A New Context-Aware Details Injection Fidelity with Adaptive Coefficients Estimation for Variational Pansharpening

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

***Abstract*—Pansharpening is related to the fusion of a low spatial resolution multispectral (MS) image retaining an abundant spectral content and a high spatial resolution panchromatic (PAN) image to obtain a product with both the abundant spectral content of the former and the high spatial resolution of the latter. Many previous researches are only focused on the global or local relationship between the PAN image and the corresponding high resolution multispectral (HRMS) image. However, we found that the relationship between PAN and HRMS images in the gradient domain can be better explored through the image context. In this paper, we propose a context-aware details injection fidelity (CDIF) with adaptive coefficients estimation, which can fully explore the complicated relationship between the PAN image and the HRMS image in the gradient domain. More specifically, we apply a clustering method to divide the pixels of an image into different context-based regions. Afterwards, the adaptive coefficients are estimated by using a regression-based method for each region. The CDIF is effective in extracting the main features from the two inputs to be fused. In addition, we integrate the CDIF with a conventional fidelity term and a total variation regularization to formulate a novel variational pansharpening model that is solved by designing an algorithm based on the alternating direction method of multiplier (ADMM) framework. Qualitative and quantitative assessments on different datasets support the effectiveness and robustness of the proposed method. The code is available at https://github.com/liangjiandeng/CDIF.**

***Index Terms*—Variational Models, Adaptive Coefficients, Context-Aware Fidelity, Pansharpening, Image Fusion, Remote Sensing.**

I. Introduction

Remote sensing has several applications, such as, detection, analysis, and forecasting. Satellites, e.g., IKONOS, QuickBird, WorldView-2, Pléiades, and WorldView-3, acquire more information. However, because of hardware limitations, we can design acquisition sensors with only one high resolution, usually penalizing the others. Thus, high spatial resolution panchromatic (PAN) images can be acquired together with data showing a greater focus on the spectral resolution as low spatial resolution multispectral (LRMS) images [1]. The goal of pansharpening (which stands for panchromatic sharpening) is to fuse the above-mentioned pairs to obtain high spatial resolution multispectral (HRMS) images.

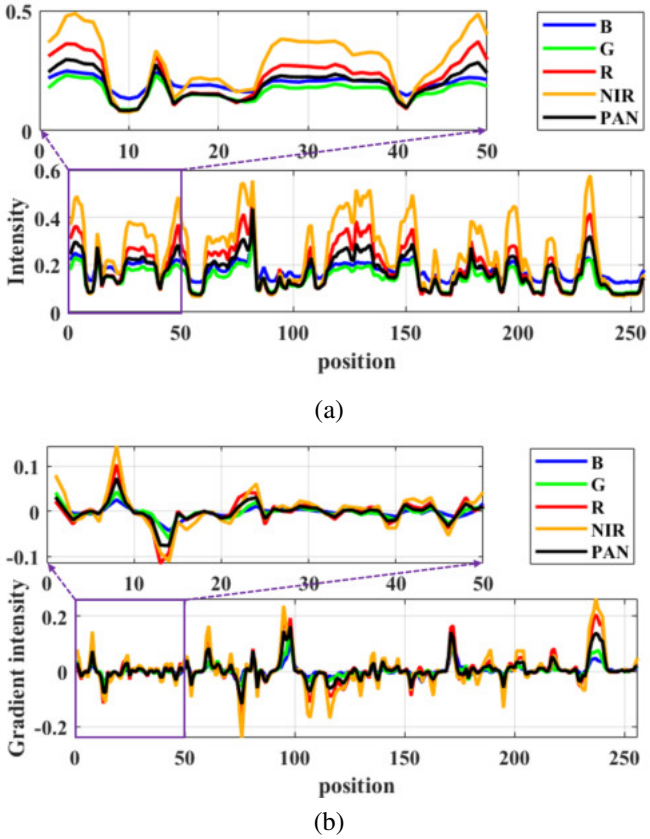


Fig. 1. (a) Plot of the intensity for both the PAN image and the HRMS data randomly choosing a row of the image; (b) Plot of the gradient intensity for both the PAN image and the HRMS data considering blue (B), green (G), red (R), and near-infrared (NIR) bands for the same row as in (a). It is worth to be remarked that the behavior of PAN and HRMS images is more similar in the gradient domain than in the intensity domain.

A. Related Works

Pansharpening methods can be roughly divided into four classes [2], [3], [4], i.e., component substitution (CS) methods, multi-resolution analysis (MRA) approaches, machine learning (ML) techniques, and variational optimization-based (VO)

The work is supported by National Natural Science Foundation of China (Grant No. 12171072, 61702083), and Key Projects of Applied Basic Research in Sichuan Province (Grant No. 2020YJ0216) and National Key Research and Development Program of China (Grant No. 2020YFA0714001).

J.-L. Xiao, T.-Z. Huang, L.-J. Deng and Z.-C. Wu are with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, China (e-mails: jinliang_xiao@163.com; tingzhuhuang@126.com; liangjian.deng@uestc.edu.cn; wuzhch97@163.com).

G. Vivone is with the National Research Council - Institute of Methodologies for Environmental Analysis, CNR-IMAA, I-85050, Tito, Italy (e-mail: gemine.vivone@imaa.cnr.it).

Corresponding Author: T.-Z. Huang, L.-J. Deng.

