Variational Pansharpening Based on Coefficient Estimation With Nonlocal Regression

Jin-Liang Xiao, Ting-Zhu Huang, *Member, IEEE*, Liang-Jian Deng, *Senior Member, IEEE*, Zhong-Cheng Wu, Xiao Wu, and Gemine Vivone, *Senior Member, IEEE*

***Abstract*— Pansharpening (which stands for panchromatic (PAN) sharpening) involves the fusion between a multispectral (MS) image with a higher spectral content than a fine spatial resolution PAN image to generate a high spatial resolution MS (HRMS) image. A widely used concept is the construction of the relationship between PAN and HRMS images by designing pixel-based coefficients. Previous pixel-based methods compute the coefficients pixel-by-pixel while suffering from inaccuracies in some areas leading to spatial distortion. However, we found that the coefficients inherit the spatial properties of the HRMS image, e.g., the local smoothness and nonlocal self-similarity, and the spatial correlation between the coefficients and the HRMS image can increase the accuracy of the estimation process. In this article, we propose a novel spatial fidelity with nonlocal regression (SFNLR) to describe the relationship between PAN and HRMS images. Unlike from the pixel-based perspective, the SFNLR can jointly use the local smoothness and nonlocal self-similarity of the coefficients for preserving spatial information. Besides, the SFNLR is integrated with a widely used spectral fidelity to formulate a new variational model for the pansharpening problem. An effective algorithm based on the alternating direction method of multiplier (ADMM) framework is designed to solve the proposed model. Qualitative and quantitative assessments on reduced and full resolution datasets from different satellites demonstrate that the proposed approach outperforms several state-of-the-art methods. The code is available at: https://github.com/Jin-liangXiao/SFNLR.**

***Index Terms*— Coefficient estimation, local smoothness, nonlocal self-similarity, pansharpening, remote sensing, variational models.**

NOMENCLATURE

Notations	Explanations
$\mathcal{X}, \mathbf{X}, \mathbf{x}, x$	Tensor, matrix, vector, scalar.
$\mathbf{X}_{(3)}$	Mode-3 unfolding of $\mathcal{X}$.
$\mathcal{X} \in \mathbb{R}^{H\times W\times S}$	HRMS image.
$\mathcal{Y} \in \mathbb{R}^{h\times w\times S}$	LRMS image.
$\widetilde{\mathcal{Y}} \in \mathbb{R}^{H\times W\times S}$	Upsampled LRMS image.
$\mathbf{P} \in \mathbb{R}^{H\times W}$	PAN image.
$\mathcal{P} \in \mathbb{R}^{H\times W\times S}$	Extended PAN image.
$\mathcal{P}_L \in \mathbb{R}^{H\times W\times S}$	Low-pass version of the extended PAN image.
$\mathcal{G} \in \mathbb{R}^{H\times W\times S}$	Adaptive coefficient.
$\odot$	Hadamard product.
$\oslash$	Elementwise division.

Manuscript received 24 May 2023; revised 4 July 2023; accepted 30 July 2023. Date of publication 15 August 2023; date of current version 29 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 12171072 and Grant 12271083, in part by the Natural Science Foundation of Sichuan Province under Grant 2022NSFSC0501, in part by the Key Projects of Applied Basic Research in Sichuan Province under Grant 2020YJ0216, and in part by the National Key Research and Development Program of China under Grant 2020YFA0714001. *(Corresponding authors: Ting-Zhu Huang; Liang-Jian Deng.)*

Jin-Liang Xiao, Ting-Zhu Huang, Liang-Jian Deng, Zhong-Cheng Wu, and Xiao Wu are with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China (e-mail: jinliang_xiao@163.com; tingzhuhuang@126.com; liangjian.deng@uestc.edu.cn; wuzhch97@163.com; wxwsx1997@gmail.com).

Gemine Vivone is with the Institute of Methodologies for Environmental Analysis, CNR-IMAA, 85050 Tito Scalo, Italy, and also with the National Biodiversity Future Center (NBFC), 90133 Palermo, Italy (e-mail: gemine.vivone@imaa.cnr.it).

Digital Object Identifier 10.1109/TGRS.2023.3305296

I. INTRODUCTION

WITH the rapid progress of satellite technology, remote sensing satellites can provide accurate reproductions of the globe surface [1]. Due to physical constraints, images with both high spectral and spatial resolution cannot be directly obtained. Thus, satellite systems, e.g., GaoFen-2, QuickBird (QB), and WorldView series, simultaneously capture low spatial resolution MS (LRMS) images and high spatial resolution panchromatic (PAN) images. PAN sharpening (pansharpening) refers to fusing the above-mentioned pairs to obtain high spatial resolution MS (HRMS) images.

Recently, pansharpening has gained comprehensive attention, and numerous methods have been proposed. These methods are approximately divided into four classes [4], [5], [6], [7], [8]: 1) component substitution (CS) methods; 2) multiresolution analysis (MRA) approaches; 3) deep learning (DL) methods; and 4) variational optimization (VO) techniques.

The CS methods separate the spatial information from different components by projecting the LRMS image into a transformed space. Subsequently, the spatial component of the LRMS image is substituted by the PAN image. Typical methods include the Brovey transform (BT) [2], the principal component analysis (PCA) [9], the intensity–hue–saturation (IHS) [10], the Gram–Schmidt (GS) spectral sharpening [11], the partial replacement adaptive CS (PRACS) [12], the band-dependent spatial detail (BDSD) [13], and GS adaptive (GSA) method [14]. In general, this class of methods tends to preserve spatial details but producing spectral distortion [15], [16].

The MRA approaches focus on complementing the LRMS image using spatial details of the PAN image, which are obtained by the direct implementation of some decompositions, e.g., smoothing filter-based intensity modulation

(SFIM) [17], decimated wavelet transform (DWT) [18], undecimated wavelet transform (UDWT) [19], "à-trous" wavelet transform (ATWT) [20], generalized Laplacian pyramid (GLP) [21], and the GLP with modulation transfer function (MTF)-matched filter and its high-pass modulation version (GLP-HPM) [3]. Unlike from CS methods, MRA approaches show superiority in preserving spectral information. Nevertheless, this superiority is usually paid by an increase in the spatial distortion [22].

In recent years, DL methods have become increasingly popular [23], [24], [25], [26], [27]. Because of the powerful feature extraction capability of convolutional neural networks, DL methods have achieved excellent results in many tasks of image processing [28], [29], [30], [31], [32], [33]. However, most DL methods, e.g., [34], [35], [36], [37], [38], [39], rely on a corpus of training data. The results may be affected by the lack of training samples, and the interpretability is often absent [40], [41], [42], [43]. Although unsupervised learning is adopted to explore the intrinsic characteristics without depending on the simulated datasets, the results of these methods heavily rely on the loss function [4], [44], and the performance is weakened compared with the supervised learning approaches.

The VO techniques generate the fused image by designing the optimization model between the HRMS image and the two inputs, i.e., the LRMS and PAN data. These techniques have shown superior ability in extracting spectral and spatial features [5], [45], [46], [47]. Ballester et al. [48] proposed the first variational model by assuming that the PAN image is the linear combination of the different bands of the HRMS image, which describes the global relationship between PAN and HRMS images. Subsequently, Fu et al. [49] proposed a variational pansharpening model with local spatial constraints. In addition, some techniques, e.g., [50], use different norms to constrain the spatial details. Various transformed domains, e.g., [51], [52], [53], were also considered. These methods were proposed by exploring effective fidelity terms, which have demonstrated great potential for the pansharpening problem [54], [55], [56]. Hence, we claim that the critical problem of VO techniques is how to build the appropriate fidelity terms.

Many existing works, e.g., [2], [57], [58], established the spatial fidelity by treating the HRMS image as a Hadamard product of the coefficients and the PAN image, i.e., $\mathcal{X} = \mathcal{G} \odot \mathcal{P}$, where $\mathcal{X} \in \mathbb{R}^{H\times W\times S}$ denotes the HRMS image, $\mathcal{G} \in \mathbb{R}^{H\times W\times S}$ is the coefficient, $\odot$ means the Hadamard product, and $\mathcal{P} \in \mathbb{R}^{H\times W\times S}$ is the extended PAN image (the fidelity is graphically shown in the first row of Fig. 1). Due to the flexibility of the coefficients, this concept has shown great potential in the pansharpening problem. However, these works only focused on the relationship between the PAN and HRMS images at pixel level, thus neglecting the latent properties of the coefficients. For example, two classic methods, i.e., BT [2] and GLP-HPM [3], respectively, explore the relationship by $\mathcal{G} = \widetilde{\mathcal{Y}} \oslash \mathcal{I}_L$ and $\mathcal{G} = \widetilde{\mathcal{Y}} \oslash \mathcal{P}_L$, where $\widetilde{\mathcal{Y}}$ is the upsampling LRMS image, $\mathcal{P}_L$ is the low-pass version of $\mathcal{P}$, $\mathcal{I}_L$ is the weighted average of each band of $\widetilde{\mathcal{Y}}$, and $\oslash$ denotes the elementwise division. Obviously, $\mathcal{G}(i, j, k)$ of BT [2] and GLP-HPM [3] are independently calculated by $\widetilde{\mathcal{Y}}(i, j, k) \oslash \mathcal{I}_L(i, j, k)$ and $\widetilde{\mathcal{Y}}(i, j, k) \oslash \mathcal{P}_L(i, j, k)$, respectively. The information of $\mathcal{G}$ at other positions [except (i, j, k)] is excluded. Thus, the pixel-based coefficient estimation is not generally aligned in some areas, especially along images' edges [see Fig. 1(a) and (b)]. Furthermore, Fig. 2 shows a toy example exploring the underlying properties of the coefficients. We find that the coefficients inherit the spatial characteristics from the HRMS image. Since the local smoothness and nonlocal similarity have been proven to be useful for retaining spatial information, e.g., structure and texture information in many image tasks [59], it motivates us to accurately estimate the coefficients by fully exploring the above properties.

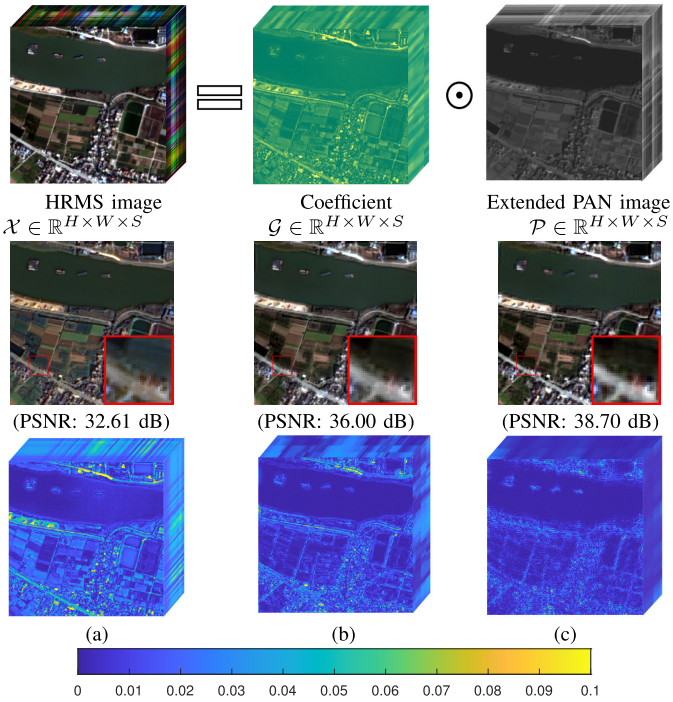


Fig. 1. (First row) Relationship between PAN and HRMS images. The HRMS image can be represented as the Hadamard product of the coefficients and the extended PAN image. (Second row) Visual results (red, green, blue bands) of BT [2], GLP-HPM [3], and the proposed method, respectively. (Third row) Residual images of the coefficients obtained by the above approaches. (a) BT [2]. (b) GLP-HPM [3]. (c) Proposed.

Bearing these concerns in mind, in this article, we propose a new variational pansharpening model based on a novel spatial fidelity with nonlocal regression (SFNLR). The SFNLR combines the local smoothness and nonlocal self-similarity of the coefficients to effectively describe the relationship between the PAN and HRMS images. Fig. 1 shows the powerful ability of the proposed fidelity to preserve spatial structure and texture information. We can see that the SFNLR is able to retain more spatial details than the pixel-based methods. Besides, the proposed model also consists of a spectral fidelity term, which can steadily obtain spectral information from the LRMS image. Afterward, an algorithm that relies on the alternating direction method of multiplier (ADMM) [60] framework is used to solve the proposed model. A series of experimental results show the performance of the proposed model compared with several state-of-the-art methods.

The *contributions* of this article are summarized as follows.

