparison of pan-sharpening algorithms: Outcome of the 2006 GRS-S Data-Fusion Contest (MIF-GLP) [43]; two OV-based methods, i.e., variational pan-sharpening with local gradient constraints (LGC) [22], pan-sharpening via a gradient-based deep network

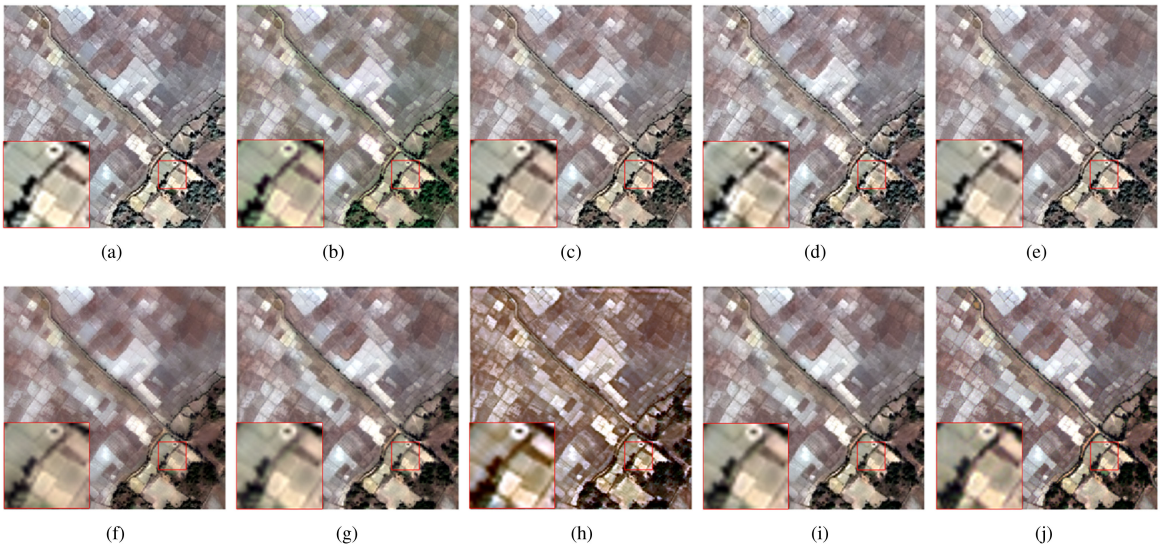

Fig. 8. Comparisons of the reduced-scale experiment on a QuickBird image (visualized using the color composite of RGB bands). (a) PCA. (b) PRACS. (c) BDSD. (d) Indusion. (e) MTF-GLP. (f) LGC. (g) GDN. (h) PNN. (i) Our. (j) Ground truth.

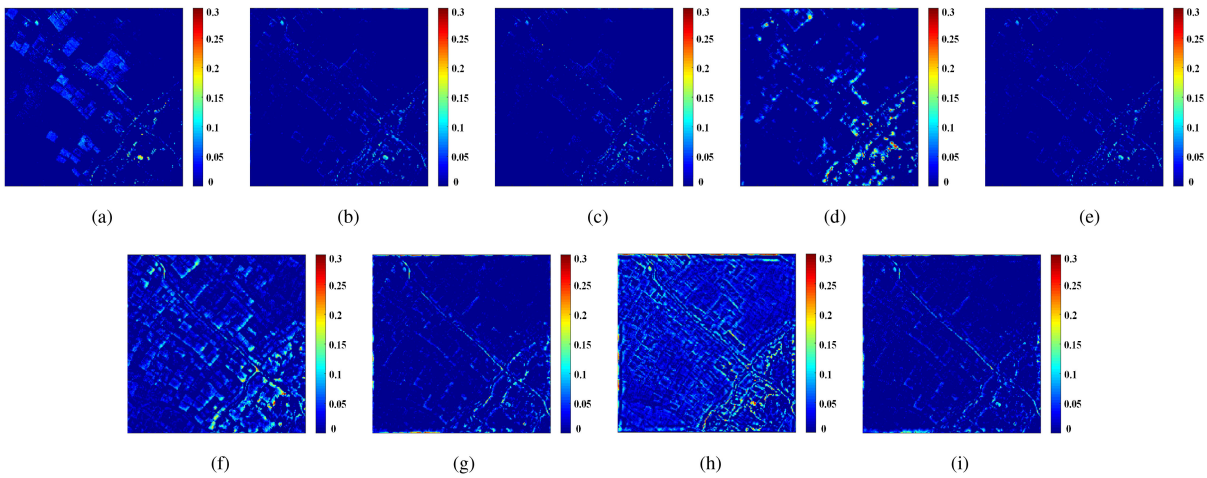

Fig. 9. Residuals between the HRMS image reconstructions and the ground truth from Fig. 8. (a) PCA. (b) PRACS. (c) BDSD. (d) Indusion. (e) MTF-GLP. (f) LGC. (g) GDN. (h) PNN. (i) Our.