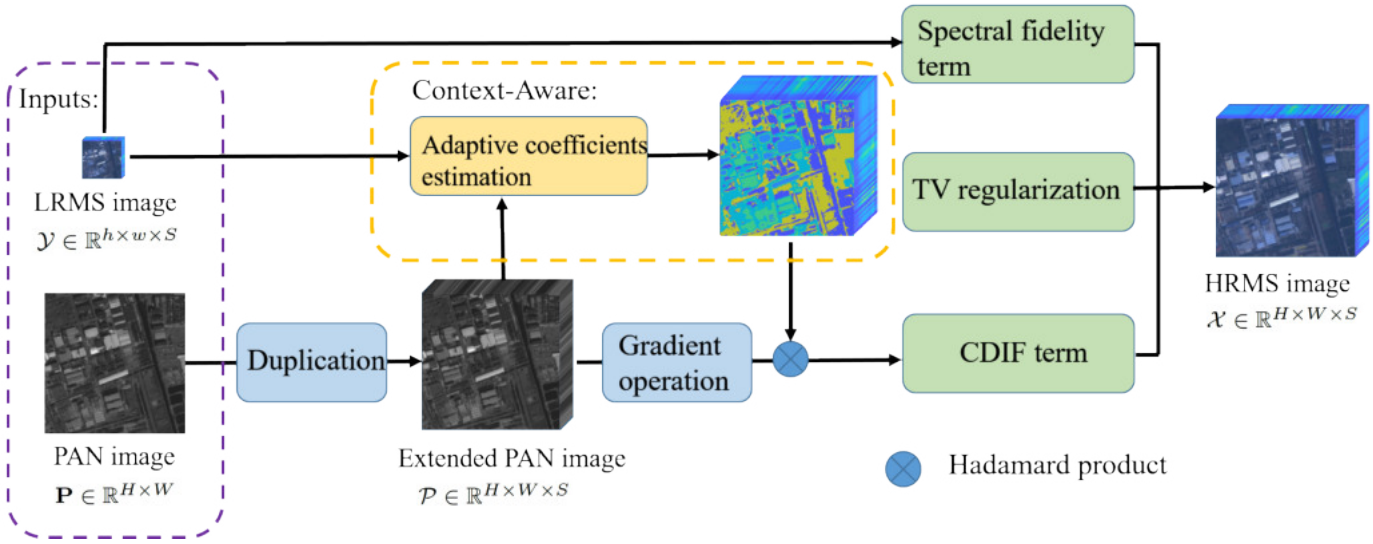


Fig. 2. The framework of our model. The details of our framework can be found in Sect. III.

methods. More details about the above-mentioned categories can be found in [1].

The CS methods, commonly considered classic approaches for pansharpening, are based on the substitution of one or more components after applying a reversible spectral transformation. This latter is applied to the MS image with the aim of separating the spatial and spectral information. The spatial information is substituted with the PAN image to enhance the spatial resolution of the MS image. Some instances of these approaches are the Brovey transform [5], the principal component analysis (PCA) [6], the intensity-hue-saturation (IHS) [7], the Gram-Schmidt (GS) spectral sharpening [8], the partial replacement adaptive component substitution (PRACS) [9], and the band-dependent spatial detail (BDSD) [10] methods. For these techniques, the tradeoff between spatial and spectral distortions is impossible to be overcome. Thus, CS methods have a good spatial fidelity, but paying it with a greater spectral distortion.

The MRA methods rely upon the injection of spatial details extracted from the PAN image into the MS image. Unlike CS methods, they can preserve spectral information, but causing (usually evident) spatial distortions. Spatial details can be extracted from different decomposition methods. Some instances of approaches in the MRA class are: the smoothing filter-based intensity modulation (SFIM) [11], the decimated wavelet transform (DWT) [12], the undecimated wavelet transform (UDWT) [13], the "à-trous" wavelet transform (ATWT) [14], the generalized Laplacian Pyramid [15], and some methods, i.e., [16] and [17], which propose strategies to extract details in an accurate way.

ML approaches have recently achieved great success in several image processing fields including pansharpening [18], [19], [20], [21], [22], [23], [24], [25]. Many ML methods for pansharpening, e.g., [26], [27], [28], [29], [30], have a strong ability in feature extraction. Hence, the relationship among MS, PAN, and HRMS images can be well-expressed by these methodologies. But, the other side of the coin is represented by i) the need of training data and computing resources, ii) a lack of robustness with respect to the changes of acquisition sensors and scenarios under analysis, and iii) the assumption usually made by these methods imposing that the relationship among MS, PAN, and HRMS images learned at reduced resolution is the same as that of at full resolution. Thus, the effectiveness of these approaches in addressing operative scenarios is often compromised. Besides, some advanced ML techniques, e.g., graph convolutional network (GCN) [31] and SpectralFormer [32], showed their potentialities in hyperspectral image processing [33]. These techniques can effectively extract feature information. However, their validity for the pansharpening problem needs to be explored and verified.

In recent years, VO methods have become more popular thanks to their flexibility [34]. They can show a good ability in modeling the relationship among MS, PAN, and HRMS images. Ballester *et al.* in [35] assumed that the PAN image is a linear combination of the different bands of the HRMS image, thus proposing the P+XS model. In this category, we can express the connections among MS, PAN, and HRMS in several domains. For example, Deng *et al.* in [36] built the model based on reproducible kernels in the Hilbert space, Fu *et al.* in [37] proposed a VO model based on a local gradient constraint, and Zhuang *et al.* in [38] used gradient domain guided image filtering for both preserving image structures and suppressing artifacts and noise. Moreover, the use of different norms is sometimes considered. For instance, Deng *et al.* in [39] proposed a pansharpening model with a hyper-Laplacian prior using a ℓ_p $(0 < p < 1)$ norm to describe the relationship between the HRMS image and the upsampled MS image in the gradient domain [40], instead, Wu *et al.* in [41] applied a $\ell_{2,1}$ norm to constrain the HRMS image and the PAN image in the gradient domain, again. VO methods can improve spatial information of the original MS image without affecting the spectral content by solving optimization problems. However, the definition of appropriated fidelity terms is a hard task,

(a) (b)

Fig. 3. (a) The HRMS image; (b) Clustering of the coefficients related to the HRMS image in (a).

resulting in a reduction of the performance when this operation is not properly performed.

B. Contributions

A context-aware details injection fidelity (CDIF) with adaptive coefficients estimation is proposed in this paper to obtain spatial structure features from the PAN image. The CDIF describes the relationship between PAN and HRMS images in the gradient domain. Unlike local and global approaches, we explore the relationship based on the image context. This latter can be drawn in an accurate way (please, see Sect. V-E3 for more details). The PAN image retains more spatial information than the original MS image. Thus, the CDIF can obtain this spatial information from the PAN image thanks to the exploitation of a context-based approach.

On the above basis, we integrate the proposed CDIF with two conventional fidelity and regularization terms to formulate a novel variational pansharpening model. The framework of the proposed model is shown in Fig. 2. Moreover, we design an algorithm that is based on the alternating direction method of multiplier (ADMM) [42] framework to effectively solve the proposed model. In our experiments, we compare this approach with some state-of-the-art methods on data acquired by different satellites, e.g., IKONOS, QuickBird, Pléiades, WorldView-2, and WorldView-3. The proposed model has shown good performance when applied to different satellite data getting a remarkable robustness.

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

- We propose a CDIF that can draw the context-aware relationship between PAN and HRMS images with adaptive coefficients estimation;
- A variational model including the proposed CDIF and two conventional fidelity and regularization terms is formulated for the task of remote sensing pansharpening. Moreover, an ADMM-based algorithm is designed to effectively solve the proposed model;
- Extensive experiments on several at reduced and at full resolution datasets demonstrate the superiority of the proposed approach compared with recent state-of-the-art pansharpening techniques.

TABLE I
SOME EXPLANATIONS ABOUT THE USED NOTATION.