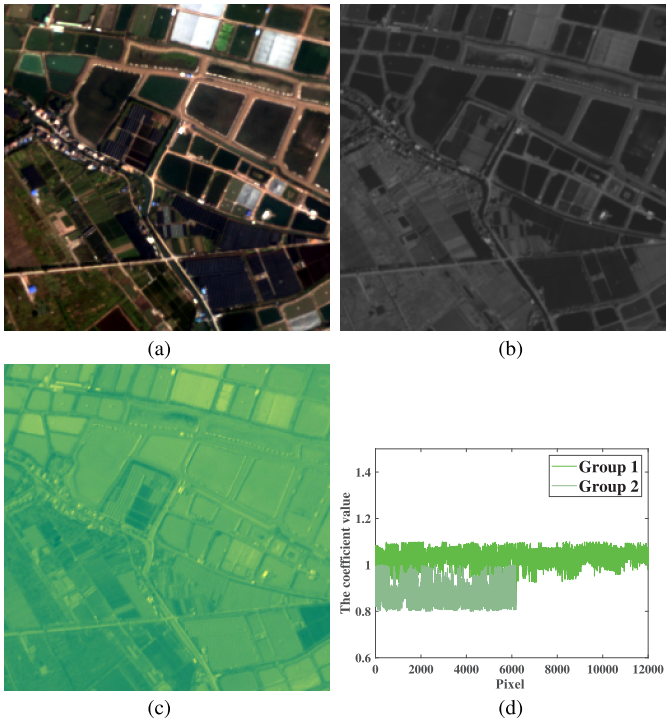


Fig. 2. We randomly chose a reduced resolution data, which includes the PAN, LRMS, and HRMS images (source: GaoFen-2). (a)–(c) HRMS image, the PAN image, and the coefficients, respectively. The coefficients have been obtained by the elementwise division between the HRMS and extended PAN images. It is clear that the coefficients are related to the spatial structure of the HRMS and PAN images. Subsequently, we cut the coefficients into overlapping patches. These patches are clustered into different groups according to the spatial similarity. (d) Distribution of the coefficient values into two of the above-mentioned groups. It can be observed that the values of the coefficients in each group are quite approximated.

1) We propose a new variational model, which consists of the SFNLR and a spectral fidelity term, to solve the pansharpening problem. Furthermore, an ADMM-based algorithm is designed to effectively solve the aforesaid model.
2) The SFNLR is proposed to depict the relationship between PAN and HRMS images based on the latent properties of the coefficients. The SFNLR can simultaneously take advantage of the local smoothness and nonlocal self-similarity of the coefficients to preserve spatial details.
3) The experimental results on both the reduced and full resolution datasets prove the superiority of the proposed method compared with recent state-of-the-art pansharpening approaches.

The remaining of this article is organized as follows. Section II presents the proposed model. Section III describes the algorithm to effectively solve the proposed model. Section IV shows the experiments and the related discussion. Finally, Section V draws concluding remarks.

II. Proposed Model

Before the presentation of the proposed model, the used notations are summarized in Nomenclature.

A. Spectral Fidelity Term

The LRMS image has a relevant spectral content paying it with a coarse spatial resolution. Many previous works, e.g., [46], [61], modeled a spectral fidelity term at the scale of the PAN image by upsampling the LRMS image. Specifically, they assumed that the upsampled LRMS image is the blurred version of the HRMS image by a blurring operator. Since the upsampling operation inevitably affects the extraction of the spectral information, Fu et al. [49] considered the LRMS image as the HRMS image after a downsampling operator to avoid the use of the upsampling operation, yet lacking the blur operator. Combined with the above analysis, the LRMS image can be treated as the degraded and decimated version of the HRMS image [62], which can be denoted as follows:

$$\mathbf{X}_{(3)}\mathbf{BS} = \mathbf{Y}_{(3)} + \xi_1 \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{HW\times HW}$ is a blurring matrix, $\mathbf{S} \in \mathbb{R}^{HW\times hw}$ represents a decimation operation, $\mathbf{Y}_{(3)}$ is the mode-3 unfolding of $\mathcal{Y}$, and ξ_1 indicates a Gaussian noise. Thus, the spectral fidelity term can be written as

$$f_{\text{spec}} = \|\tau(\mathcal{X}) - \mathcal{Y}\|_F^2 \tag{2}$$

where τ denotes both the blurring and decimation operations, i.e., $\mathbf{BS}$ in (1), and $\|\cdot\|_F$ is the Frobenius norm. This spectral fidelity preserves the spectral information and also provides a constraint for the spatial low-frequency component of the HRMS image.

B. Proposed SFNLR

The PAN image contains many spatial information, and strong structural similarities between PAN and HRMS images can be found [49]. Hence, many methods were proposed for building the relationship between the PAN and HRMS images to extract the spatial features. A feasible scheme is to model the above relationship by pixel-based coefficients [63], which can be expressed as follows:

$$\mathcal{X} = \mathcal{G} \odot \mathcal{P} + \xi_2 \tag{3}$$

where ξ_2 denotes a zero-mean Gaussian noise, $\mathcal{P}$ represents the extended PAN image, i.e., the PAN image after a histogram matching procedure, and the size of the coefficient $\mathcal{G}$ is the same as that of the HRMS image. The relationship can be represented by the coefficient $\mathcal{G}$. However, the HRMS image, i.e., the fused product $\mathcal{X}$ in (3), is what we want to find. This motivates us to search for the surrogate of the HRMS image. "Invariance among scales" is widely used in several remote sensing tasks. Exploiting this classical hypothesis, (3) can be converted into the following lower spatial resolution form:

$$\mathbf{X}_{(3)}\mathbf{B} = \mathbf{G}_{(3)} \odot \left(\mathbf{P}_{(3)}\mathbf{B}\right) + \xi_3 \tag{4}$$

where $\mathbf{P}_{(3)}$ and $\mathbf{G}_{(3)}$ are the mode-3 unfolding of the extended PAN image $\mathcal{P}$ and the coefficient $\mathcal{G}$, respectively, and ξ_3 is a zero-mean Gaussian noise. In other words, (4) is built by applying the same spatial blur operation to both the sides of (3). According to (1), we can use an upsampling version of

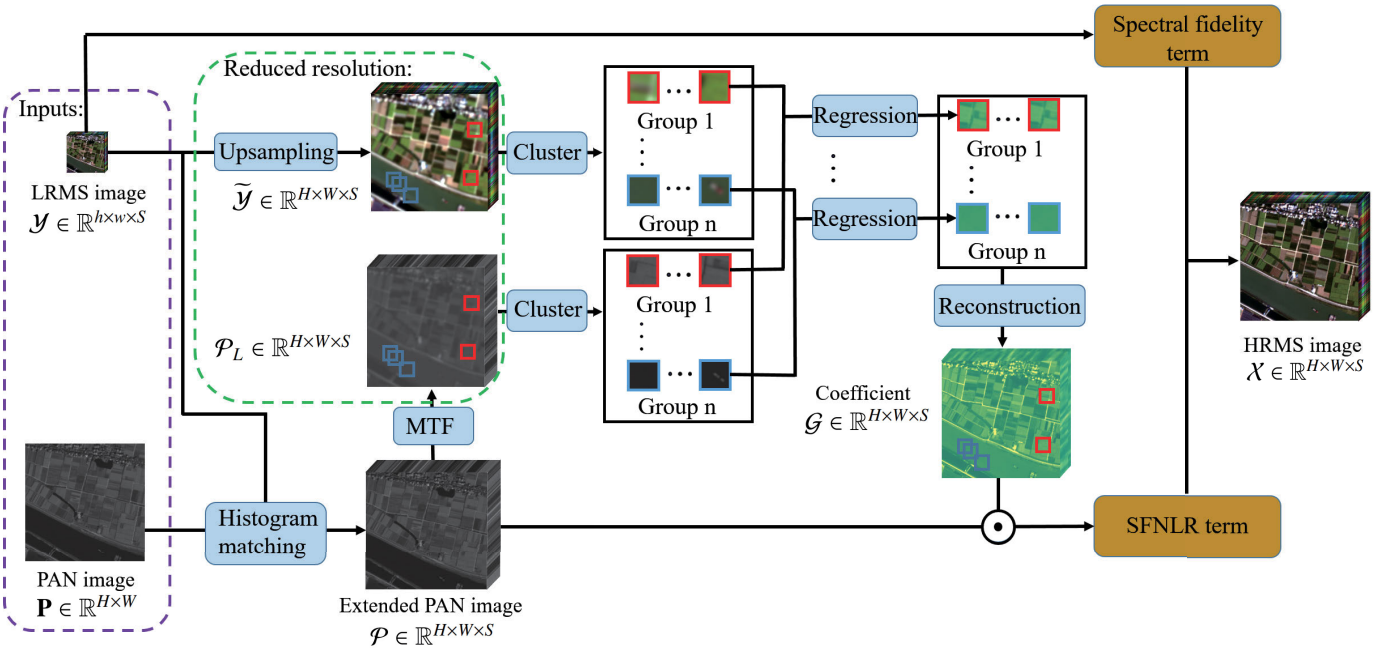


Fig. 3. Flowchart of the proposed model including two fidelity terms, i.e., the spectral fidelity term and the proposed SFNLR term. The SFNLR term is based on the nonlocal regression, which involves the clustering of similar patches and the coefficient estimation on similar patches. More details can be found in Section II.

the LRMS image to surrogate $\mathbf{X}_{(3)}\mathbf{B}$ in (4). Thus, the adaptive coefficient $\mathcal{G}$ can be estimated as follows:

$$\mathcal{G} = \widetilde{\mathcal{Y}} \oslash \mathcal{P}_L \tag{5}$$

where $\widetilde{\mathcal{Y}}$ is the upsampling LRMS image by a polynomial interpolator with 23 coefficients [64], and $\mathcal{P}_L$ is the low-pass version of the extended PAN image $\mathcal{P}$ by a Gaussian filter matched with the MTF of the multispectral (MS) sensor [1]. Obviously, $\mathcal{P}_L$ is the tensor form of $\mathbf{P}_{(3)}\mathbf{B}$ when the MTF is the blur operation in (4). Both $\widetilde{\mathcal{Y}}$ and $\mathcal{P}_L$ are the lower spatial resolution versions of the HRMS image and the extended PAN image, respectively. However, the direct pixel-based computation by (5) is pixel-dependent, which only uses the information of a single pixel, thus ignoring other properties related to the coefficient $\mathcal{G}$.

Based on the scheme (3), we found that the coefficient, i.e., $\mathcal{G}$, inherits the spatial properties, e.g., the local smoothness and nonlocal self-similarity, of the HRMS image. To verify it, we chose a reduced resolution dataset, which includes the PAN and HRMS images, to simulate the coefficient. As shown in Fig. 2, similar spatial structures with the HRMS image exist in the coefficient. The local smoothness and nonlocal self-similarity are two crucial properties of images for retaining spatial information [65]. The local smoothness, i.e., the local structure of images, is relatively stable, thus motivating us to simultaneously estimate the coefficients at a small patch (strongly reducing errors in the estimation procedure). Besides, the nonlocal self-similarity property means that these patches in the coefficient tend to redundantly repeat themselves many times. Hence, the local smoothness and nonlocal self-similarity suggest that the coefficient can be computed using similar nonlocal patches.

Specifically, the extended PAN image is separated first into overlapping patches. We cluster similar patches into n groups. Since the extended PAN image contains the spatial information of the HRMS image, we can learn the nonlocal self-similarity from the extended PAN image. Thus, according to the above clustering results on the extended PAN image, the coefficient $\mathcal{G}$ is divided into the same groups, denoted $\{\mathcal{R}_j\mathcal{G}\}_{j=1}^{n}$, where $\mathcal{R}_j\mathcal{G} \in \mathbb{R}^{p\times p\times n_j}$ (p and n_j are the patch size and patch number, respectively). Similarly, $\widetilde{\mathcal{Y}}$ and $\mathcal{P}_L$ can be divided in the same way, denoted as $\mathcal{R}_j\widetilde{\mathcal{Y}}$ and $\mathcal{R}_j\mathcal{P}_L$, respectively. In addition, as shown in Fig. 2(d), we found that the coefficient values in each group are approximately close. Therefore, we assume the coefficient values in each location of the group satisfy the following overdetermined equations:

$$\mathcal{R}_j\widetilde{\mathcal{Y}} = \mathcal{R}_j\mathcal{G} \odot \mathcal{R}_j\mathcal{P}_L, \quad j = 1, 2, \ldots, n \tag{6}$$

where the values of $\{\mathcal{R}_j\mathcal{G}\}$ are constant. Thus, we can calculate $\mathcal{R}_j\mathcal{G}$ by the regression framework, i.e., exploiting the ordinary least-squares method. The regression approach, which is combined with the patch-based nonlocal strategy, can alleviate the inaccuracy of the estimation to preserve spatial information. The calculation formula of $\mathcal{R}_j\mathcal{G}$ is as follows:

$$\mathcal{R}_j\mathcal{G} = \mathrm{Reg}(\mathcal{R}_j\widetilde{\mathcal{Y}}, \mathcal{R}_j\mathcal{P}_L) \tag{7}$$

where $\mathrm{Reg}(\mathcal{A}, \mathcal{B})$ denotes the robust regression of all the corresponding values in $\mathcal{A}$ and $\mathcal{B}$ [66], [67]. Then, the coefficient $\mathcal{G}$ can be reconstructed by $\mathcal{R}_j\mathcal{G}$, $j = 1, 2, \ldots, n$, based on the clustering information.