prior (GDN) [29]; and PNN [15]. We introduce both qualitative results and quantitative metrics to assess the pan-sharpened images using different methods. For reduced-scale experiments, the quantitative metrics include spectral angle mapper (SAM) [38], universal image quality index averaged over the bands (QAVE) [40], Q8 for eight-band image, Q4 for four-band image, relative dimensionless global error in synthesis (ERGAS) [39], and spatial correlation coefficient (SCC) [41]. And we use the quality without reference (QNR) [42], which is composed of the spectral distortion index D_λ and the spatial distortion index D_s to assess the pan-sharpened images in original-scale experiments. In particular, the ideal values for ERGAS, SAM, D_λ, and D_s are 0, whereas the ideal values for QAVE, SCC, Q4, Q8, and QNR are 1. The results of our all experiments are listed in Tables I–V, and in each comparison group, the best performance is marked in bold; the second best is labeled as underline. For visualization, we show the RGB bands of the pan-sharpened images but conduct experiments in all spectral bands.

TABLE I
QUALITY METRICS OF DIFFERENT METHODS ON A WORLDVIEW-2 DATASET (BOLD: THE BEST; UNDERLINE: THE SECOND BEST)

	Q	SAM	ERGAS	SCC	Q8
PCA	0.6733	8.1513	6.3220	0.8310	0.7070
PRACS	0.7707	6.5198	4.9398	0.8018	0.7864
BDSD	0.8294	6.8618	4.4390	0.8771	0.8378
Indusion	0.7723	6.6630	5.7619	0.8320	0.7678
MIF-GLP	0.8217	6.1887	4.5261	0.8789	0.8264
LGC	0.7901	5.7793	5.0304	0.8786	0.7960
GDN	0.8539	5.5575	4.0437	0.9122	0.8405
PNN	0.8559	6.0539	3.8063	0.9166	0.8446
Proposed	**0.8858**	**4.5310**	**3.1765**	**0.9438**	**0.8747**
Reference	1	0	0	1	1