Notation	Explanation
$\mathcal{X}, \mathbf{X}, \mathbf{x}, x$	Tensor, matrix, vector, scalar
$\mathcal{X} \in \mathbb{R}^{H\times W\times S}$	The HRMS image
$\mathbf{X} \in \mathbb{R}^{S\times HW}$	The mode-3 unfolding of $\mathcal{X}$
$\mathcal{X}_i \in \mathbb{R}^{H\times W}$	The i-th band of the HRMS image $\mathcal{X}$
$\mathcal{Y} \in \mathbb{R}^{h\times w\times S}$	The LRMS image
$\mathbf{Y} \in \mathbb{R}^{S\times hw}$	The mode-3 unfolding of the LRMS image $\mathcal{Y}$
$\mathbf{P} \in \mathbb{R}^{H\times W}$	The PAN image
$\mathcal{P} \in \mathbb{R}^{H\times W\times S}$	The extended PAN image, $\mathbf{P}$, with S bands
$\overline{\mathbf{P}} \in \mathbb{R}^{S\times HW}$	The mode-3 unfolding of the extended PAN image
$\mathbf{G} \in \mathbb{R}^{S\times 2HW}$	The adaptive coefficient
$\circ$	Hadamard product
$\oslash$	The element-wise division
$\nabla_i, i = 1, 2, 3.$	The gradient operation along the i-th direction
$\nabla = \begin{bmatrix} \nabla_1 \\ \nabla_2 \end{bmatrix}$	The gradient operation along the spatial directions
$\nabla\mathcal{X} \in \mathbb{R}^{2H\times W\times S}$	$\mathcal{X}$ in the gradient domain
$\nabla\mathbf{X} \in \mathbb{R}^{S\times 2HW}$	The mode-3 unfolding of $\nabla\mathcal{X}$

C. Organization

The rest of the paper is organized as follows. Sect. II briefly introduces the notation and the motivations behind this work. In Sect. III, we give an interpretation of the proposed model. The algorithm designed to solve the proposed model is described in Sect. IV. Instead, Sect. V is devoted to the experimental results with a particular emphasis on analyzing the parameters and discussing some details about the proposed method. Finally, concluding remarks are drawn in Sect. VI.

II. Notation and Motivations

The notation is summarized in Tab. I. Instead, the motivations behind this work are introduced in the following section.

A. Motivations

A widely used assumption performed in the literature is that the PAN image and the HRMS image share common spatial structures. Starting from this consideration, spatial structures are extracted from the PAN image using several methodologies. Many previous works consider the relationship between the PAN and the HRMS images in a direct way. Just for instance, the P+XS method [35] assumes that the PAN image is the weighted sum of the HRMS bands. However, this direct approach can lead to a reduction of the overall accuracy of the approach. Thus, we propose to extract the spatial information from the PAN image, but working in the gradient domain. Information in the gradient domain has demonstrated its usefulness for low-level vision tasks, including pansharpening, since it can describe, in a better way, dominant image structures to represent crucial image features. As shown in Fig. 1, it is evident that the PAN image approximates, in a better way, the HRMS bands in the gradient domain with respect to the original intensity domain. Hence, we can obtain spatial features in an easier way working in the gradient domain than the original one. This cue motivated us

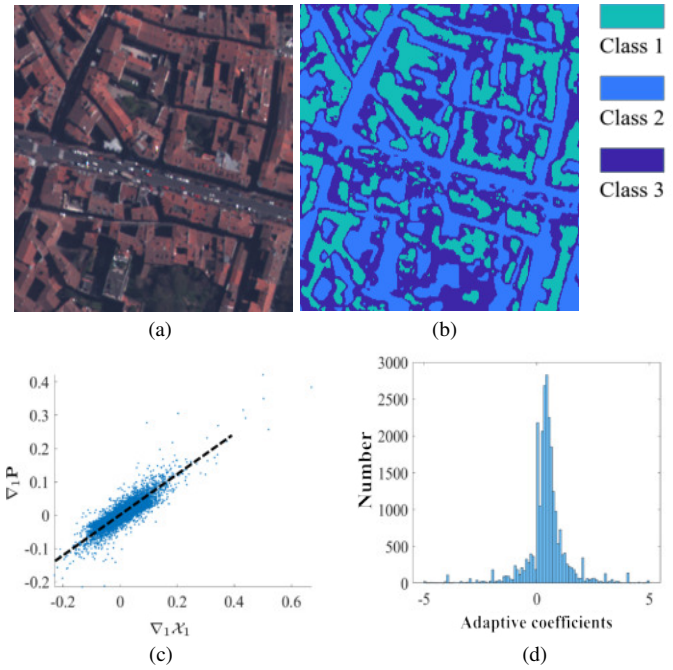


Fig. 4. (a) A close-up of the HRMS image of the Pléiades dataset; (b) Context-aware regions extracted from (a); (c) Scatter plot of $\nabla_1\mathcal{X}_1$ and $\nabla_1\mathbf{P}$ for pixels belonging to Class 1 in (b); (d) A histogram of $\nabla_1\mathcal{X}_1 \oslash \nabla_1\mathbf{P}$ for pixels belonging to Class 1 in (b). It is worth to be remarked that the distribution of $\nabla_1\mathcal{X}_1$ and $\nabla_1\mathbf{P}$ is well described by a directly proportional function, i.e., the black dotted line in (c). For other regions and bands, there is a similar directly proportional relationship between HRMS and PAN images in the gradient domain. Thus, the values of $\nabla\mathbf{X} \oslash \nabla\overline{\mathbf{P}}$ are approximately the same inside each region and for each band.

to explore the relationship between PAN and HRMS images in the gradient domain, unlike what is usually proposed in the literature.

Many previous methods are focused on the global or the local relationship between the PAN image and the HRMS image. For instance, Wu *et al.* in [41] consider a global relationship, instead, Fu *et al.* in [37] investigate on a local linear relationship among image patches. However, global and local relationships are not generally enough to accurately describe the complex relationship between the two input data, see, e.g., Fig. 1 (b) where the relationship between each band of the HRMS image and the PAN image is pixel-dependent. Thus, in this paper, we propose a context-aware method for the following reasons. First, we found that the relationship between PAN and HRMS images in the gradient domain is closely related to the image context. To corroborate it, we chose a reduced resolution dataset to calculate the coefficients related to the HRMS and the PAN images in the gradient domain. Afterwards, we clustered them. As shown in Fig. 3, there is a remarkable similarity between the clustered results and the image context in the original data. Second, a simple linear model can be exploited in this case to describe the relationship between each band of the HRMS image and the PAN image. Indeed, as shown in Figs. 4 (c) and (d), the relationship between the PAN image and the HRMS image is approximately linear inside the same context-aware region. Therefore, the above-mentioned observations motivate us to formulate the context-aware details injection fidelity.