C. Final Model

Combining the above-mentioned terms, the final model can be expressed as

$$\min_{\mathcal{X}} \ \|\tau(\mathcal{X}) - \mathcal{Y}\|_F^2 + \lambda_1\|\mathcal{X} - \mathcal{G} \odot \mathcal{P}\|_F^2 \tag{8}$$

where λ_1 is the positive balance parameter. The flowchart of the proposed model can be found in Fig. 3. It is worth to be remarked that the influence of the fidelity terms is enough.

In addition, the calculation of regularization terms, e.g., the total variation (TV) regularization, is usually time-consuming and introduces further parameters to tune. Thus, we only apply fidelity terms to constrain the spectral and spatial information of the HRMS image for conciseness of the model (considering only one parameter in the final model). More discussions about the regularization term can be found in Section IV-D2. Although the final model is convex, the direct calculation will lead to a high computational burden. Therefore, we designed an algorithm based on the ADMM [60] framework to effectively solve the final model. Section III will be about the proposed ADMM-based algorithm.

III. Proposed Algorithm

In this section, we design an effective algorithm based on the ADMM [60] framework to solve the final model (8). The ADMM [60] is an algorithm that solves convex optimization problems by breaking them into smaller pieces, each of which are then easier to handle. ADMM can be viewed as an attempt to blend the benefits of dual decomposition and augmented Lagrangian methods for constrained optimization. We introduce the auxiliary variables $\mathbf{M}$ by $\mathbf{M} = \mathbf{X}_{(3)}\mathbf{B}$. Thus, the final model is converted into the following format:

$$\begin{aligned} &\min_{\mathbf{M},\mathcal{X}} \left\|\mathbf{MS} - \mathbf{Y}_{(3)}\right\|_F^2 + \lambda_1 \|\mathcal{X} - \mathcal{G} \odot \mathcal{P}\|_F^2 \\ &\text{s.t. } \mathbf{M} = \mathbf{X}_{(3)}\mathbf{B}. \end{aligned} \tag{9}$$

From the above format, the augmented Langrangian function is as follows:

$$\begin{aligned} \mathcal{L} = \left\|\mathbf{MS} - \mathbf{Y}_{(3)}\right\|_F^2 + \frac{\eta_1}{2}\left\|\mathbf{X}_{(3)}\mathbf{B} - \mathbf{M} + \frac{\mathbf{J}_{(3)}}{\eta_1}\right\|_F^2 \\ + \lambda_1 \|\mathcal{X} - \mathcal{G} \odot \mathcal{P}\|_F^2 + \text{const} \end{aligned} \tag{10}$$

where $\mathbf{J}_{(3)}$ is the unfolding form of the Lagrange multipliers $\mathcal{J}$, η_1 represents a penalty parameter, and const denotes a constant, which is independent of the variables. Hence, the whole model is transformed into the following three subproblems.

A. M Subproblem

According to (10), the $\mathbf{M}$ subproblem at the $k+1$th iteration step is expressed as the following minimization problem:

$$\min_{\mathbf{M}} \left\|\mathbf{MS} - \mathbf{Y}_{(3)}\right\|_F^2 + \frac{\eta_1}{2}\left\|\mathbf{X}_{(3)}^k\mathbf{B} - \mathbf{M} + \frac{\mathbf{J}_{(3)}^k}{\eta_1}\right\|_F^2 \tag{11}$$

where $\mathbf{X}_{(3)}^k$ denotes $\mathbf{X}_{(3)}$ at the kth iteration step. $\mathbf{MSS}^{\mathrm{T}}$ can be represented in the following elementwise multiplication form:

$$\mathbf{MSS}^{\mathrm{T}} = \mathbf{M} \odot \mathbf{D}_{\mathbf{SS}^{\mathrm{T}}} \tag{12}$$

where $\mathbf{S}^{\mathrm{T}}$ represents the transpose of the matrix, $\mathbf{S}$, and $\mathbf{D}_{\mathbf{SS}^{\mathrm{T}}}$ is a sparse matrix only containing zeros and ones. The graphical representation of (12) is shown in Fig. 4. More details about (12) are presented in [58]. Thus, we can obtain $\mathbf{M}$ by

$$\mathbf{M}^{k+1} = \frac{2\mathbf{Y}_{(3)}\mathbf{S}^{\mathrm{T}} + \eta_1\mathbf{X}_{(3)}^k\mathbf{B} + \mathbf{J}_{(3)}^k}{2\mathbf{D}_{\mathbf{SS}^{\mathrm{T}}} + \eta_1\mathbf{1}} \tag{13}$$

where $\mathbf{1}$ is an all-ones matrix.

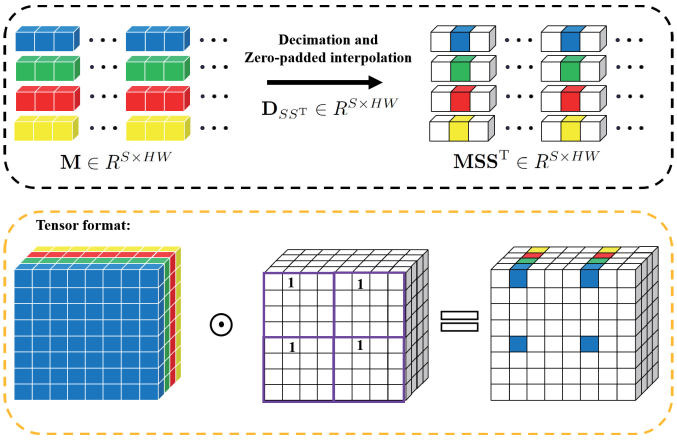


Fig. 4. Graphical representation of (12). The white cube indicates a zero value. The first row shows the processing from $\mathbf{M}$ to $\mathbf{MSS}^{\mathrm{T}}$. The decimation and zero-padded interpolation operations on $\mathbf{M}$ are equal to the elementwise multiplication between $\mathbf{M}$ and $\mathbf{D}_{\mathbf{SS}^{\mathrm{T}}}$. The second row, i.e., the orange box, is the tensor format of the first row when $\mathbf{M} \in \mathbb{R}^{4\times 64}$. It is worth noting that the tensor version of $\mathbf{D}_{\mathbf{SS}^{\mathrm{T}}}$ is produced from sparse matrix (i.e., the purple area), whose entries are 1 only in one position.

B. $\mathbf{X}_{(3)}$ Subproblem

There are matrix and tensor forms of $\mathcal{X}$ in (10). For simplicity, the subproblem in the tensor unfolding form is

$$\begin{aligned} \min_{\mathbf{X}_{(3)}} \lambda_1\left\|\mathbf{X}_{(3)} - \mathbf{G}_{(3)} \odot \mathbf{P}_{(3)}\right\|_F^2 \\ + \frac{\eta_1}{2}\left\|\mathbf{X}_{(3)}\mathbf{B} - \mathbf{M}^{k+1} + \frac{\mathbf{J}_{(3)}^k}{\eta_1}\right\|_F^2. \end{aligned} \tag{14}$$

The fast Fourier transform (FFT) algorithm is an effective method to solve the above subproblem. Thus, we have

$$\mathbf{X}_{(3)}^{k+1} = \mathcal{F}^{-1}\left(\frac{\mathcal{F}\left(\mathbf{H}^{k+1}\right)}{\mathcal{F}\left(2\lambda_1 + \eta_1\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)}\right) \tag{15}$$

with

$$\mathbf{H}^{k+1} = 2\lambda_1\left(\mathbf{G}_{(3)} \odot \mathbf{P}_{(3)}\right) + \eta_1\mathbf{M}^{k+1}\mathbf{B}^{\mathrm{T}} - \mathbf{J}_{(3)}^k\mathbf{B}^{\mathrm{T}} \tag{16}$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ represent the FFT and its inverse transformation, respectively.

C. Updating Multipliers

Finally, the multiplier $\mathbf{J}_{(3)}$ is updated as follows:

$$\mathbf{J}_{(3)}^{k+1} = \mathbf{J}_{(3)}^k + \eta_1\left(\mathbf{X}_{(3)}^{k+1}\mathbf{B} - \mathbf{M}^{k+1}\right). \tag{17}$$

The relative change (RelCha) and the number of iterations are used as termination conditions of the algorithm. RelCha is defined as

$$\text{RelCha} = \left\|\mathbf{X}_{(3)}^{k+1} - \mathbf{X}_{(3)}^k\right\|_F / \left\|\mathbf{X}_{(3)}^k\right\|_F. \tag{18}$$

The whole solving algorithm is summarized in Algorithm 1, where k_{mit} denotes the maximum iteration, r is the scale ratio between the LRMS and PAN images, ε refers to a tolerance value, and Ψ indicates the bicubic interpolation.

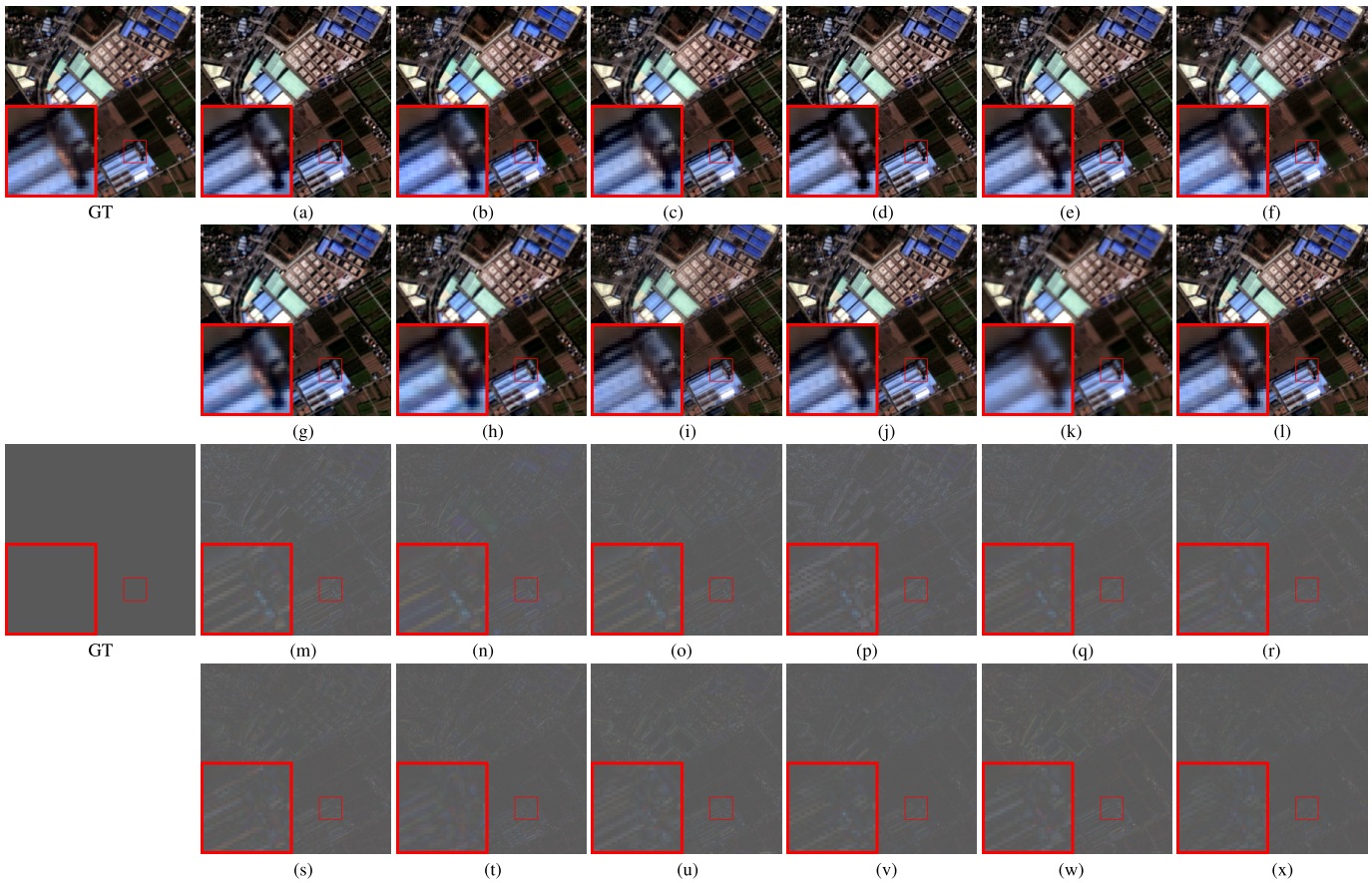


Fig. 5. Fusion results with close-ups for different approaches on a four-band image (source: GaoFen-2 satellite) at reduced resolution (size of the PAN image: 256 × 256). (a)–(l) Visual results in true colors of GSA [14], C-BDSD [68], BDSD-PC [66], GLP-HPM [3], GLP-Reg-FS [69], C-GLP [67], PNN [70], TPNN [71], HQBP [72], CDIF [53], BAGDC [61], and the proposed method. (m)–(x) Corresponding residual images for the displayed red, green, and blue bands. We added 0.35 to aid visual inspection.