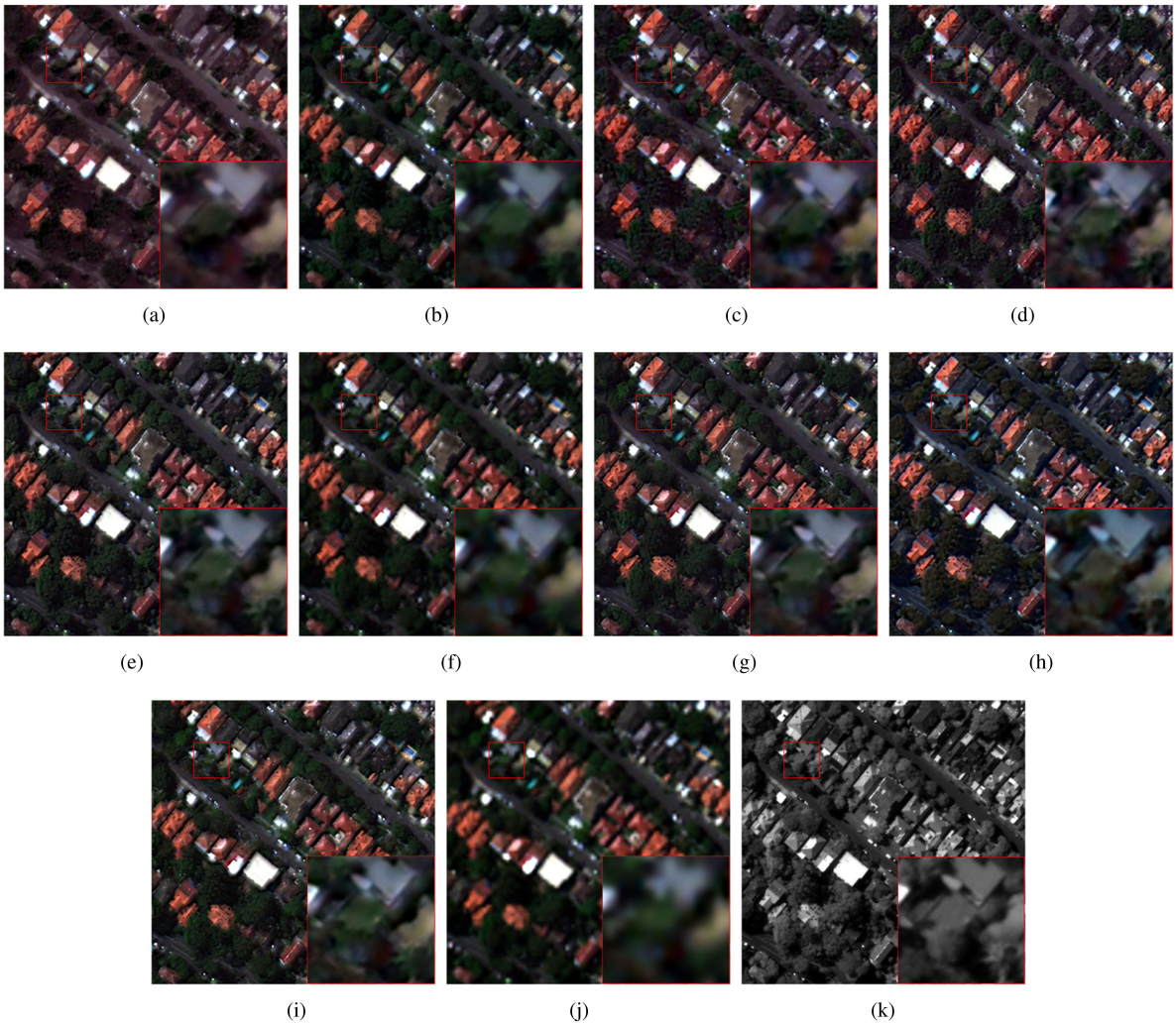

Fig. 10. Comparisons of the original-scale experiment on a WorldView-2 image (visualized using the color composite of RGB bands). (a) PCA. (b) PRACS. (c) BDSD. (d) Indusion. (e) MTF-GLP. (f) LGC. (g) GDN. (h) PNN. (i) Our. (j) LRMS. (k) PAN.

TABLE II
QUALITY METRICS OF DIFFERENT METHODS ON A QUICKBIRD DATASET (BOLD: THE BEST; UNDERLINE: THE SECOND BEST)

	Q	SAM	ERGAS	SCC	Q4
PCA	0.8389	2.9378	2.8175	0.9368	0.8176
PRACS	0.8877	2.6247	2.0704	0.9439	0.8841
BDSD	0.8933	2.5817	2.0556	0.9470	0.8841
Indusion	0.8185	2.9294	3.2058	0.8925	0.8065
MIF-GLP	0.8860	2.5410	2.1250	0.9430	0.8816
LGC	0.8470	2.5517	2.6892	0.9272	0.8496
GDN	0.9086	2.2495	1.9908	0.9475	0.8832
PNN	0.8859	3.0532	2.8033	0.9261	0.8641
Proposed	**0.9165**	**2.0270**	**1.7467**	**0.9606**	**0.9019**
Reference	1	0	0	1	1

TABLE III
QUALITY METRICS OF DIFFERENT METHODS ON A WORLDVIEW-2 DATASET AT THE ORIGINAL SCALE (BOLD: THE BEST; UNDERLINE: THE SECOND BEST)

	D_λ	D_s	QNR
PCA	0.1087	0.2405	0.6877
PRACS	0.0855	0.1979	0.7417
BDSD	0.1302	0.1557	0.7413
Indusion	0.1294	0.4092	0.5051
MIF-GLP	0.1700	0.2400	0.6383
LGC	**0.0258**	0.3787	0.6051
GDN	0.1166	0.1688	0.7418
PNN	0.1050	0.3832	0.5459
Proposed	0.0939	**0.1517**	**0.7733**
Reference	0	0	1

A. *Reduced-Scale Experiment*

The Wald protocol is used in the simulated experiments due to the lack of HRMS images. On the basis of this protocol, we degrade the original MS images to yield the input images, and the original MS image is regarded to be a ground truth, which will be used to compare with the pan-sharpened images. We simulate our pan-sharpening scheme on 100 images from WorldView-2 and QuickBird, respectively.

1) WorldView-2 Data: We take a dataset acquired from the WorldView-2 sensor, which has eight bands. We first crop