III. The Proposed Model

A. The Spectral Fidelity Term

Many previous methods directly up-sample the MS image to obtain the same size image as the PAN image [43]. Thus, a fidelity term is built based on the upsampled version of the MS image and the PAN image. However, the spectral content obtained by exploiting this fidelity term is often inaccurate. For this reason, we consider a different spectral fidelity term relied upon the following model:

$$\mathbf{XBS} = \mathbf{Y} + \xi_1, \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{HW\times HW}$ denotes a blurring matrix, $\mathbf{S} \in \mathbb{R}^{HW\times hw}$ represents the decimation operation, and ξ_1 indicates a zero-mean Gaussian noise. The blur operation is equivalent to the convolution of the HRMS image with the point spread function (PSF) of the MS sensor [44], [45]. According to (1), the spectral fidelity term can be expressed as:

$$f_{spec} = \|\mathbf{XBS} - \mathbf{Y}\|_F^2, \tag{2}$$

where $\|\cdot\|_F$ is the Frobenius norm.

B. The Proposed Context-Aware Details Injection Fidelity

For the pansharpening problem, fidelity terms are at the basis of each model extracting spatial and spectral information from the two inputs, i.e., the PAN and the MS images. Many previous methods, e.g., [37], [40], prefer exploring the global or local relationship between PAN and HRMS images in the gradient domain. However, we found that the relationship between PAN and HRMS images in the gradient domain can be better described in a context-aware manner, see Figs. 3 and 4. Therefore, we assume that:

$$\nabla\mathbf{X} = \mathbf{G} \circ \nabla\overline{\mathbf{P}} + \xi_2, \tag{3}$$

where ∇ denotes the gradient operation along the two spatial directions, $\nabla\mathbf{X} \in \mathbb{R}^{S\times 2HW}$ stands for the mode-3 unfolding of $\nabla\mathcal{X}$, $\mathbf{G} \in \mathbb{R}^{S\times 2HW}$ represents the adaptive coefficients whose estimation is related to the image context, $\overline{\mathbf{P}}$ is the mode-3 unfolding of the extended PAN image (obtained by duplicating $\mathcal{P}$ along the spectral direction), $\circ$ indicates the Hadamard product, and ξ_2 is a zero-mean Gaussian noise. Hence, the CDIF is described by the following equation:

$$f_{CDIF} = \left\|\nabla\mathbf{X} - \mathbf{G} \circ \nabla\overline{\mathbf{P}}\right\|_F^2. \tag{4}$$

The CDIF with adaptive coefficients estimation is effective in extracting the features from the two inputs of the fusion process.

However, we cannot directly calculate $\mathbf{G}$ by (3) because the HRMS image, $\mathbf{X}$, is what we want to find. Thus, the adaptive coefficients $\mathbf{G}$ cannot be directly obtained by the following relationship:

$$\mathbf{G} = \nabla\mathbf{X} \oslash \nabla\overline{\mathbf{P}}. \tag{5}$$

To address this problem, we can exploit the LRMS image, $\mathbf{Y}$, which is instead known. We use an up-sampling operation to

Algorithm 1 The estimation of the adaptive coefficients $\mathbf{G}$.

Input: The upsampled version of the LRMS image, $\widetilde{\mathbf{Y}}$, the extended PAN image, $\overline{\mathbf{P}}$, the number of clusters, k.

1: Calculate $\mathbf{XB}$ via (6).
2: Calculate $\overline{\mathbf{P}}\mathbf{B}$ via (8).
3: **for** $j = 1 \rightarrow k$ **do**
4: Calculate $\mathbf{g}^j$ via (10).
5: **end for**

Output: $\mathbf{G}$ collects $\mathbf{g}^j$ for all the regions j.

simulate the inverse operation of the decimation operation, $\mathbf{S}$, and, thus, we have that:

$$\mathbf{XB} \approx \widetilde{\mathbf{Y}} = \Phi(\mathbf{Y}), \tag{6}$$

where $\widetilde{\mathbf{Y}}$ denotes the mode-3 unfolding of the upsampled version of the LRMS image and $\Phi(\cdot)$ is the upsampling function implemented by considering a polynomial interpolator with 23 coefficients [46], commonly exploiting for pansharpening [2], [47]. Thus, the original relationship in (3) can be reported at reduced resolution considering the classical hypothesis performed in developing pansharpening algorithms, that is the "invariance among scales". Thus, we approximately have that:

$$\nabla(\mathbf{XB}) \approx \mathbf{G} \circ \nabla(\overline{\mathbf{P}}\mathbf{B}), \tag{7}$$

where

$$\overline{\mathbf{P}}\mathbf{B} = \mathtt{MTF}(\overline{\mathbf{P}}) \tag{8}$$

and $\mathtt{MTF}(\cdot)$ is a Gaussian filter matched with the modulation transfer function (MTF) of the PAN image [2], [46], [48].

$\mathbf{G}$ links the PAN image and the HRMS image. To estimate it, we search for similar pixels' locations using a clustering method, i.e., the k-means clustering algorithm. The k-means approach needs to set the value of k (i.e., the number of clusters). To balance the computational burden and the clustering accuracy of the k-means algorithm, we set k to 5. Thus, all the pixels in the image are divided into k regions denoted by $\omega^j, j = 1, 2, ..., k$. As shown in Fig. 4 (c), the values of the coefficients $\nabla\mathbf{X} \oslash \nabla\overline{\mathbf{P}}$ are approximately the same inside each region and for each band, which indicates that $\mathbf{G}$ should have values around 1 considering (3). Accordingly, we have that:

$$\mathbf{g}^j \circ \mathbf{c}^{(n)} = \mathbf{d}^{(n)}, \ \forall n \in \omega^j, \tag{9}$$

where $\mathbf{c}^{(n)}$ and $\mathbf{d}^{(n)}$ are the pixels of $\nabla(\overline{\mathbf{P}}\mathbf{B})$ and $\nabla(\mathbf{XB})$, respectively, inside region j, and $\mathbf{g}^j$ denotes the adaptive coefficients for the j-th region, i.e., $\mathbf{G}$ evaluated for the j-th region. The overdetermined equation (9) holds for all the pixels n belonging to the j-th region, ω^j. The regression-based method, i.e., the ordinary least-squares, is selected to solve (9) for each spectral band. Therefore, we can estimate $\mathbf{g}^j$, as follows:

$$\mathbf{g}^j = \mathtt{Reg}(\{\mathbf{d}^{(n)}, \mathbf{c}^{(n)}\}_{n \in \omega^j}), \tag{10}$$

where $\mathtt{Reg}(\cdot, \cdot)$ denotes the ordinary least-squares regression using the inputs $\mathbf{d}^{(n)}$ and $\mathbf{c}^{(n)}$ for all $n \in \omega^j$ and separately applying it to all the spectral bands, thus estimating the coefficients of the vector $\mathbf{g}^j$ [49].