Algorithm 1 ADMM-Based Solver for the Proposed Pansharpening Model (8)

Input: The LRMS image $\mathcal{Y}$, the extended PAN image $\mathcal{P}$, the adaptive coefficient $\mathcal{G}$, λ_1, η_1, r, k_{mit}, ε.

Initialization: $\mathbf{X}_{(3)}^0 = \Psi(\mathcal{Y}, r)$, $\mathbf{M}^0 = \mathbf{J}_{(3)}^0 = \mathbf{0}$, $k = 0$.

1: **while** $k < k_{mit}$ and $RelCha > \varepsilon$ **do**
2: Update $\mathbf{M}^{k+1}$ via (13).
3: Update $\mathbf{X}_{(3)}^{k+1}$ via (15).
4: Update Lagrange multiplier $\mathbf{J}_{(3)}^{k+1}$ via (17).
5: $k = k + 1$.
6: **end while**

Output: The fused HRMS image $\mathbf{X}_{(3)}$.

IV. Experimental Results

In this section, we compare the proposed method with some state-of-the-art approaches to verify the superiority of the proposed model. The experiments consist of qualitative and quantitative assessments on datasets from different sensors, i.e., GaoFen-2, QB, and the WorldView series. For fairness, all the methods in the benchmark are run in MATLAB (R2020a) on a computer of 32-Gb RAM, Intel[1] Core[2] i7-8700K CPU at 3.70 GHz and NVIDIA GeForce GTX 1650. Besides, we fine-tune the parameters of all the approaches to achieve better results. Afterward, more analyses about parameters, the influence of the fidelity and regularization terms, the running time of the different components of the final model, and the nonlocal strategy are conducted. It is worth noting that we set k_{mit}, r, and ε to 100, 4, and 2×10^{-5}, respectively, in our experiments.

To evaluate the performance of all the methods, we apply a variety of objective metrics for comprehensive assessment. For reduced resolution data, the peak signal-to-noise ratio (PSNR), the structural similarity index (SSIM) [73], the spectral angle mapper (SAM) [74], the spatial correlation coefficient (SCC) [75], the erreur relative globale adimensionnelle de synthèse (ERGAS) [76], and the $Q4$ index [77] (or the $Q8$ [78] metric for eight-band datasets) are used. For full resolution data, we chose the hybrid quality with no reference (HQNR) [15], which consists of a spectral quality index, D_λ, and a spatial quality index, D_s.

A. Benchmark

To validate the effectiveness of the proposed model, we compare it with the following representative methods belonging to different pansharpening classes.

1) *CS Methods:*

[1]Registered trademark.
[2]Trademarked.

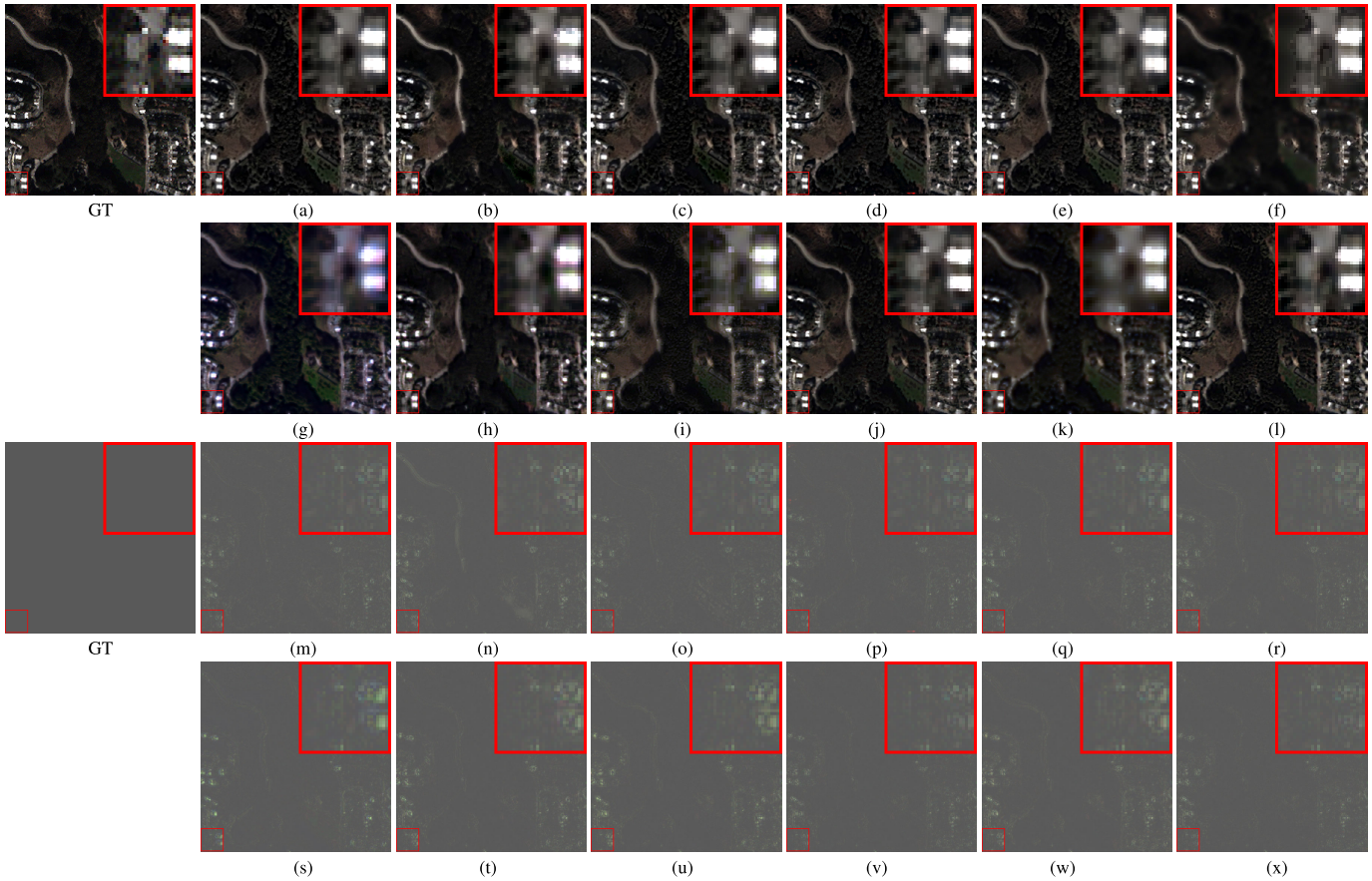


Fig. 6. Fusion results with close-ups for different approaches on a four-band image (source: QB satellite) at reduced resolution (size of the PAN image: 256 × 256). (a)–(l) Visual results in true colors of GSA [14], C-BDSD [68], BDSD-PC [66], GLP-HPM [3], GLP-Reg-FS [69], C-GLP [67], PNN [70], TPNN [71], HQBP [72], CDIF [53], BAGDC [61], and the proposed method. (m)–(x) Corresponding residual images for the displayed red, green, and blue bands. We added 0.35 to aid visual inspection.

a) *GSA:* Gram–Schmidt adaptive method [14].
b) *C-BDSD:* Context-adaptive band-dependent spatial-detail method [68].
c) *BDSD-PC:* Robust band-dependent spatial-detail method [66].

2) *MRA Methods:*
a) *GLP-HPM:* GLP with MTF-matched filter and high-pass modulation injection method [3].
b) *GLP-Reg-FS:* GLP with a full-scale regression method [69].
c) *C-GLP:* The context-based GLP [67].

3) *DL Methods:*
a) *PNN:* Pansharpening neural network [70].
b) *TPNN:* Target-adaptive pansharpening neural network [71].

4) *VO Methods:*
a) *HQBP:* High-quality Bayesian pansharpening method [72].
b) *CDIF:* Context-aware details injection fidelity [53].
c) *BAGDC:* Model based on band-adaptive gradient and detail correction [61].

B. Reduced Resolution Validation

We compare the proposed method with other advanced approaches on reduced resolution data from different sensors, i.e., GaoFen-2, QB, and WorldView-3.

1) GaoFen-2 Dataset: The GaoFen-2 sensor captures the PAN and LRMS images at a spatial resolution of 0.8 and 3.2 m, respectively. We compare the proposed method with the methods in the benchmark on 82 test samples (size of the PAN image: 256 × 256). As shown in Table I, the quantitative comparison shows that the proposed method gets the best results. For visual comparison, we randomly chose an image from the above pairs, as shown in Fig. 5. We can note that Fig. 5(l) is visually close to the ground truth (GT). Fig. 5(k) shows spatial distortion. Instead, in Fig. 5(a)–(c), spectral distortion is clear. Compared with the other methods, the proposed approach better preserves spatial and spectral information. Furthermore, residual images are used to clearly show the differences with respect to the GT. The proposed approach produces excellent outcomes, especially on the edges of the houses in the scene.

2) QB Dataset: The QB satellite captures PAN and LRMS images at 0.7- and 2.8-m resolution, respectively. We chose 48 test examples from the QB satellite (size of the PAN image: 256 × 256). The quantitative comparison among the different approaches can be found in Table I. The proposed method shows the best results compared with the other techniques. About the visual appearance, the proposed method also obtains the best performance, as shown in Fig. 6. More specifically, Fig. 6(i) obtained from HQBP [72] ignores some spatial details

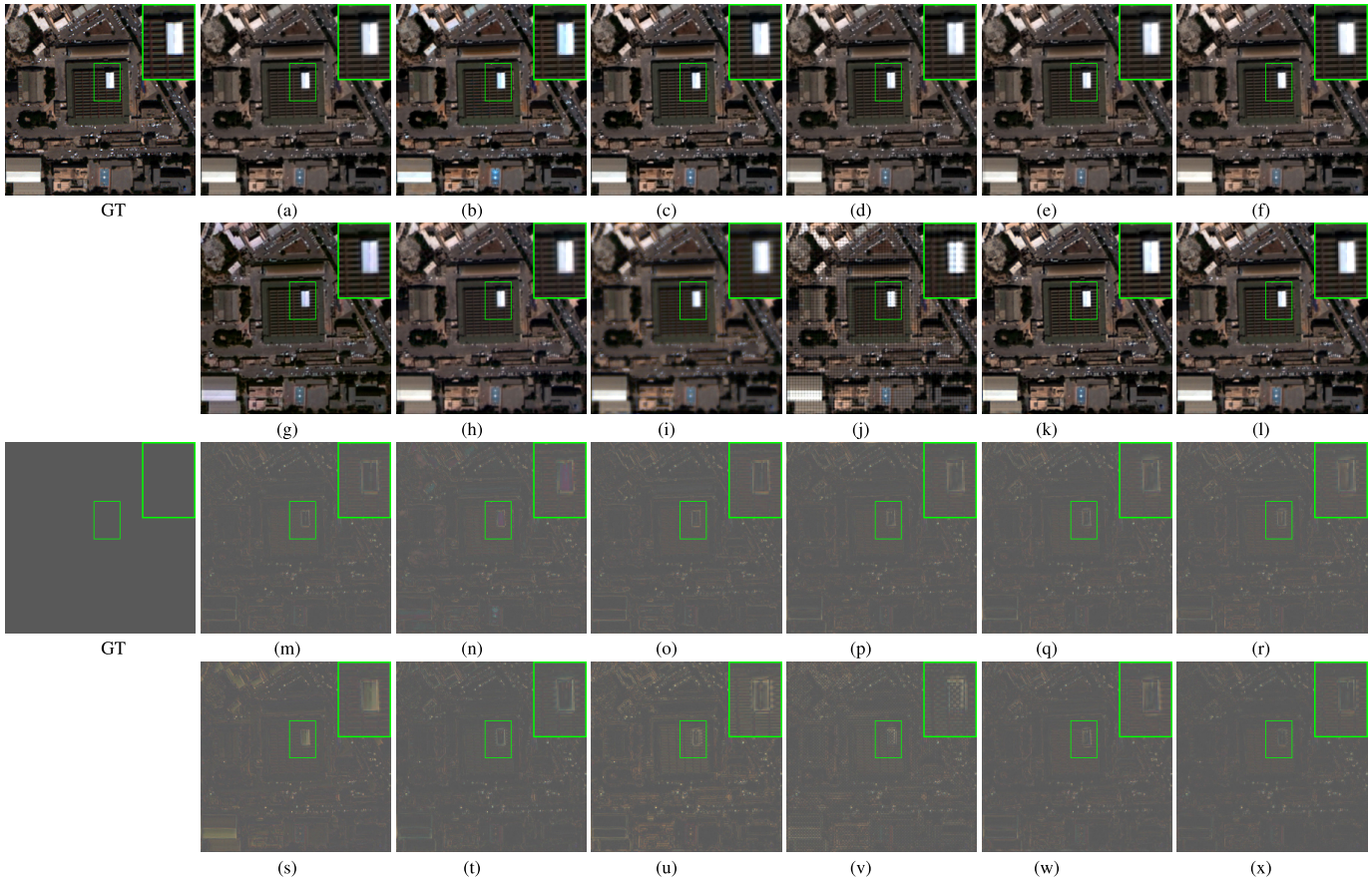


Fig. 7. Fusion results with close-ups for different approaches on an eight-band image from the Tripoli dataset (source: WorldView-3 satellite) at reduced resolution (size of the PAN image: 256 × 256). (a)–(l) Visual results in true colors of GSA [14], C-BDSD [68], BDSD-PC [66], GLP-HPM [3], GLP-Reg-FS [69], C-GLP [67], PNN [70], TPNN [71], HQBP [72], CDIF [53], BAGDC [61], and the proposed method. (m)–(x) Corresponding residual images for the displayed red, green, and blue bands. We added 0.35 to aid visual inspection.

causing spatial distortion. Fig. 6(g), obtained from PNN [70], is spectrally distorted. Instead, the proposed method can keep both the spatial and spectral information. Finally, the residual images also validate the high performance of the proposed approach.