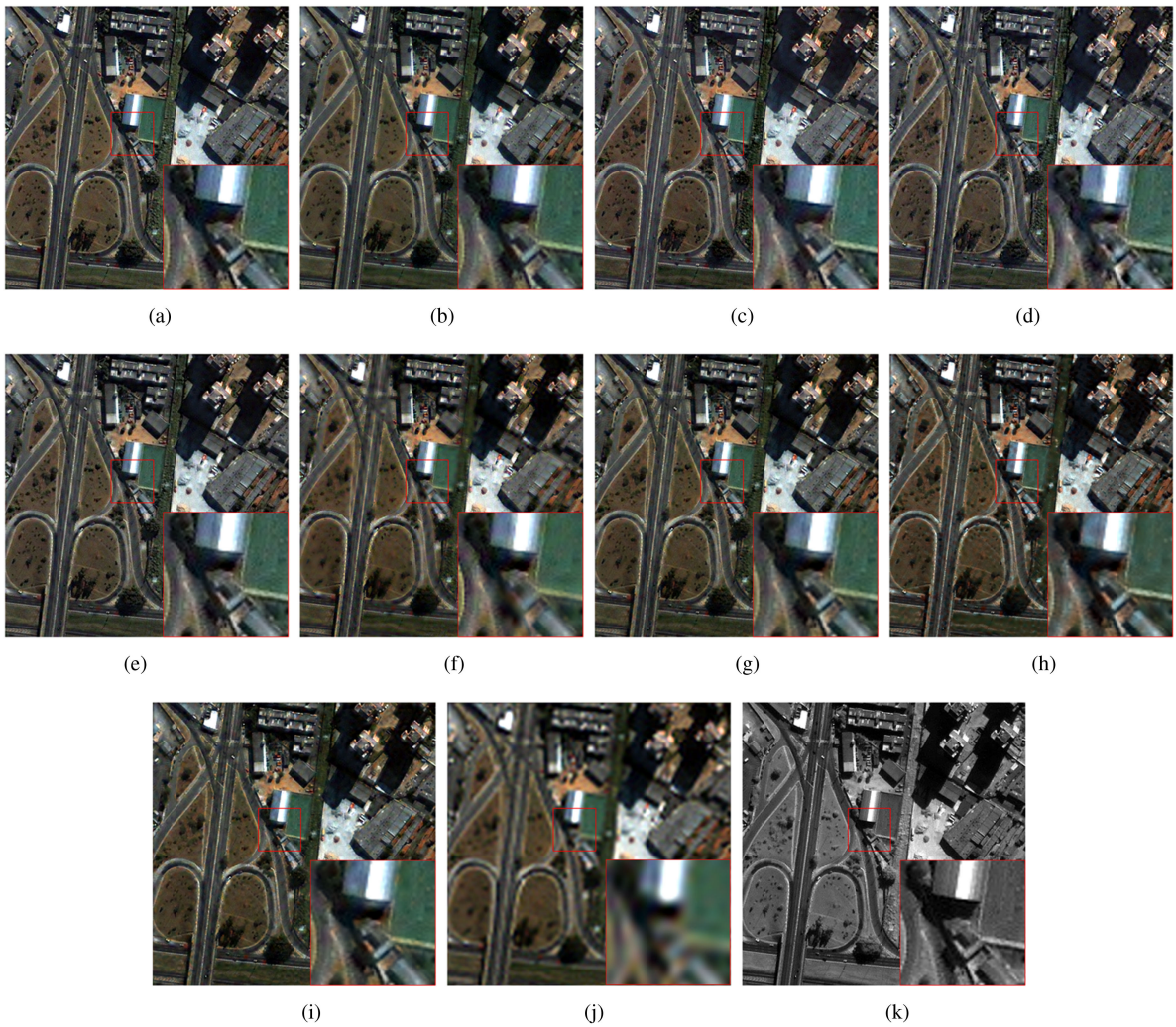

Fig. 11. Original-scale experiment on a QuickBird image (visualized using the color composite of RGB bands). (a) PCA. (b) PRACS. (c) BDSD. (d) Indusion. (e) MTF-GLP. (f) LGC. (g) GDN. (h) PNN. (i) Our. (j) LRMS. (k) PAN.

TABLE IV
QUALITY METRICS OF DIFFERENT METHODS ON A QUICKBIRD DATASET AT THE ORIGINAL SCALE (BOLD: THE BEST; UNDERLINE: THE SECOND BEST)

	D_λ	D_s	QNR
PCA	0.1339	0.2414	0.6603
PRACS	0.1052	0.2247	0.6960
BDSD	0.1054	0.2148	0.7063
Indusion	0.1768	0.4813	0.4138
MIF-GLP	0.1625	0.2428	0.6388
LGC	**0.0338**	0.5424	0.4420
GDN	0.1267	<u>0.1845</u>	<u>0.7167</u>
PNN	0.1120	0.2813	0.6412
Proposed	<u>0.0832</u>	**0.1449**	**0.7873**
Reference	0	0	1

TABLE V
QUALITY METRICS OF DIFFERENT VARIANTS OF THE PROPOSED METHOD ON A WORLDVIEW-2 DATASET (BOLD: THE BEST; UNDERLINE: THE SECOND BEST)

	Q	SAM	ERGAS	SCC	Q8
-RL	0.8776	4.6772	3.3294	0.9405	0.8698
-DF	<u>0.8811</u>	<u>4.5996</u>	<u>3.2636</u>	<u>0.9410</u>	<u>0.8724</u>
-BN	0.8594	5.1812	3.7765	0.9213	0.8541
Proposed	**0.8858**	**4.5310**	**3.1765**	**0.9438**	**0.8747**
Reference	1	0	0	1	1