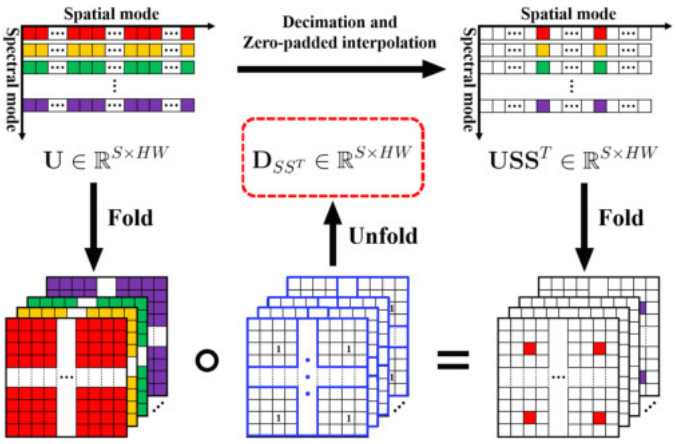


Fig. 5. The graphic representation of (22) for a scale ratio equal to 4. The white squares with a blank content indicate zero values. The first row shows the processing from $\mathbf{U}$ to $\mathbf{USS}^{\mathrm{T}}$, which is equal to an element-wise multiplication between $\mathbf{D}_{SS^{\mathrm{T}}}$ and $\mathbf{U}$, i.e., the second row. It is worth noting that $\mathbf{D}_{SS^{\mathrm{T}}}$ is produced from sparse matrices (i.e., the blue squares), whose entries are 1 only in one position [44].

Finally, $\mathbf{G}$ is obtained by collecting $\mathbf{g}^j$ (defined in (10)) for all the regions j. The adaptive coefficient process is summarized in Algorithm 1.

C. The Total Variation Regularization

The total variation (TV) regularization is a conventional regularization term in the field of image processing [50]. This regularization can keep the piecewise-constant of the result. The anisotropic TV regularization at pixel (i, j, k) is defined as follows:

$$\begin{aligned} \|\nabla_1 \mathcal{X}(i,j,k)\|_1 &= \|\mathcal{X}(i+1,j,k) - \mathcal{X}(i,j,k)\|_1, \\ \|\nabla_2 \mathcal{X}(i,j,k)\|_1 &= \|\mathcal{X}(i,j+1,k) - \mathcal{X}(i,j,k)\|_1, \\ \|\nabla_3 \mathcal{X}(i,j,k)\|_1 &= \|\mathcal{X}(i,j,k+1) - \mathcal{X}(i,j,k)\|_1. \end{aligned} \tag{11}$$

We apply TV regularization to describe the properties of the HRMS images, i.e., the piecewise-constant and the sparsity in the gradient domain. The TV regularization term is as follows:

$$f_{TV} = \beta_1 \|\nabla_1 \mathbf{X}\|_1^2 + \beta_2 \|\nabla_2 \mathbf{X}\|_1^2 + \beta_3 \|\nabla_3 \mathbf{X}\|_1^2, \tag{12}$$

where $\nabla_i \mathbf{X}$ is the mode-3 unfolding of $\nabla_i \mathcal{X}$ and β_1, β_2, and β_3 are positive parameters.

D. The Proposed Model

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

$$\begin{aligned} \min_{\mathbf{X}} \ & \|\mathbf{XBS} - \mathbf{Y}\|_F^2 + \lambda \left\|\nabla\mathbf{X} - \mathbf{G} \circ \nabla\overline{\mathbf{P}}\right\|_F^2 \\ & + \beta_1 \|\nabla_1 \mathbf{X}\|_1^2 + \beta_2 \|\nabla_2 \mathbf{X}\|_1^2 + \beta_3 \|\nabla_3 \mathbf{X}\|_1^2. \end{aligned} \tag{13}$$

This model is convex, but the direct calculation requires a huge computational burden. Thus, we designed an algorithm based on the ADMM [42] framework.

IV. The Proposed Algorithm

The ADMM [42] framework is one of the widely used methods to solve structured convex optimization problems. We designed a fast and effective ADMM-based algorithm to solve the proposed model. Let us introduce the auxiliary variables $\mathbf{U}$, $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{H}_3$, where $\mathbf{U}=\mathbf{XB}$, $\mathbf{H}_1=\nabla_1\mathbf{X}$, $\mathbf{H}_2=\nabla_2\mathbf{X}$, and $\mathbf{H}_3=\nabla_3\mathbf{X}$. The optimization model can be reconstructed as

$$\begin{aligned}\min_{\mathbf{X},\mathbf{U},\mathbf{H}_1,\mathbf{H}_2,\mathbf{H}_3}\ & \|\mathbf{US}-\mathbf{Y}\|_F^2+\lambda\left\|\nabla\mathbf{X}-\mathbf{G}\circ\nabla\overline{\mathbf{P}}\right\|_F^2\\ &+\beta_1\|\mathbf{H}_1\|_1^2+\beta_2\|\mathbf{H}_2\|_1^2+\beta_3\|\mathbf{H}_3\|_1^2\\ s.t.\quad & \mathbf{U}=\mathbf{XB},\ \mathbf{H}_1=\nabla_1\mathbf{X},\\ & \mathbf{H}_2=\nabla_2\mathbf{X},\ \mathbf{H}_3=\nabla_3\mathbf{X}.\end{aligned}\tag{14}$$

Thus, the augmented Lagrangian function is as follows

$$\begin{aligned}\mathcal{L}=&\|\mathbf{US}-\mathbf{Y}\|_F^2+\lambda\left\|\nabla\mathbf{X}-\mathbf{G}\circ\nabla\overline{\mathbf{P}}\right\|_F^2\\ &+\frac{\eta_1}{2}\left\|\mathbf{XB}-\mathbf{U}+\frac{\mathbf{\Lambda_1}}{\eta_1}\right\|_F^2\\ &+\beta_1\|\mathbf{H}_1\|_1^2+\beta_2\|\mathbf{H}_2\|_1^2+\beta_3\|\mathbf{H}_3\|_1^2\\ &+\frac{\eta_2}{2}\left\|\nabla_1\mathbf{X}-\mathbf{H}_1+\frac{\mathbf{\Lambda_2}}{\eta_2}\right\|_F^2+\frac{\eta_3}{2}\left\|\nabla_2\mathbf{X}-\mathbf{H}_2+\frac{\mathbf{\Lambda_3}}{\eta_3}\right\|_F^2\\ &+\frac{\eta_4}{2}\left\|\nabla_3\mathbf{X}-\mathbf{H}_3+\frac{\mathbf{\Lambda_4}}{\eta_4}\right\|_F^2+const,\end{aligned}\tag{15}$$

where $\mathbf{\Lambda_1}$, $\mathbf{\Lambda_2}$, $\mathbf{\Lambda_3}$, and $\mathbf{\Lambda_4}$ represent the Lagrange multipliers, and $const$ denotes a constant that is independent from the variables $\mathbf{U}$, $\mathbf{X}$, and $\mathbf{H}_i$, $i=1,2,3$. The optimization problem is solved by dealing with five sub-problems. Each sub-problem is solved by minimizing one variable under the condition that the others are fixed.