3) WorldView-3 Tripoli Dataset: The WorldView-3 sensor captures PAN and LRMS images at a spatial resolution of 0.31 and 1.24 m, respectively. In Table I, we show the quantitative results on the Tripoli dataset, which contains 42 images (size of the PAN image: 256 × 256). As shown in Table I, the proposed method achieves the best results, thus demonstrating the superiority and robustness of the proposed method. We chose an image from the Tripoli dataset for the visual analysis. As shown in Fig. 7, despite C-BDSD [68] being a nonlocal method, it shows spectral distortion. As displayed in Fig. 7(d)–(f), these results produced by MRA approaches cause more spatial distortion. The results from the proposed method have a better visual appearance than the others. The VO method can well preserve spectral and spatial information. From Fig. 7(m)–(x), residual images indicate that the sharpened product from the proposed method is closer to the GT.

C. Full Resolution Assessment

This section focuses on the full resolution assessment. The Stockholm dataset (size of the PAN image: 400 × 400, source: WorldView-2) consists of eight image pairs. Instead, 16 image pairs are from the Alice dataset (size of the PAN image: 256 × 256, source: WorldView-4). The WorldView-2 and WorldView-4 satellites obtain PAN and LRMS images at a spatial resolution of (0.46, 1.84 m) and (0.31, 1.24 m), respectively. As shown in Table II, it is clear that the proposed method achieves the best accuracy for the overall quality metric, HQNR. For the other evaluation indexes, the proposed approach also gets good and balanced results. We visually compare the results of the different methods in Fig. 8. The images in Fig. 8(b) and (c) are affected by spatial distortion, especially for vegetated areas. Fig. 8(g) instead shows spectral distortion. On the other hand, our model gets good spatio-spectral results. About the Alice dataset, Fig. 8(p) is affected by severe spatial and spectral distortions for some pixels. Fig. 8(n), (u), and (v) is spatially distorted, as shown in the close-ups (red boxes), i.e., in the bottom-left corner of the images. Our method instead preserves spatial details, as shown in Fig. 8(x).

D. Discussion

1) Analysis of Parameters: In this section, we analyze the sensitivity of the parameters on a QB test case (size of the PAN image: 256 × 256). There are only two parameters, i.e., λ_1 and η_1, in the proposed algorithm. We show the curve of a parameter by fixing the other parameter. For a better

TABLE I

QUANTITATIVE RESULTS FOR 82 IMAGES FROM THE GAOFEN-2 (GF-2) SATELLITE, 48 IMAGES FROM QB SATELLITE, AND 42 IMAGES FROM THE TRIPOLI DATASET (SOURCE: WORLDVIEW-3, WV-3), RESPECTIVELY. "VALUE1" DENOTES MEAN VALUES, AND "VALUE2" INDICATES STANDARD DEVIATION VALUES OF THE METRICS IN THE FORM "VALUE1 ± VALUE2." ALL THE METHODS ARE EXECUTED ON CPU EXCEPT FOR TPNN [71] RUNNING ON GPU. (BOLD: BEST; UNDERLINE: SECOND BEST)

Dataset	Method	PSNR	SSIM	SAM	SCC	ERGAS	Q4/Q8	Runtime(s)
GF-2	GSA [14]	33.67 ± 2.04	0.884 ± 0.03	1.85 ± 0.31	0.947 ± 0.02	1.995 ± 0.33	0.878 ± 0.04	0.023 ± 0.002
	C-BDSD [68]	33.21 ± 1.96	0.900 ± 0.03	1.98 ± 0.34	0.948 ± 0.02	2.083 ± 0.37	0.892 ± 0.03	0.219 ± 0.017
	BDSD-PC [66]	33.97 ± 2.09	0.893 ± 0.03	1.85 ± 0.33	0.952 ± 0.02	1.931 ± 0.33	0.892 ± 0.03	**0.019 ± 0.001**
	GLP-HPM [3]	33.78 ± 2.09	0.880 ± 0.04	1.78 ± 0.33	0.948 ± 0.02	1.945 ± 0.31	0.881 ± 0.04	0.044 ± 0.002
	GLP-Reg-FS [69]	34.30 ± 1.78	0.892 ± 0.03	1.80 ± 0.32	0.953 ± 0.02	1.838 ± 0.26	0.890 ± 0.04	0.036 ± 0.002
	C-GLP [67]	34.24 ± 1.85	0.895 ± 0.03	1.82 ± 0.32	0.952 ± 0.02	1.913 ± 0.26	0.895 ± 0.03	0.294 ± 0.080
	PNN [70]	32.74 ± 1.60	0.865 ± 0.03	2.17 ± 0.37	0.939 ± 0.02	2.296 ± 0.34	0.868 ± 0.04	0.280 ± 0.025
	TPNN [71]	35.08 ± 1.74	0.919 ± 0.02	1.93 ± 0.37	0.960 ± 0.02	1.755 ± 0.28	0.919 ± 0.03	2.558 ± 0.047
	HQBP [72]	34.59 ± 1.88	0.896 ± 0.03	1.60 ± 0.30	0.956 ± 0.02	1.778 ± 0.28	0.896 ± 0.03	10.70 ± 0.435
	CDIF [53]	36.08 ± 2.13	0.923 ± 0.02	1.41 ± 0.28	0.968 ± 0.01	1.521 ± 0.29	0.930 ± 0.03	12.36 ± 0.464
	BAGDC [61]	34.26 ± 1.69	0.891 ± 0.024	2.02 ± 0.357	0.953 ± 0.018	1.896 ± 0.232	0.892 ± 0.027	0.286 ± 0.018
	Proposed	**36.40 ± 2.10**	**0.924 ± 0.02**	**1.37 ± 0.26**	**0.970 ± 0.01**	**1.467 ± 0.27**	**0.934 ± 0.02**	15.61 ± 2.057
QB	GSA [14]	32.50 ± 3.34	0.855 ± 0.04	7.75 ± 1.92	0.901 ± 0.02	7.575 ± 1.00	0.797 ± 0.15	0.025 ± 0.009
	C-BDSD [68]	32.08 ± 3.94	0.860 ± 0.04	7.78 ± 1.96	0.887 ± 0.05	8.068 ± 1.87	0.801 ± 0.15	0.239 ± 0.036
	BDSD-PC [66]	32.59 ± 3.39	0.859 ± 0.04	7.75 ± 1.97	0.902 ± 0.02	7.494 ± 1.00	0.805 ± 0.15	**0.021 ± 0.009**
	GLP-HPM [3]	31.98 ± 2.76	0.864 ± 0.03	7.45 ± 1.77	0.882 ± 0.06	12.06 ± 13.6	0.808 ± 0.14	0.045 ± 0.004
	GLP-Reg-FS [69]	32.42 ± 3.43	0.855 ± 0.04	7.53 ± 1.83	0.901 ± 0.02	7.622 ± 1.02	0.796 ± 0.14	0.037 ± 0.004
	C-GLP [67]	32.09 ± 3.66	0.844 ± 0.04	7.50 ± 1.78	0.890 ± 0.02	8.077 ± 1.17	0.786 ± 0.13	0.306 ± 0.072
	PNN [70]	31.05 ± 2.88	0.853 ± 0.03	8.16 ± 1.16	0.879 ± 0.02	8.535 ± 0.78	0.782 ± 0.14	0.297 ± 0.167
	TPNN [71]	32.07 ± 3.57	0.848 ± 0.04	7.72 ± 1.79	0.889 ± 0.02	8.084 ± 1.08	0.790 ± 0.15	2.588 ± 0.184
	HQBP [72]	31.36 ± 4.02	0.808 ± 0.06	8.16 ± 2.02	0.879 ± 0.02	8.980 ± 1.48	0.743 ± 0.12	11.36 ± 0.171
	CDIF [53]	33.22 ± 3.55	0.870 ± 0.04	7.22 ± 1.74	0.916 ± 0.02	7.093 ± 0.95	0.822 ± 0.14	12.02 ± 0.437
	BAGDC [61]	31.59 ± 3.75	0.807 ± 0.05	8.19 ± 1.98	0.880 ± 0.02	8.612 ± 1.25	0.746 ± 0.16	0.259 ± 0.031
	Proposed	**33.30 ± 3.49**	**0.873 ± 0.03**	**7.19 ± 1.72**	**0.917 ± 0.02**	**7.000 ± 0.91**	**0.831 ± 0.13**	16.97 ± 2.143
WV-3	GSA [14]	31.31 ± 1.66	0.882 ± 0.08	4.31 ± 1.77	0.937 ± 0.07	3.412 ± 1.61	0.891 ± 0.12	0.087 ± 0.005
	C-BDSD [68]	29.21 ± 3.07	0.875 ± 0.08	5.24 ± 1.84	0.900 ± 0.09	4.658 ± 2.61	0.879 ± 0.11	0.749 ± 0.035
	BDSD-PC [66]	31.38 ± 1.66	0.889 ± 0.08	4.34 ± 1.74	0.939 ± 0.06	3.376 ± 1.50	0.897 ± 0.12	**0.054 ± 0.004**
	GLP-HPM [3]	30.95 ± 1.72	0.881 ± 0.07	4.37 ± 1.70	0.935 ± 0.06	5.351 ± 1.03	0.893 ± 0.11	0.139 ± 0.005
	GLP-Reg-FS [69]	30.90 ± 1.63	0.876 ± 0.07	4.35 ± 1.72	0.936 ± 0.06	3.541 ± 1.45	0.886 ± 0.11	0.105 ± 0.005
	C-GLP [67]	30.62 ± 1.72	0.869 ± 0.07	4.49 ± 1.69	0.932 ± 0.06	3.668 ± 1.42	0.884 ± 0.11	0.864 ± 0.202
	PNN [70]	27.87 ± 1.25	0.864 ± 0.07	6.14 ± 1.45	0.921 ± 0.06	5.231 ± 0.10	0.870 ± 0.10	0.281 ± 0.014
	TPNN [71]	29.30 ± 1.53	0.848 ± 0.07	5.09 ± 1.68	0.925 ± 0.06	4.197 ± 1.32	0.865 ± 0.12	2.988 ± 0.193
	HQBP [72]	29.03 ± 1.78	0.788 ± 0.05	4.60 ± 1.67	0.916 ± 0.05	4.464 ± 1.33	0.836 ± 0.10	37.23 ± 1.112
	CDIF [53]	31.90 ± 1.58	0.890 ± 0.07	**4.05 ± 1.73**	0.946 ± 0.06	3.168 ± 1.41	0.904 ± 0.11	30.96 ± 0.815
	BAGDC [61]	31.12 ± 1.80	0.885 ± 0.08	4.39 ± 1.76	0.936 ± 0.07	3.462 ± 1.54	0.898 ± 0.11	0.768 ± 0.050
	Proposed	**32.02 ± 1.69**	**0.895 ± 0.07**	4.12 ± 1.79	**0.947 ± 0.06**	**3.130 ± 1.42**	**0.911 ± 0.10**	27.45 ± 2.490
	Ideal value	**+∞**	**1**	**0**	**1**	**0**	**1**	**0**

visualization, we consider the quality metrics SAM, ERGAS, and $Q4$, standardized as [metric$-$**Mean**(metric)]/**Std**(metric), where **Mean**$(\cdot)$ and **Std**$(\cdot)$ denote the average and the standard deviation operators, respectively. As shown in Fig. 9, to balance the different terms of the proposed model, we set $\lambda_1 = 0.00001$ and $\eta_1 = 0.0001$ in the experimental analysis.