$200 \times 200 \times 8$ part of the multispectral images to be the ground truth; then, we generate the LRMS images with size of $50 \times 50 \times 8$ by using the Wald protocol and simulate the PAN images with size of 200×200. Fig. 6 shows the pan-sharpened results using different methods, and their corresponding residuals are shown in Fig. 7. From Figs. 6 and 7, it is clear that LGC and PCA suffer from different blurring artifacts. GDN, PNN, and BDSD preserve spatial information well, but they fail in protecting spectral information and lead to distinct color distortions. PRACS and Indusion have different levels of spatial distortions. MIF-GLP is relatively well balanced in spatial and spectral details, but its residual image is less competitive to that of our method. The proposed method performs outstanding spectral and spatial fusion. For the quantitative evaluation, we present

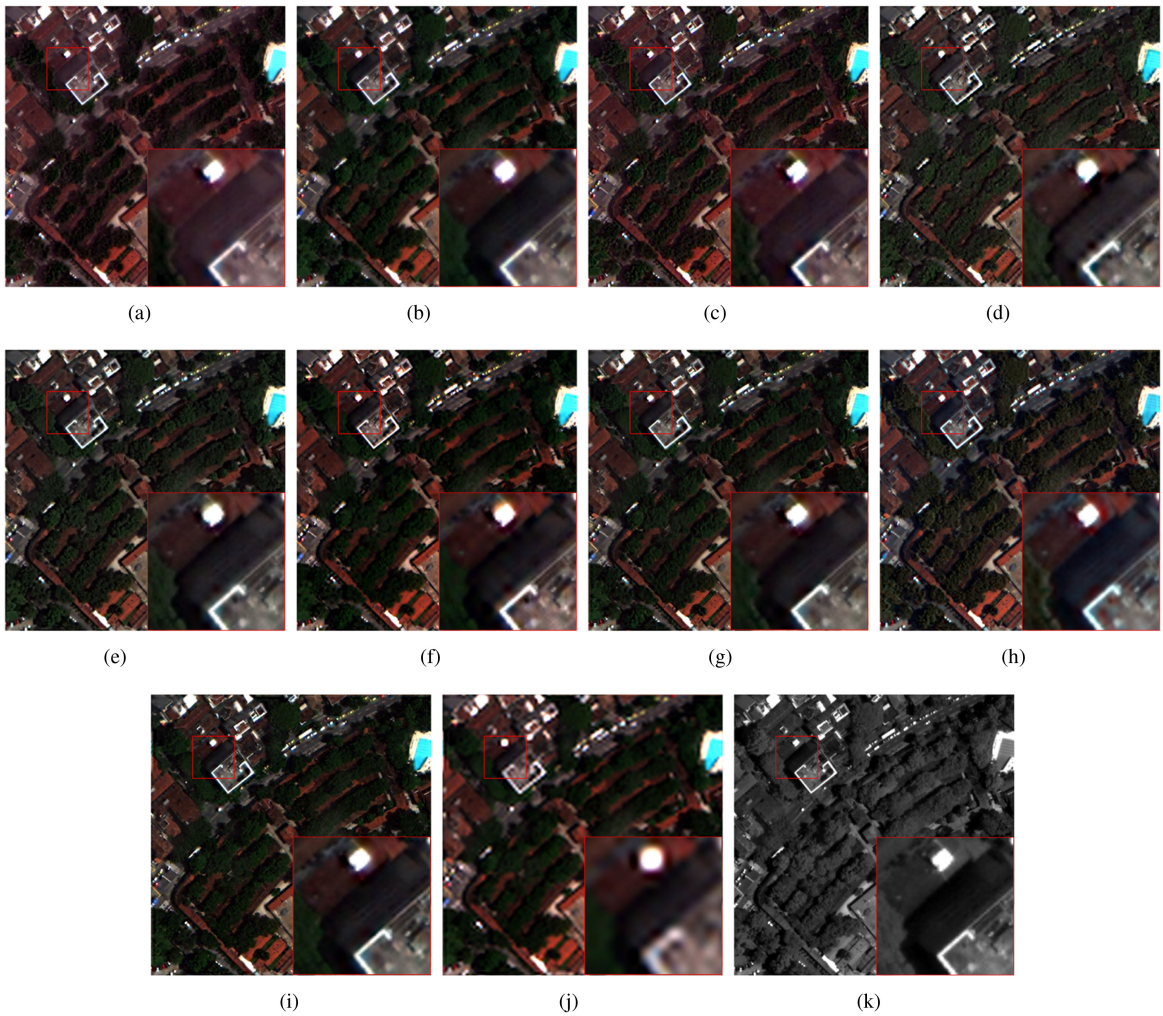

Fig. 12. Original-scale experiment on a WorldView-3 image (visualized using the color composite of RGB bands). (a) PCA. (b) PRACS. (c) BDSD. (d) Indusion. (e) MTF-GLP. (f) LGC. (g) GDN. (h) PNN. (i) Our. (j) LRMS. (k) PAN.

the mean of 100 images using different methods in Table I, and it is observed that our method yields better results of all metrics compared with other methods.