A. $\mathbf{X}$-sub-problem

According to (15), the $\mathbf{X}$-sub-problem is expressed as the following optimization problem:

$$\begin{aligned}\min_{\mathbf{X}}\ &\lambda\left\|\nabla\mathbf{X}-\mathbf{G}\circ\nabla\overline{\mathbf{P}}\right\|_F^2+\frac{\eta_1}{2}\left\|\mathbf{XB}-\mathbf{U}+\frac{\mathbf{\Lambda_1}}{\eta_1}\right\|_F^2\\ &+\frac{\eta_2}{2}\left\|\nabla_1\mathbf{X}-\mathbf{H}_1+\frac{\mathbf{\Lambda_2}}{\eta_2}\right\|_F^2+\frac{\eta_3}{2}\left\|\nabla_2\mathbf{X}-\mathbf{H}_2+\frac{\mathbf{\Lambda_3}}{\eta_3}\right\|_F^2\\ &+\frac{\eta_4}{2}\left\|\nabla_3\mathbf{X}-\mathbf{H}_3+\frac{\mathbf{\Lambda_4}}{\eta_4}\right\|_F^2.\end{aligned}\tag{16}$$

There are only Frobenius norms in the optimization model above. Thus, it can be solved by the fast Fourier transform (FFT) algorithm under the periodic boundary condition:

$$\mathbf{X}^{k+1}:=\mathcal{F}^{-1}\left(\frac{\mathcal{F}\left(\mathbf{W}^k+\mathbf{Q}^k\right)}{\mathcal{F}\left(\mathbf{O}\right)}\right),\tag{17}$$

with

$$\mathbf{W}^k=2\lambda\nabla^{\mathrm{T}}\left(\mathbf{G}\circ\nabla\overline{\mathbf{P}}\right)+\eta_1\mathbf{U}^k\mathbf{B}^{\mathrm{T}}+\sum_{i=2}^{4}\left(\eta_i\nabla_{i-1}^{\mathrm{T}}\mathbf{H}_{i-1}^k\right),\tag{18}$$

Algorithm 2 The ADMM-based solver for the proposed pansharpening model (13).

Input: The LRMS image, $\mathbf{Y}$, the extended PAN image, $\overline{\mathbf{P}}$, the adaptive coefficients, $\mathbf{G}$, λ, β_1, β_2, β_3, η_1, η_2, η_3, η_4, r, k_{mit}, and ε.

Initialization: $\mathbf{X}^0=\Psi(\mathbf{Y},r),\mathbf{U}^0=\mathbf{\Lambda}_1^0=\mathbf{0},\mathbf{H}_1^0=\mathbf{H}_2^0=\mathbf{H}_3^0=\mathbf{\Lambda}_2^0=\mathbf{\Lambda}_3^0=\mathbf{\Lambda}_4^0=\mathbf{0},k=0$

while $k<k_{mit}$ and $RelCha>\varepsilon$ **do**
- Update $\mathbf{X}^{k+1}$ via (17).
- Update $\mathbf{U}^{k+1}$ via (23).
- Update $\mathbf{H}_i^{k+1}$ via (25), i=1,2,3.
- Update Lagrange multiplier $\mathbf{\Lambda}_j^{k+1}$ via (26), j=1,2,3,4.
- $k=k+1$.

end while

Output: The fused HRMS image $\mathbf{X}$

$$\mathbf{Q}^k=-{\mathbf{\Lambda_1}}^k\mathbf{B}^{\mathrm{T}}-\sum_{i=2}^{4}\left(\nabla_{i-1}^{\mathrm{T}}\mathbf{\Lambda}_i^k\right),\tag{19}$$

$$\mathbf{O}=2\lambda\nabla^{\mathrm{T}}\nabla+\eta_1\mathbf{BB}^{\mathrm{T}}+\sum_{i=2}^{4}\eta_i\nabla_{i-1}^{\mathrm{T}}\nabla_{i-1}\tag{20}$$

where $\mathcal{F}\left(\cdot\right)$ and $\mathcal{F}^{-1}\left(\cdot\right)$ denotes the fast Fourier transform and its inverse transformation, respectively, $\left(\cdot\right)^{\mathrm{T}}$ is the transpose operator, and the division is element-wise.

B. $\mathbf{U}$-sub-problem

The $\mathbf{U}$-sub-problem is as follows:

$$\min_{\mathbf{U}}\ \|\mathbf{US}-\mathbf{Y}\|_F^2+\frac{\eta_1}{2}\left\|\mathbf{XB}-\mathbf{U}+\frac{\mathbf{\Lambda_1}}{\eta_1}\right\|_F^2.\tag{21}$$

The solution to this optimization problem relies on handling the decimation operation, $\mathbf{S}$. $\mathbf{USS}^{\mathrm{T}}$ can be seen as an element-wise multiplication $\mathbf{D}_{SS^{\mathrm{T}}}$ on $\mathbf{U}$:

$$\mathbf{USS}^{\mathrm{T}}=\mathbf{U}\circ\mathbf{D}_{SS^{\mathrm{T}}},\tag{22}$$

as shown in Fig. 5 [44]. Thus, the method based on the element-wise division can efficiently solve the $\mathbf{U}$-sub-problem:

$$\mathbf{U}^{k+1}:=\frac{2\mathbf{YS}^{\mathrm{T}}+\eta_1\mathbf{X}^{k+1}\mathbf{B}+{\mathbf{\Lambda_1}}^k}{2\mathbf{D}_{SS^{\mathrm{T}}}+\eta_1\mathbf{1}}.\tag{23}$$

C. $\mathbf{H}_i$-sub-problem

For $\mathbf{H}_i$-sub-problems, the optimization problem can be formalized as follows:

$$\min_{\mathbf{H}_i}\ \frac{\eta_{i+1}}{2}\left\|\nabla_i\mathbf{X}-\mathbf{H}_i+\frac{\mathbf{\Lambda}_{i+1}}{\eta_{i+1}}\right\|_F^2+\beta_i\|\mathbf{H}_i\|_1^2.\tag{24}$$

The above problem can be solved by the soft-thresholding strategy [51]. Thus, we have that:

$$\mathbf{H}_i^{k+1}=\mathtt{Soft}\left(\mathbf{X}^{k+1}+\frac{\mathbf{\Lambda}_{i+1}^k}{\eta_{i+1}},\frac{\beta_i}{\eta_{i+1}}\right),\tag{25}$$

where $\mathtt{Soft}\left(a,b\right):=sign(a)\cdot max(|a|-b,0)$, i =1,2,3.