2) Influence of Fidelity and Regularization Terms: The conventional variational model usually consists of fidelity and regularization terms. In this section, we analyze the influence of fidelity and regularization terms. As an instance, we chose the TV regularization, which is widely used in many pansharpening models, e.g., [50], [53]. The TV regularization at pixel (i, j, k) is defined as follows:

$$\begin{aligned}\|\nabla_1\mathcal{X}(i,j,k)\|_1 &= \|\mathcal{X}(i+1,j,k) - \mathcal{X}(i,j,k)\|_1\\ \|\nabla_2\mathcal{X}(i,j,k)\|_1 &= \|\mathcal{X}(i,j+1,k) - \mathcal{X}(i,j,k)\|_1\\ \|\nabla_3\mathcal{X}(i,j,k)\|_1 &= \|\mathcal{X}(i,j,k+1) - \mathcal{X}(i,j,k)\|_1 \end{aligned} \quad (19)$$

where $\|\cdot\|_1$ is the ℓ_1 norm. We analyze the following four submodels on a test case from the QB dataset (size of the PAN image: 256 × 256).

Submodel-I: SFNLR + TV

$$\min_{\mathcal{X}} \ \lambda_1\|\mathcal{X} - \mathcal{G}\odot\mathcal{P}\|_F^2 + \beta_1\|\nabla_1\mathcal{X}\|_1 + \beta_2\|\nabla_2\mathcal{X}\|_1 + \beta_3\|\nabla_3\mathcal{X}\|_1 \quad (20)$$

where β_1, β_2, and β_3 are weight coefficients.

Submodel-II: Spectral fidelity + TV

$$\min_{\mathcal{X}} \ \|\tau(\mathcal{X})\|_F^2 + \beta_1\|\nabla_1\mathcal{X}\|_1 + \beta_2\|\nabla_2\mathcal{X}\|_1 + \beta_3\|\nabla_3\mathcal{X}\|_1. \quad (21)$$

Submodel-III: Spectral fidelity + SFNLR

$$\min_{\mathcal{X}} \ \|\tau(\mathcal{X}) - \mathcal{Y}\|_F^2 + \lambda_1\|\mathcal{X} - \mathcal{G}\odot\mathcal{P}\|_F^2. \quad (22)$$

Submodel-IV: Spectral fidelity + SFNLR + TV

$$\min_{\mathcal{X}} \ \|\tau(\mathcal{X}) - \mathcal{Y}\|_F^2 + \lambda_1\|\mathcal{X} - \mathcal{G}\odot\mathcal{P}\|_F^2 + \beta_1\|\nabla_1\mathcal{X}\|_1 + \beta_2\|\nabla_2\mathcal{X}\|_1 + \beta_3\|\nabla_3\mathcal{X}\|_1. \quad (23)$$

As shown in Table III, the proposed SFNLR shows high performance. The first three submodels perform well, and

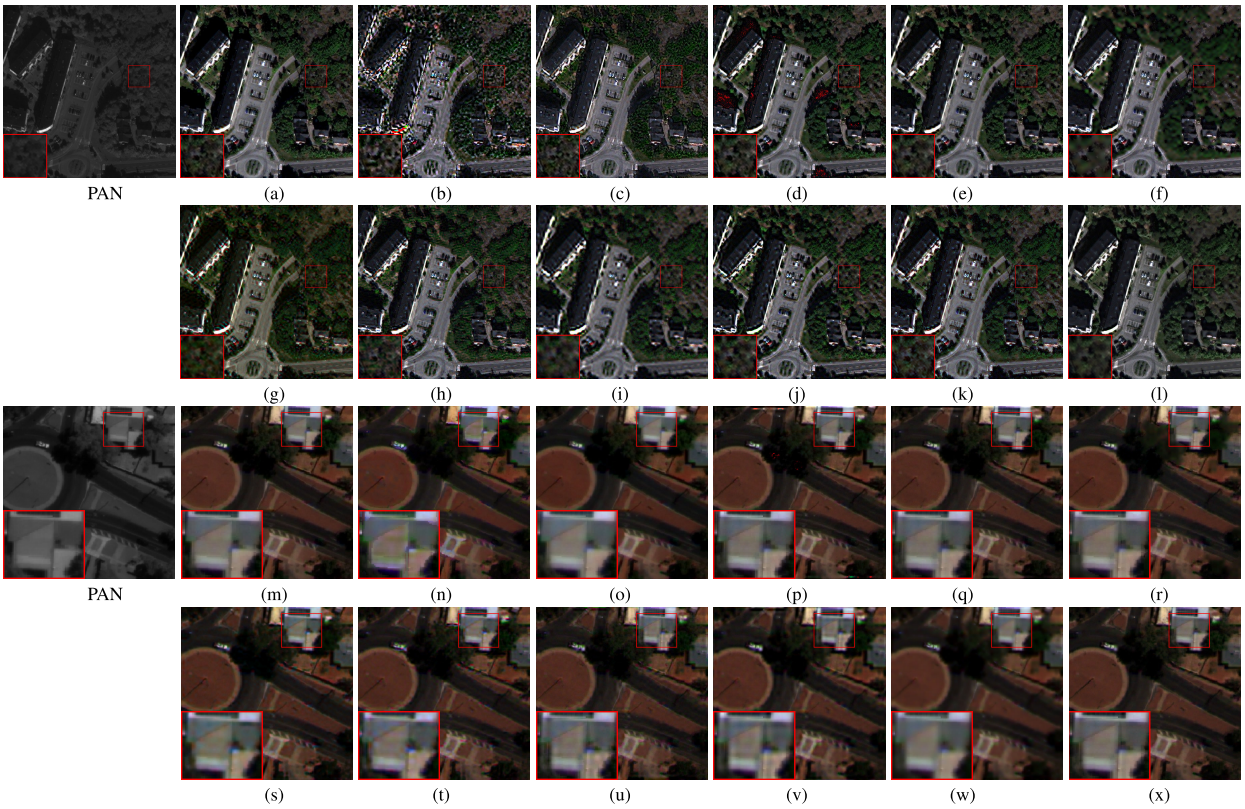


Fig. 8. Fusion results with close-ups for different approaches on the Stockholm dataset (size of the PAN image: 400 × 400, source: WorldView-2 satellite) at full resolution. (a)–(l) Visual results in true colors of GSA [14], C-BDSD [68], BDSD-PC [66], GLP-HPM [3], GLP-Reg-FS [69], C-GLP [67], PNN [70], TPNN [71], HQBP [72], CDIF [53], BAGDC [61], and the proposed method. (m)–(x) Fusion results with close-ups for the above-mentioned approaches on the Alice dataset (size of the PAN image: 256 × 256, source: WorldView-4 satellite).

TABLE II

QUANTITATIVE RESULTS FOR ALL THE COMPARED METHODS IN THE BENCHMARK. (a) EIGHT IMAGES FROM THE FULL RESOLUTION STOCKHOLM DATASET (SIZE OF THE PAN IMAGE: 400 × 400, SOURCE: WORLDVIEW-2, WV-2); (b) 64 IMAGES FROM THE FULL RESOLUTION ALICE DATASET (SIZE OF THE PAN IMAGE: 256 × 256, SOURCE: WORLDVIEW-4, WV-4). "VALUE1" DENOTES MEAN VALUES, AND "VALUE2" DENOTES STANDARD DEVIATION VALUES OF THE METRICS IN THE FORM "VALUE1 ± VALUE2." ALL THE METHODS ARE EXECUTED ON CPU EXCEPT FOR TPNN [71] RUNNING ON GPU. (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	*(a) Full WV-2 Stockholm Dataset*				*(b) Full WV-4 Alice Dataset*			
	$\mathbf{D}_\lambda$	$\mathbf{D}_s$	**HQNR**	**Runtime(s)**	$\mathbf{D}_\lambda$	$\mathbf{D}_s$	**HQNR**	**Runtime(s)**
GSA [14]	0.192 ± 0.045	0.088 ± 0.021	0.737 ± 0.045	0.113 ± 0.006	0.202 ± 0.037	0.164 ± 0.020	0.667 ± 0.033	0.027 ± 0.006
C-BDSD [68]	0.335 ± 0.066	0.162 ± 0.033	0.558 ± 0.065	1.280 ± 0.040	0.244 ± 0.039	0.057 ± 0.015	0.713 ± 0.037	0.264 ± 0.038
RBDSD [66]	0.247 ± 0.051	0.039 ± 0.025	0.724 ± 0.052	**0.075 ± 0.005**	0.207 ± 0.021	0.136 ± 0.034	0.685 ± 0.037	**0.021 ± 0.007**
GLP-HPM [3]	0.165 ± 0.037	0.066 ± 0.024	0.780 ± 0.038	0.195 ± 0.005	0.149 ± 0.026	0.109 ± 0.012	0.759 ± 0.025	0.048 ± 0.002
GLP-Reg-FS [69]	0.182 ± 0.045	0.080 ± 0.023	0.752 ± 0.044	0.148 ± 0.006	0.149 ± 0.029	0.111 ± 0.010	0.757 ± 0.028	0.038 ± 0.002
C-GLP [67]	0.181 ± 0.041	0.050 ± 0.034	0.778 ± 0.040	1.112 ± 0.243	0.161 ± 0.034	0.051 ± 0.038	0.796 ± 0.027	0.276 ± 0.064
PNN [70]	0.220 ± 0.047	0.053 ± 0.010	0.739 ± 0.042	0.335 ± 0.027	0.152 ± 0.026	0.055 ± 0.011	0.801 ± 0.026	0.299 ± 0.108
TPNN [71]	0.167 ± 0.035	**0.035 ± 0.017**	0.804 ± 0.035	2.759 ± 0.037	**0.132 ± 0.026**	0.047 ± 0.011	0.827 ± 0.026	2.578 ± 0.108
HQBP [72]	0.171 ± 0.034	0.041 ± 0.008	0.795 ± 0.033	75.50 ± 0.457	0.144 ± 0.030	0.036 ± 0.008	0.825 ± 0.029	11.77 ± 0.095
CDIF [53]	**0.085 ± 0.015**	0.110 ± 0.019	0.814 ± 0.022	77.25 ± 2.100	0.136 ± 0.030	0.064 ± 0.022	0.809 ± 0.030	11.38 ± 0.340
BAGDC [61]	0.104 ± 0.015	0.070 ± 0.019	0.834 ± 0.024	1.865 ± 0.097	0.144 ± 0.027	0.072 ± 0.019	0.795 ± 0.032	0.351 ± 0.022
Proposed	0.100 ± 0.022	0.059 ± 0.013	**0.847 ± 0.028**	71.31 ± 8.181	0.133 ± 0.029	**0.036 ± 0.007**	**0.847 ± 0.028**	14.09 ± 1.633
Ideal value	**0**	**0**	**1**	**0**	**0**	**0**	**1**	**0**

Submodel-IV achieves the best results at the cost of an increment in the computational burden of the algorithm and the tuning of more parameters. It is clear that the results of the proposed model, i.e., Submodel-III, are as good as the ones of Submodel-IV, but requiring less computational resources. Hence, we can apply the model without the regularization term to reduce the number of parameters to tune and the running times.

3) Analysis About the Nonlocal Strategy: In this article, we adaptively estimate the coefficients with a nonlocal regression, which can improve the accuracy of the estimation of the coefficients by jointly exploring the local smoothness and nonlocal self-similarity of spatial structures. In Table IV, we compare three strategies (pixel-based, local, and patch-based nonlocal) on a test case from the GaoFen-2 dataset (size of the PAN image: 256 × 256). It is clear to

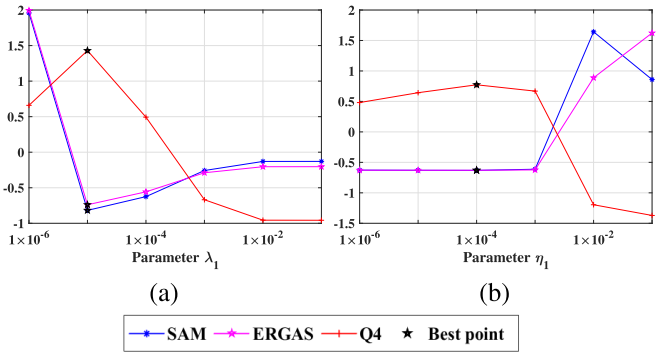


Fig. 9. Robustness analysis (using SAM, ERGAS, $Q4$ metrics) for the parameters (a) λ_1 and (b) η_1 on a QB test case.