2) QuickBird Data: We test a dataset acquired from the QuickBird sensor with four bands. The dataset is cropped by $200 \times 200 \times 4$ part of the multispectral images as the ground truths, and we generate the LRMS images with the size of $50 \times 50 \times 4$ by using the same way as the WorldView-2 dataset and generate the PAN images with the size of 200×200. Fig. 8 illustrates the pan-sharpened results of different approaches, and their corresponding residuals are shown in Fig. 9. We find that PCA, PNN, and PRACS produce serious color distortions, although image spatial information is well preserved, and LGC introduces some blurring artifacts. Indusion can protect image spectral information well and fail to preserve image spatial information. GDN, MIF-GLP, and BDSD show a good ability to preserve spatial and spectral information, but our method holds better spectral information than these competitive methods. In Table II, we report the mean of 100 images from the QuickBird sensor by using different methods. It is noticeable that our method yields more promising results of all metrics.

B. Original-Scale Experiment

We evaluate different pan-sharpening methods at the original-scale images acquired from WorldView-2 and QuickBird satellites, respectively. Since our inference network is trained in reduced scale, for the sake of assessing the ability of transferring to the original scale, the raw MS and PAN images are input into our method to yield full-resolution results.

We show the pan-sharpened results using different methods on the eight-band satellite image in Fig. 10. From Fig. 10, we can see that PCA and PNN have spectral distortions similar to their performance in reduced-scale experiments. BDSD and MIF-GLP lose the spectral information compared with other methods. On the contrary, LGC, Indusion, and PRACS generate different blurring artifacts. GDN performs a good tradeoff between spectral and spatial preservations. Furthermore, our method has better spectral and spatial fusion than GDN. In Table III, we show the mean values of 50 images from the WorldView-2 satellite by using different methods. It is combined with Table III that our method is well balanced in spectral and spatial preservations, and our fusion results have more promising performance.