D. Updating Multipliers

Finally, the multipliers $\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$, $\mathbf{\Lambda}_3$, and $\mathbf{\Lambda}_4$ are updated as follows:

$$\begin{pmatrix} \mathbf{\Lambda}_1^{k+1} \\ \mathbf{\Lambda}_2^{k+1} \\ \mathbf{\Lambda}_3^{k+1} \\ \mathbf{\Lambda}_4^{k+1} \end{pmatrix} := \begin{pmatrix} \mathbf{\Lambda}_1^{k} \\ \mathbf{\Lambda}_2^{k} \\ \mathbf{\Lambda}_3^{k} \\ \mathbf{\Lambda}_4^{k} \end{pmatrix} + \begin{pmatrix} \eta_1 \left(\mathbf{X}^{k+1}\mathbf{B} - \mathbf{U}^{k+1} \right) \\ \eta_2 \left(\nabla_1 \mathbf{X}^{k+1} - \mathbf{H}_1^{k+1} \right) \\ \eta_3 \left(\nabla_2 \mathbf{X}^{k+1} - \mathbf{H}_2^{k+1} \right) \\ \eta_4 \left(\nabla_3 \mathbf{X}^{k+1} - \mathbf{H}_3^{k+1} \right) \end{pmatrix}. \tag{26}$$

The reconstructed model, i.e., (14), is convex for each variable satisfying the hypotheses under the convergence theorem in [42]. Therefore, the convergence of the whole algorithm can be guaranteed. The relative change ($RelCha$) and the number of iterations are used as termination conditions of the algorithm. The $RelCha$ is defined as:

$$RelCha = \left\| \mathbf{X}^{k+1} - \mathbf{X}^{k} \right\|_F / \left\| \mathbf{X}^{k} \right\|_F . \tag{27}$$

The whole optimization algorithm can be summarized in Algorithm 2, where k_{mit} denotes the maximum iteration, r is the scale ratio between MS and PAN images, ε is a tolerance value, and Ψ denotes the bicubic interpolation.

V. Experimental Results

This section is devoted to the demonstration of the effectiveness of the proposed algorithm. We compare the proposed method with some state-of-the-art methods on data coming from different sensors, i.e., IKONOS, QuickBird, Pléiades, WorldView-2, and WorldView-3. To make the experimental analysis more representative, we chose methods from different pansharpening classes. Two assessment procedures, i.e., at reduced resolution and at full resolution, are considered in this paper to show the high performance of the proposed approach. All the methods in the benchmark are run on the same software and hardware platforms, i.e., MATLAB (R2016b) and a computer of 16Gb RAM with AMD Ryzen7-4800H, NVIDIA GeForce GTX 1650, and Radeon Graphics 2.90 GHz. The procedure to adjust the parameters is also described in this section together with an ablation study in the discussions section, even comparing the proposed approach with the direct estimation and the equivalent methodology based on patches. It is worth to be remarked that k_{mit}, r, and ε are set to 100, 4, and 2×10^{-5}, respectively, in all the experiments.

A. Quality Metrics

There are many metrics to quantitatively measure the performance at reduced resolution and at full resolution. We use the peak signal-to-noise ratio (PSNR) [52], the structural similarity index (SSIM) [52], the spectral angle mapper (SAM) [53], the spatial correlation coefficient (SCC) [54], the erreur relative globale adimensionnelle de synthèse (ERGAS) [55], and the $Q2^n$ index [56], [57] (i.e., the Q4 for 4-bands MS images or the Q8 for 8-bands MS data) for the assessment at reduced resolution. Instead, at full resolution, where no reference image is available for validation, the quality with no reference (QNR) [58] index consisting of a spectral quality index, D_λ, and a spatial quality index, D_s, is adopted.

B. Benchmark

We compare the proposed technique with some state-of-the-art pansharpening approaches. For CS methods, we choose for comparison the partial replacement adaptive component substitution (PRACS) [9], the robust band-dependent spatial detail (RBDSD) [49]. The generalized Laplacian Pyramid (GLP) framework is a widely used approach for MRA methods. In this class, the context-based GLP (C-GLP) [59] and the GLP with a full scale regression model (GLP-Reg-FS) [43] are considered. Moreover, the compared methods also include the VO method with local gradient constraints (LGC) [37], the high-quality Bayesian pansharpening (HQBP) [60], the pansharpening neural network (PNN) [19] and its target-adaptive version (TPNN) [61]. It is worth to be noted that we used the PNN and the TPNN distributed in [1] with parameters shared by authors. All compared methods are summarized as follows:

- CS methods: PRACS [9] and RBDSD [49].
- MRA methods: C-GLP [59] and GLP-Reg-FS [43].
- VO methods: LGC [37] and HQBP [60].
- ML methods: PNN [19] and TPNN [61].

C. The Reduced Resolution Assessment

This section is devoted to the comparison of the proposed method with the adopted benchmark on data at reduced resolution.

1) The Pléiades Test Case: The Pléiades dataset consists of two images (the PAN image and the four-band MS image with a spatial resolution of 0.5 m and 2 m, respectively) acquired by the Pléiades satellite mission. The results are shown in Fig. 6. The images in Figs. 6 (a), (e), and (g) introduce some spatial artifacts. Moreover, the images in Figs. 6 (b), (h), and (i), especially the one obtained by the proposed method, can effectively alleviate the spectral and spatial distortions. The residual images are depicted in Figs. 6 (k)-(t). It is easy to see that the proposed method gets the best results compared with the other approaches (see Fig. 6(s)). Tab. II reports the quantitative comparison for the different methods. These results corroborate the above-mentioned analysis showing that the proposed method gets the best index values with respect to the techniques in the proposed benchmark.