TABLE III

QUANTITATIVE RESULTS ABOUT THE INFLUENCE OF FIDELITY AND REGULARIZATION TERMS USING REDUCED RESOLUTION DATA (SOURCE: QB). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q4	Runtime(s)
Submodel-I	31.73	0.840	11.80	0.854	8.567	0.704	**9.97**
Submodel-II	32.15	0.850	11.34	0.862	8.269	0.722	24.0
Submodel-III	34.76	**0.896**	7.217	**0.925**	6.105	**0.859**	20.1
Submodel-IV	**34.78**	**0.896**	**7.197**	**0.925**	**6.098**	**0.859**	25.1
Ideal value	$+\infty$	**1**	**0**	**1**	**0**	**1**	**0**

TABLE IV

QUANTITATIVE RESULTS ABOUT THREE WAYS TO ESTIMATE THE ADAPTIVE COEFFICIENTS USING GAOFEN-2 DATA. (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q4	Runtime(s)
Pixel	35.89	0.926	1.314	0.981	1.372	0.951	**3.98**
Local	36.32	0.927	1.204	0.983	1.296	0.955	4.11
Nonlocal	**36.64**	**0.933**	**1.136**	**0.984**	**1.242**	**0.959**	9.84
Ideal value	$+\infty$	**1**	**0**	**1**	**0**	**1**	**0**

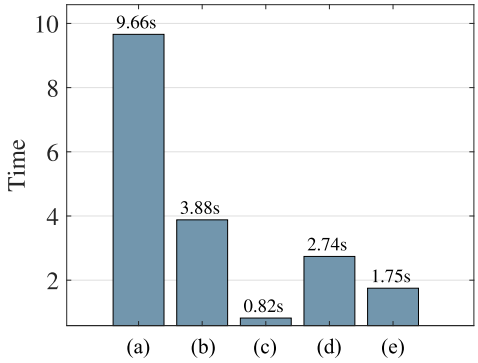


Fig. 10. Computational analysis for each component of the proposed model. (a)–(e) Running times of clustering, coefficient computation, $\mathbf{M}$ subproblem, $\mathbf{X}_{(3)}$ subproblem, and updating multiplier, respectively.

see that the patch-based nonlocal strategy shows the best performance.

4) Running Times for Each Component: The final model can be divided into five main parts, i.e., clustering, coefficient computation, $\mathbf{M}$ subproblem, $\mathbf{X}_{(3)}$ subproblem, and updating multiplier. Fig. 10 shows the time comparison for the above-mentioned components on a GaoFen-2 test case (size of the PAN image: 256 × 256). We find that the clustering takes up the most time compared with the other components. Except for clustering, the proposed model can iteratively update each subproblem with a very reduced computational burden.

5) Discussion on Patch Size and the Number of Similar Patches: The patch size and the number of similar patches are defined before the nonlocal estimation [81]. In this section, we discuss the patch size and the number of similar patches for data acquired by different sensors (i.e., QB, GaoFen-2, and WV-3). From Fig. 11, one can see that the PSNR value tends to be stable when the patch size is bigger than 4 for the WorldView-3 test case. However, the PSNR value decreases when the patch size increases for the QB test case, and the PSNR value increases when the patch size increases for the GaoFen-2 test case. On other hand, the PSNR value reaches the peak when the clustering number is 150 for the QB test case. In addition, it is clear that the ERGAS value is stable varying the patch size and the number of clusters. The larger the patch size and the number of clusters, the longer the computational time. To balance the running time and performance among data acquired by different sensors, we set the patch size and the number of clusters to 5 and 150, respectively, in all our experiments.

6) Comparison With Advanced DL Approaches: In this section, we compare the proposed method with three advanced DL approaches, i.e., FusionNet [79], GTP-PNet [35], and LAGConv [80], on the PanCollection dataset,[3] which is an open collection of pansharpening training–test datasets. All the DL approaches are trained with WV-3 data and tested on WV-3 and WV-2 data (size of the PAN image: 256 × 256). The quantitative results are reported in Table V. We can observe that DL approaches show their superiority for WV-3 data. Since these DL approaches are trained on the WV-3 dataset, their ability to extract features and fuse data from the WV-3 dataset is strong. However, this superiority is due to the huge amount of training data from the same satellite. When these approaches are executed on WV-2 data, their performance is limited as shown in Table V. On the contrary, our VO method removes the need for training data and shows a better generalization ability with stable performance on testing data from different satellites. Thus, compared with the state-of-the-art DL approaches, our method is more practical and stable.

7) Analysis About Coefficient Estimation: The proposed coefficient estimation method is based on the "Invariance among scales," whose effectiveness has been demonstrated in several approaches, e.g., [53], [66]. According to the classical hypothesis, the coefficients are estimated using reduced resolution information. As shown in Table VI, the estimated coefficients can make the extended PAN image close to the HRMS image. In this section, we analyze three ways to use this reduced resolution information on the test data from the GaoFen-2 dataset (size of the PAN image: 256 × 256). A direct approach, named "Redirect," is given by the upsampling of the coefficients calculated at LRMS scale. In addition, a VO model can be built as follows:

$$\min_{\mathbf{G}} \|\mathbf{GS}-\mathbf{N}\|_F^2+\alpha_1\|\nabla_1\mathbf{G}\|_1+\alpha_2\|\nabla_2\mathbf{G}\|_1+\alpha_3\|\nabla_3\mathbf{G}\|_1 \tag{24}$$

where $\mathbf{N} \in \mathbb{R}^{S\times hw}$ represents the coefficient estimated at LRMS scale, and $\mathbf{S} \in \mathbb{R}^{HW\times hw}$ models the decimation oper-

[3]https://liangjiandeng.github.io/PanCollection.html

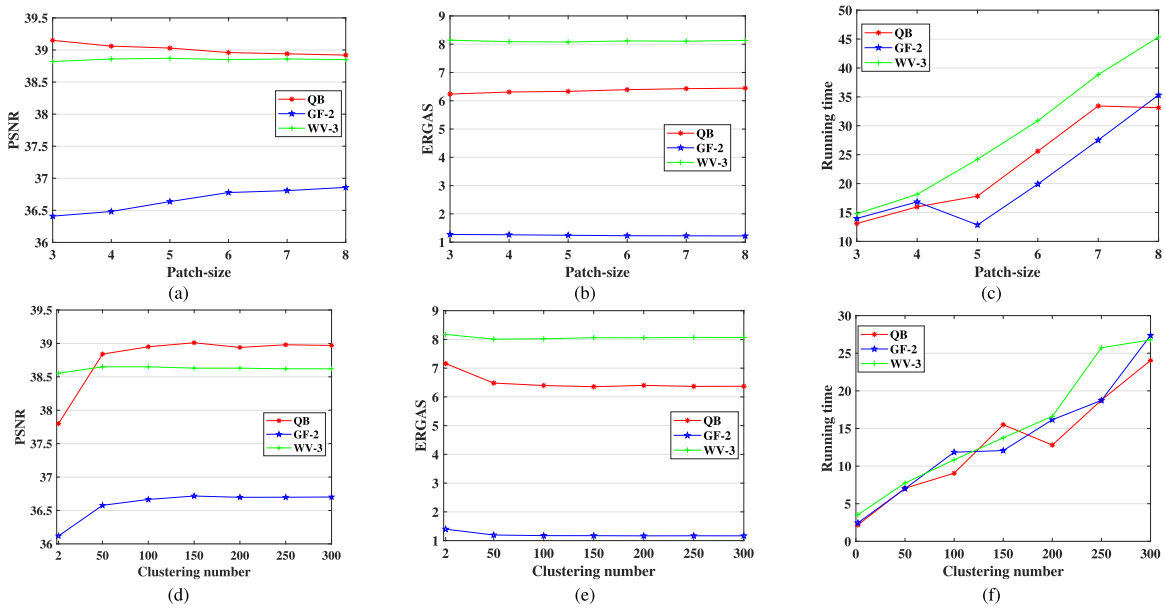


Fig. 11. Changes of the PSNR, the ERGAS, and the running time metrics varying (a)–(c) patch size and (d)–(f) clustering number, where QB, GF-2, and WV-3 indicate QB, GaoFen-2, and WorldView-3, respectively (size of the PAN image: 256 × 256).

TABLE V

QUANTITATIVE RESULTS FOR 20 TEST IMAGES FROM THE WORLDVIEW-3 (WV-3) SATELLITE AND 20 TEST IMAGES FROM WORLDVIEW-2 (WV-2), RESPECTIVELY. THE SIZE OF THE PAN IMAGE IS 256 × 256. ALL THE DL METHODS ARE TRAINED WITH DATA FROM WV-3 SATELLITE. ALL THE METHODS ARE EXECUTED ON GPU EXCEPT FOR THE PROPOSED METHOD RUNNING ON CPU. (BOLD: BEST; UNDERLINE: SECOND BEST)

Dataset	Method	PSNR	SSIM	SAM	SCC	ERGAS	Q8
WV-3	FusionNet [79]	38.09 ± 2.41	0.967 ± 0.01	3.29 ± 0.64	0.978 ± 0.02	2.433 ± 0.60	0.905 ± 0.09
	GTP-PNet [35]	37.44 ± 2.68	0.964 ± 0.01	3.50 ± 0.67	0.976 ± 0.02	2.622 ± 0.61	0.903 ± 0.09
	LAGConv [80]	**38.56 ± 2.60**	**0.971 ± 0.01**	**3.08 ± 0.49**	**0.981 ± 0.02**	**2.283 ± 0.58**	**0.911 ± 0.09**
	Proposed	33.70 ± 2.82	0.901 ± 0.03	4.74 ± 1.48	0.942 ± 0.03	4.585 ± 1.45	0.827 ± 0.11
WV-2	FusionNet [79]	28.54 ± 2.04	0.778 ± 0.04	6.41 ± 0.77	0.907 ± 0.02	5.121 ± 0.52	0.789 ± 0.08
	GTP-PNet [35]	29.37 ± 1.82	0.815 ± 0.03	6.29 ± 0.97	0.933 ± 0.01	4.668 ± 0.43	0.813 ± 0.03
	LAGConv [80]	27.93 ± 1.98	0.812 ± 0.04	6.92 ± 0.40	0.930 ± 0.02	5.292 ± 0.30	0.812 ± 0.09
	Proposed	**30.51± 1.77**	**0.838 ± 0.03**	**5.59 ± 0.75**	**0.939 ± 0.02**	**4.176 ± 0.69**	**0.850 ± 0.09**
	Ideal value	**+∞**	**1**	**0**	**1**	**0**	**1**

TABLE VI

QUANTITATIVE RESULTS ABOUT THREE DIFFERENT WAYS OF APPLYING THE "INVARIANCE AMONG SCALES" TO ESTIMATE THE ADAPTIVE COEFFICIENTS USING GAOFEN-2 DATA. (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q4	Runtime(s)
$\mathcal{P}$	30.06	0.917	7.042	0.923	4.756	0.887	-
Re-Direct	32.20	0.932	5.252	0.954	3.670	0.921	**0.68**
Re-VO	32.06	0.930	5.347	0.951	3.748	0.919	6.01
Proposed	**33.81**	**0.949**	**5.229**	**0.949**	**3.426**	**0.947**	11.4
Ideal value	**+∞**	**1**	**0**	**1**	**0**	**1**	**0**

ator. This VO estimation method, denoted as "Re-VO," uses local smoothing prior to obtain the final coefficient. As demonstrated in Table VI, compared with the other approaches, the proposed estimation method can estimate the coefficient with a higher accuracy.

V. CONCLUSION

In this article, we proposed a novel variational model for the pansharpening problem. More specifically, the SFNLR has been proposed to establish the relationship between the PAN and HRMS images. The SFNLR is based on latent properties, e.g., the local smoothness and the nonlocal self-similarity, of the coefficients to preserve spatial information. We divided first the extended PAN image into overlapping patches. Afterward, we clustered the similar patches into n groups and the coefficient $\mathcal{G}$ is separated into the same groups. Then, we calculated the adaptive coefficients by estimating $\mathcal{G}$ in each group using a nonlocal regression framework. Besides, we designed an ADMM-based algorithm to effectively solve the proposed model. The experiments demonstrated the effectiveness of the proposed method. We achieved state-of-the-art results on both the reduced and full resolution datasets. Moreover, discussions on the proposed approach have been conducted to analyze the parameters, the influence of the fidelity and regularization terms, the nonlocal strategy, the running time for each component of the model, the patch size and the number of similar patches, comparison with advanced DL approaches, and analysis about coefficient estimation methods. Finally, the coefficient estimation is a topic that still deserves future developments. Indeed, more properties, e.g., low-rank

and sparse features, of the coefficients can be considered in the formulated problem.

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] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images. II. Channel ratio and 'chromaticity' transformation techniques," *Remote Sens. Environ.*, vol. 22, no. 3, pp. 343–365, 1987.

[3] G. Vivone, R. Restaino, M. D. Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 930–934, May 2014.

[4] L.-J. Deng et al., "Machine learning in pansharpening: A benchmark, from shallow to deep networks," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 3, pp. 279–315, Sep. 2022, doi: 10.1109/MGRS.2022.3187652.