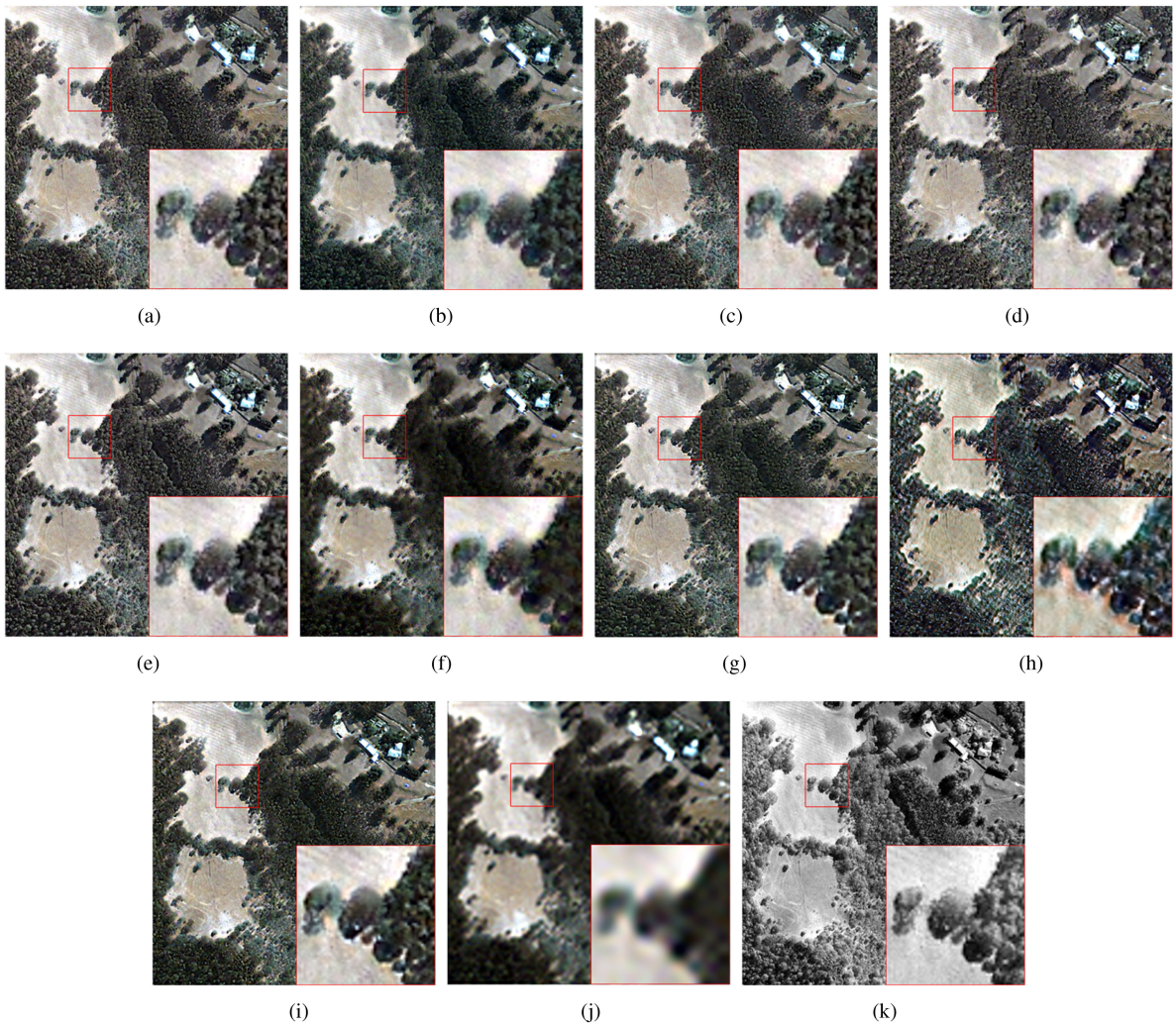

Fig. 13. Original-scale experiment on an IKONOS image (visualized using the color composite of RGB bands). (a) PCA. (b) PRACS. (c) BDSD. (d) Indusion. (e) MTF-GLP. (f) LGC. (g) GDN. (h) PNN. (i) Our. (j) LRMS. (k) PAN.

Meanwhile, we show the fused results of different methods on a four-band satellite image in Fig. 11. By comparing the visual results in Fig. 11, PCA, PRACS, and BDSD produce some spectral distortions. LGC, Indusion, MIF-GLP, and PNN generate blurring edges. GDN has the competitive visual results of sharp edges and spectral information, but it cannot outperform our method. We evaluate the average values of 50 images acquired from the QuickBird satellite for different methods in Table IV, and it is observed that the proposed method yields results of better spectral and spatial fusion compared with other methods in four-band satellite images.

C. Extension

To demonstrate the generalization of our model across different satellites, we perform on the WorldView-3 and IKONOS images by using our network parameters trained on WorldView-2 and QuickBird datasets. We present the visual results at the original scale in Figs. 12 and 13. As can be seen from Figs. 12 and 13, the proposed method consistently yields better spectral and spatial fusion. PCA and PNN suffer from obvious color distortions. BDSD, Indusion, and MTF-GLP exhibit distinct degradations in the spectral domain. LGC introduces clear blurring artifacts. Although GDN and PRACS perform a stabilization in spectral and spatial domains, they cannot outperform our method. Experimental results prove that our model obtains superior robustness across different satellites.

D. Ablation Study

To demonstrate the effect of each component of our inference network, we compare variants of the proposed network and the following statements hold: 1) the MCNN removes residual learning operation (-RL); 2) the MCNN removes dilated convolution operation (-DF); and 3) the MCNN removes batch normalization operation (-BN). We carry out the test on the above WorldView-2 dataset in reduced scale. In Table V, one can see that residual learning operation, dilated convolution operation, and BN operation can improve quantitative pan-sharpened performance, and our inference network composed of these three components outperforms the three variants (-RL, -DF, and -BN) in all quality metrics.

V. Conclusion

In this article, we have proposed a Bayesian pan-sharpening model, which contains the spectral preservation term, the spatial preservation term, and the prior term. We attempt to design an MCNN to enforce the spatial fusion of PAN and LRMS images in a nonlinear manner, and our inference network is composed of efficient multiscale recursive blocks. And our CNN output is integrated into a variational-based optimization framework for pan-sharpening. In addition, a multiorder gradient prior is introduced to reconstruct better edges and details. Compared with several leading pan-sharpening methods at reduced and original resolutions, our method yields better results of both structural fusion and spectral preservation. And our model is more generalization to different satellites, since the proposed CNN is performed in multiorder gradient domains.

References

[1] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, Jul. 2014.

[2] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu, "Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges," *Inf. Fusion*, vol. 46, pp. 102–113, Mar. 2019.

[3] 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.

[4] W. Carper, T. Lillesand, and R. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, 1990.

[5] Q. Xu, B. Li, Y. Zhang, and L. Ding, "High-fidelity component substitution pansharpening by the fitting of substitution data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7380–7392, Nov. 2014.

[6] H. R. Shahdoosti and H. Ghassemian, "Combining the spectral PCA and spatial PCA fusion methods by an optimal filter," *Inf. Fusion*, vol. 27, pp. 150–160, Jan. 2016.

[7] P. Kwarteng and A. Chavez, "Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 3, pp. 339–348, 1989.

[8] S. G. 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.

[9] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*. New York, NY, USA: Springer, Jan. 1995, pp. 281–299.

[10] L. Alparone, S. Baronti, B. Aiazzi, and A. Garzelli, "Spatial methods for multispectral pansharpening: Multiresolution analysis demystified," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2563–2576, May 2016.