[5] G. Vivone et al., "A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 1, pp. 53–81, Mar. 2021.

[6] 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.

[7] L. Loncan et al., "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.

[8] X. Meng et al., "A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 1, pp. 18–52, Mar. 2021.

[9] P. S. Chavez and A. Y. Kwarteng, "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.

[10] W. J. Carper, T. M. Lillesand, and P. W. 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, Jan. 1990.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] A. Arienzo, G. Vivone, A. Garzelli, L. Alparone, and J. Chanussot, "Full-resolution quality assessment of pansharpening: Theoretical and hands-on approaches," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 3, pp. 168–201, Sep. 2022.

[16] A. Arienzo, B. Aiazzi, L. Alparone, and A. Garzelli, "Reproducibility of pansharpening methods and quality indexes versus data formats," *Remote Sens.*, vol. 13, no. 21, p. 4399, Oct. 2021.

[17] J. G. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, Jan. 2000.

[18] 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.

[19] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*. Berlin, Germany: Springer, 1995, pp. 281–299.

[20] M. J. Shensa, "The discrete wavelet transform: Wedding the a trous and Mallat algorithms," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992.

[21] G. Vivone, R. Restaino, and J. Chanussot, "A regression-based high-pass modulation pansharpening approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 984–996, Feb. 2018.

[22] C. S. Yilmaz, V. Yilmaz, O. Gungor, and J. Shan, "Metaheuristic pansharpening based on symbiotic organisms search optimization," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 167–187, Dec. 2019.

[23] 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.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[24] D. Hong et al., "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615.

[25] P. Guo, P. Zhuang, and Y. Guo, "Bayesian pan-sharpening with multiorder gradient-based deep network constraints," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 950–962, 2020.

[26] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.

[27] M. Jiang, H. Shen, J. Li, Q. Yuan, and L. Zhang, "A differential information residual convolutional neural network for pansharpening," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 257–271, May 2020.

[28] H. Zhou, J. Hou, Y. Zhang, J. Ma, and H. Ling, "Unified gradient- and intensity-discriminator generative adversarial network for image fusion," *Inf. Fusion*, vol. 88, pp. 184–201, Dec. 2022.

[29] L. Tang, J. Yuan, H. Zhang, X. Jiang, and J. Ma, "PIAFusion: A progressive infrared and visible image fusion network based on illumination aware," *Inf. Fusion*, vols. 83–84, pp. 79–92, Jul. 2022.

[30] J.-F. Hu, T.-Z. Huang, L.-J. Deng, H.-X. Dou, D. Hong, and G. Vivone, "Fusformer: A transformer-based fusion network for hyperspectral image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3194257.

[31] S.-Q. Deng, L.-J. Deng, X. Wu, R. Ran, D. Hong, and G. Vivone, "PSRT: Pyramid shuffle-and-reshuffle transformer for multispectral and hyperspectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5503715, doi: 10.1109/TGRS.2023.3244750.

[32] X. Meng, Q. Liu, F. Shao, and S. Li, "Spatio-temporal–spectral collaborative learning for spatio-temporal fusion with land cover changes," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5704116.

[33] X. Meng, H. Shen, Q. Yuan, H. Li, L. Zhang, and W. Sun, "Pansharpening for cloud-contaminated very high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2840–2854, May 2019.

[34] J.-F. Hu, T.-Z. Huang, L.-J. Deng, T.-X. Jiang, G. Vivone, and J. Chanussot, "Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7251–7265, Dec. 2022, doi: 10.1109/TNNLS.2021.3084682.

[35] H. Zhang and J. Ma, "GTP-PNet: A residual learning network based on gradient transformation prior for pansharpening," *ISPRS J. Photogramm. Remote Sens.*, vol. 172, pp. 223–239, Feb. 2021.

[36] T.-J. Zhang, L.-J. Deng, T.-Z. Huang, J. Chanussot, and G. Vivone, "A triple-double convolutional neural network for panchromatic sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 9, 2022, doi: 10.1109/TNNLS.2022.3155655.

[37] H. Zhang, H. Wang, X. Tian, and J. Ma, "P2Sharpen: A progressive pansharpening network with deep spectral transformation," *Inf. Fusion*, vol. 91, pp. 103–122, Mar. 2023.

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

[39] R. Ran, L.-J. Deng, T.-X. Jiang, J.-F. Hu, J. Chanussot, and G. Vivone, "GuidedNet: A general CNN fusion framework via high-resolution guidance for hyperspectral image super-resolution," *IEEE Trans. Cybern.*, vol. 53, no. 7, pp. 4148–4161, Jul. 2023, doi: 10.1109/TCYB.2023.3238200.

[40] Q. Liu, X. Meng, F. Shao, and S. Li, "Supervised-unsupervised combined deep convolutional neural networks for high-fidelity pansharpening," *Inf. Fusion*, vol. 89, pp. 292–304, Jan. 2023.

[41] M. Gong, J. Ma, H. Xu, X. Tian, and X.-P. Zhang, "D2TNet: A ConvLSTM network with dual-direction transfer for pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5409114.

[42] K. Shen, X. Yang, S. Lolli, and G. Vivone, "A continual learning-guided training framework for pansharpening," *ISPRS J. Photogramm. Remote Sens.*, vol. 196, pp. 45–57, Feb. 2023.

[43] M. Zhou et al., "Spatial-frequency domain information integration for pan-sharpening," in *Computer Vision—ECCV*. Berlin, Germany: Springer, 2022, pp. 274–291.

[44] H. Xu, J. Ma, J. Yuan, Z. Le, and W. Liu, "RFNet: Unsupervised network for mutually reinforcing multi-modal image registration and fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 19647–19656.

[45] Y. Yang, L. Wu, S. Huang, W. Wan, W. Tu, and H. Lu, "Multiband remote sensing image pansharpening based on dual-injection model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1888–1904, 2020.

[46] Y. Yang, C. Wan, S. Huang, H. Lu, and W. Wan, "Pansharpening based on low-rank fuzzy fusion and detail supplement," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5466–5479, 2020.

[47] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, and G. Vivone, "A framelet sparse reconstruction method for pansharpening with guaranteed convergence," *Inverse Problems Imag.*, vol. 17, no. 6, pp. 1277–1300, 2023.

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

[49] X. Fu, Z. Lin, Y. Huang, and X. Ding, "A variational pan-sharpening with local gradient constraints," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10257–10266.

[50] L.-J. Deng, M. Feng, and X.-C. Tai, "The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior," *Inf. Fusion*, vol. 52, pp. 76–89, Dec. 2019.

[51] L.-J. Deng, G. Vivone, W. Guo, M. D. 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.

[52] Z.-C. Wu et al., "A new variational approach based on proximal deep injection and gradient intensity similarity for spatio-spectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6277–6290, 2020.

[53] J.-L. Xiao, T.-Z. Huang, L.-J. Deng, Z.-C. Wu, and G. Vivone, "A new context-aware details injection fidelity with adaptive coefficients estimation for variational pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5408015.

[54] Y. Yang, H. Lu, S. Huang, and W. Tu, "Pansharpening based on joint-guided detail extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 389–401, 2021.

[55] X. Meng et al., "A blind full-resolution quality evaluation method for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5401916.

[56] P. Zhuang, Q. Liu, and X. Ding, "Pan-GGF: A probabilistic method for pan-sharpening with gradient domain guided image filtering," *Signal Process.*, vol. 156, pp. 177–190, Mar. 2019.

[57] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas," in *Proc. 2nd GRSS/ISPRS Joint Workshop Remote Sens. Data Fusion over Urban Areas*, May 2003, pp. 90–94.

[58] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, J.-F. Hu, and G. Vivone, "VO+Net: An adaptive approach using variational optimization and deep learning for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5401016, doi: 10.1109/TGRS.2021.3066425.

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

[60] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[61] H. Lu, Y. Yang, S. Huang, W. Tu, and W. Wan, "A unified pansharpening model based on band-adaptive gradient and detail correction," *IEEE Trans. Image Process.*, vol. 31, pp. 918–933, 2022, doi: 10.1109/TIP.2021.3137020.

[62] T. Xu, T.-Z. Huang, L.-J. Deng, X.-L. Zhao, and J. Huang, "Hyperspectral image superresolution using unidirectional total variation with tucker decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4381–4398, 2020.

[63] G. Vivone and J. Chanussot, "Fusion of short-wave infrared and visible near-infrared WorldView-3 data," *Inf. Fusion*, vol. 61, pp. 71–83, Sep. 2020.

[64] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.

[65] T. Xu, T.-Z. Huang, L.-J. Deng, and N. Yokoya, "An iterative regularization method based on tensor subspace representation for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5529316, doi: 10.1109/TGRS.2022.3176266.

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

[67] G. Vivone, S. Marano, and J. Chanussot, "Pansharpening: Context-based generalized Laplacian pyramids by robust regression," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6152–6167, Sep. 2020.

[68] A. Garzelli, "Pansharpening of multispectral images based on nonlocal parameter optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2096–2107, Apr. 2015.

[69] G. Vivone, R. Restaino, and J. Chanussot, "Full scale regression-based injection coefficients for panchromatic sharpening," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3418–3431, Jul. 2018.

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

[71] 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.

[72] 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.

[73] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[74] R. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, vol. 1, Jun. 1992, pp. 147–149.

[75] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005.

[76] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. Mann 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.

[77] 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.

[78] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 662–665, Oct. 2009.

[79] L.-J. Deng, G. Vivone, C. Jin, and J. Chanussot, "Detail injection-based deep convolutional neural networks for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6995–7010, Aug. 2021.

[80] Z.-R. Jin, T.-J. Zhang, T.-X. Jiang, G. Vivone, and L.-J. Deng, "LAGConv: Local-context adaptive convolution kernels with global harmonic bias for pansharpening," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2022, vol. 36, no. 1, pp. 1113–1121.

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

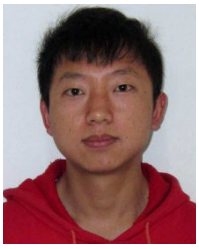
Jin-Liang Xiao received the B.S. degree in mathematics from the School of Mathematics and Statistics, Yunnan University (YNU), Kunming, China, in 2019. He is currently pursuing the Ph.D. degree with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China.

From 2016 to 2018, he was a joint-training B.S. Student with the School of Mathematical Sciences, Fudan University, Shanghai, China. His research interests include image fusion and low-rank image processing.

Ting-Zhu Huang (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computational mathematics from the Department of Mathematics, Xi'an Jiaotong University, Xi'an, China, in 1986, 1992, and 2001, respectively.

He is currently a Professor with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China. His research interests include scientific computation and applications, numerical algorithms for image processing, numerical linear algebra, preconditioning technologies, and matrix analysis with applications.

Dr. Huang is an Editor of *The Scientific World Journal*, *Advances in Numerical Analysis*, the *Journal of Applied Mathematics*, the *Journal of Pure and Applied Mathematics: Advances in Applied Mathematics*, and the *Journal of Electronic Science and Technology*, China.



Liang-Jian Deng (Senior Member, IEEE) received the B.S. and Ph.D. degrees in applied mathematics from the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2010 and 2016, respectively.

From 2013 to 2014, he was a joint-training Ph.D. Student with Case Western Reserve University, Cleveland, OH, USA. In 2017, he was a Post-Doctoral Researcher with Hong Kong Baptist University (HKBU), Hong Kong. In addition, he has stayed at the Isaac Newton Institute for Mathematical Sciences, University of Cambridge, Cambridge, U.K., and HKBU, for short visits. He is currently a Professor with the School of Mathematical Sciences, UESTC. His research interests include use of partial differential equations (PDEs), optimization modeling, and deep learning to address several tasks in image processing, and computer vision, e.g., resolution enhancement and restoration.

Zhong-Cheng Wu received the B.S. degree in information and computing science from the Anhui University of Finance and Economics (AUFE), Bengbu, China, in 2019. He is currently pursuing the Ph.D. degree with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China.

His research interests include image processing, tensor learning, and pansharpening.

Xiao Wu received the B.S. degree in intelligent science and technology from the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications (CQUPT), Chongqing, China, in 2019, and the M.S. degree from the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2023. He is currently pursuing the Ph.D. degree with UESTC under Prof. Ting-Zhu Huang.

His research interests include machine learning and deep learning for computer vision, image processing, and image fusion.

Gemine Vivone (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees (summa cum laude) and the Ph.D. degree in information engineering from the University of Salerno, Fisciano, Italy, in 2008, 2011, and 2014, respectively.

He is currently a Researcher with the Institute of Methodologies for Environmental Analysis (IMAA), National Research Council (CNR), Tito, Italy, and the National Biodiversity Future Center (NBFC), Palermo, Italy. His main research interests include statistical signal processing, detection of remotely sensed images, data fusion, and tracking algorithms.