[11] R. Restaino, M. D. Mura, G. Vivone, and J. Chanussot, "Context-adaptive pansharpening based on image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 753–766, Dec. 2016.

[12] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Roug, "A variational model for P+ XS image fusion," *Int. J. Comput. Vis.*, vol. 69, no. 1, pp. 43–58, Aug. 2006.

[13] F. Fang, F. Li, C. Shen, and G. Zhang, "A variational approach for pansharpening," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2822–2834, Apr. 2013.

[14] T. Wang, F. Fang, F. Li, and G. Zhang, "High-quality Bayesian pansharpening," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 227–239, Jan. 2019.

[15] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, Jul. 2016, Art. no. 594.

[16] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multidepth 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.

[17] 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. Vis.*, Oct. 2017, pp. 5449–5457.

[18] 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.

[19] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.

[20] G. Vivone *et al.*, "Pansharpening based on semiblind deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1997–2010, Apr. 2015.

[21] Y. Jiang, L. Chen, W. Wang, X. Ding, and Y. Huang, "A compressed sensing-based pan-sharpening using joint data fidelity and blind blurring kernel estimation," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 5042–5046.

[22] X. Fu, Z. Lin, Y. Huang, and X. Ding, "A variational pan-sharpening with local gradient constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10265–10274.

[23] X. Wang, "Laplacian operator-based edge detectors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 886–890, May 2007.

[24] Y. Weiss and W. T. Freeman, "What makes a good model of natural images?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[25] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Jan. 2009, pp. 1033–1041.

[26] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, 2008.

[27] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease," in *Reconstruction, Segmentation and Analysis of Medical Images*. Cham, Switzerland: Springer, Apr. 2017, pp. 95–102.

[28] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3929–3938.

[29] F. Ye, Y. Guo, and P. Zhuang, "Pan-sharpening via a gradient-based deep network prior," *Signal Process., Image Commun.*, vol. 74, pp. 322–331, 2019.

[30] D. Lee, J. Yoo, S. Tak, and J. C. Ye, "Deep residual learning for accelerated MRI using magnitude and phase networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1985–1995, Sep. 2018.

[31] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, Jul. 2018, pp. 286–301.

[32] S. Lefkimmiatis, "Universal denoising networks: A novel CNN architecture for image denoising," in *Proc. IEEE Conf. Comput. Vis.Pattern Recognit.*, Nov. 2017, pp. 3204–3213.

[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Int. Conf. Mach. Learn.*, vol. 37, pp. 448–456, Jul. 2015.

[34] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, Jun. 2010, pp. 807–814.

[35] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 675–678.

[36] L. Wald and T. Ranchin, "Fusion of images and raster-maps of different spatial resolutions by encrustation: An improved approach," *Comput., Environ. Urban Syst.*, vol. 19, no. 2, pp. 77–87, Nov. 1995.

[37] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, Jan. 2011.

[38] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.

[39] L. Wald, *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*, Paris, France: École des Mines de Paris, 2002.

[40] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 313–317, Oct. 2004.

[41] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, 1998.

[42] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.

[43] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.
[44] M. M. Khan, J. Chanussot, L. Condat, and A. Montanvert, "Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 98–102, Jan. 2008.
[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Aug. 2016, pp. 770–778.
[46] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.
[47] C. Padwick, M. Deskevich, F. Pacifici, and S. Smallwood, "WorldView-2 pan-sharpening," in *Proc. ASPRS Annu. Conf.*, San Diego, CA, USA, Apr. 2010, vol. 2630.
[48] Y. Jiang, X. Ding, D. Zeng, Y. Huang, and J. Paisley, "Pan-sharpening with a hyper-Laplacian penalty," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 540–548.

Penghao Guo is working toward a postgraduate degree with the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China.

His research interests include image processing and Bayesian machine learning.

Peixian Zhuang received the Ph.D. degree in sparse representation, compressed sensing, and Bayesian machine learning from Xiamen University, Xiamen, China, in 2016.

He is currently a Lecturer with the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China. His research interests include image processing and Bayesian machine learning.

Dr. Zhuang was the recipient of the Best Ph.D. Thesis Award in Fujian Province in 2017.

Yecai Guo received the Ph.D. degree in underwater acoustic engineering from Northwestern Polytechnical University, Xi'an, China, in 2003.

He is currently a Professor with the Nanjing University of Information Science and Technology, Nanjing, China. His research interests include meteorological telecommunication technology, underwater communication theory, and their applications.

Dr. Guo was a National Winner of National Outstanding Doctoral Dissertations in 2006, the training object of "Six Talents Peak" in Jiangsu province in 2008, and one of the Leaders of the advantage discipline "sensor network and modern meteorological equipment" of Colleges and Universities of Jiangsu Province in 2009.