2) The QuickBird Test Case: The QuickBird dataset consists of a PAN image (spatial resolution of 0.7 m) and four MS bands (spatial resolution of 2.8 m) captured in the visible near-infrared spectrum. Fig. 7 depicts the fused results obtained by the compared approaches. Figs. 7 (a), (e), and (f) are fuzzier than the others and characterized by an evident spatial distortion. Instead, Fig. 7 (g) has a greater spatial accuracy, but paying it with a relevant spectral distortion. Figs. 7 (b), (h), and (i) have a better visual appearance compared with the other images. By observing the small red box in the upper-left corner of the picture, the proposed method can better represent the edges of the houses. Moreover, the residuals in Fig. 7 point out the superiority of the proposed method. The best result is shown in Fig. 7 (s), especially having a look at the upper-left area of the image. Finally, Tab. III reports the quantitative comparison among the methods in the benchmark. Again,

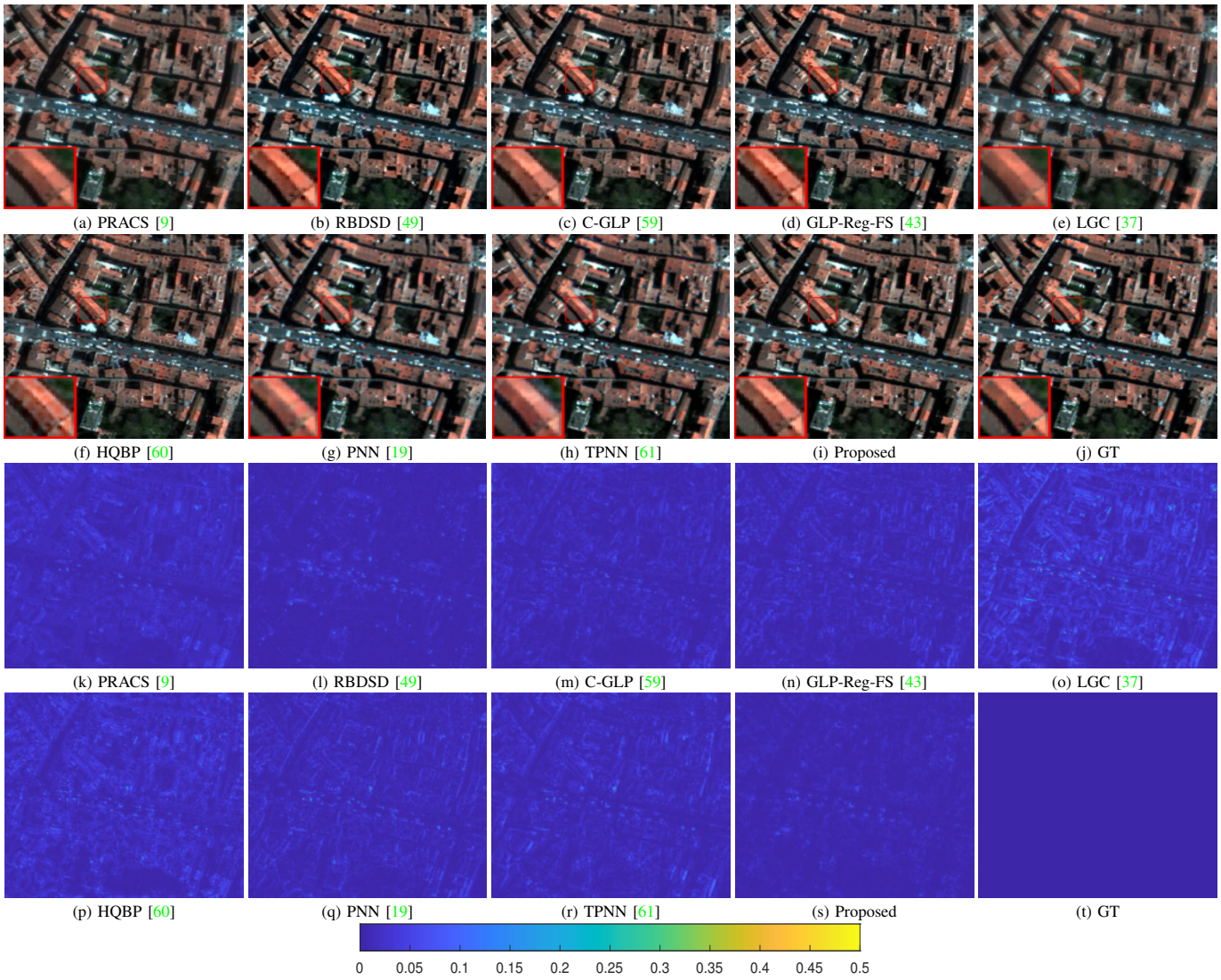


Fig. 6. Fusion results with close-ups for the compared approaches on the four-band Pléiades dataset at reduced resolution (size of the PAN image: 256 × 256). (a)-(j) Visual results in true colors of PRACS [9], RBDSD [49], C-GLP [59], GLP-Reg-FS [43], LGC [37], HQBP [60], PNN [19], TPNN [61], the proposed method, and the ground-truth (GT), respectively. (k)-(t) Residual images for the displayed red band.

TABLE II
QUANTITATIVE RESULTS FOR THE TEST CASE IN FIG. 6. TPNN [61] IS EXECUTED ON GPU (G). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q4	Time(s)
PRACS [9]	33.36	0.926	4.106	0.968	3.167	0.890	0.046
RBDSD [49]	<u>36.16</u>	<u>0.960</u>	<u>3.276</u>	<u>0.975</u>	<u>2.409</u>	0.941	<u>0.044</u>
C-GLP [59]	34.33	0.937	4.456	0.963	2.980	0.958	0.336
GLP-Reg-FS [43]	35.14	0.954	3.411	0.970	2.582	0.914	**0.038**
LGC [37]	31.98	0.910	3.897	0.958	3.595	0.918	7.722
HQBP [60]	33.62	0.918	4.456	0.965	3.055	0.946	13.38
PNN [19]	35.36	0.952	3.245	0.973	2.505	0.969	12.04
TPNN [61]	35.24	0.950	3.292	0.973	2.511	<u>0.968</u>	0.065(G)
Proposed	**37.35**	**0.965**	**2.884**	**0.981**	**2.040**	**0.978**	13.44
Ideal value	$+\infty$	**1**	**0**	**1**	**0**	**1**	-

it corroborates the previous analysis. Indeed, the proposed method shows the best indexes in Tab. III.

3) The IKONOS Toulouse Test Case: The IKONOS Toulouse dataset consists of a PAN image (spatial resolution of 1 m) and a four-band MS image (spatial resolution of 4 m) acquired by the IKONOS sensor in the visible near-infrared spectrum over the city of Toulouse, France. The visual results are shown in Fig. 8. Figs. 8 (c) and (e) suffer from spectral distortions and Figs. 8 (e), (f), and (h) are affected by spatial distortions. Instead, in Figs. 8 (b) and (i), the best results are shown. Tab. IV reports the quantitative results pointing out that the proposed approach is always in the first two positions in the ranking.

4) The WorldView-3 Test Case: The WorldView-3 dataset consists of two images: a PAN image with a spatial resolution of 0.31 m and an eight-band MS image with a spatial resolution of 1.24 m acquired by the WorldView-3 sensor in the visible near-infrared spectrum. Figs. 9 (a)-(j) show the true colors images of the fused products. It is clear to see that the proposed method obtains more details than the other approaches. In Figs. 9 (a), (c), (f), and (g), some spatial details

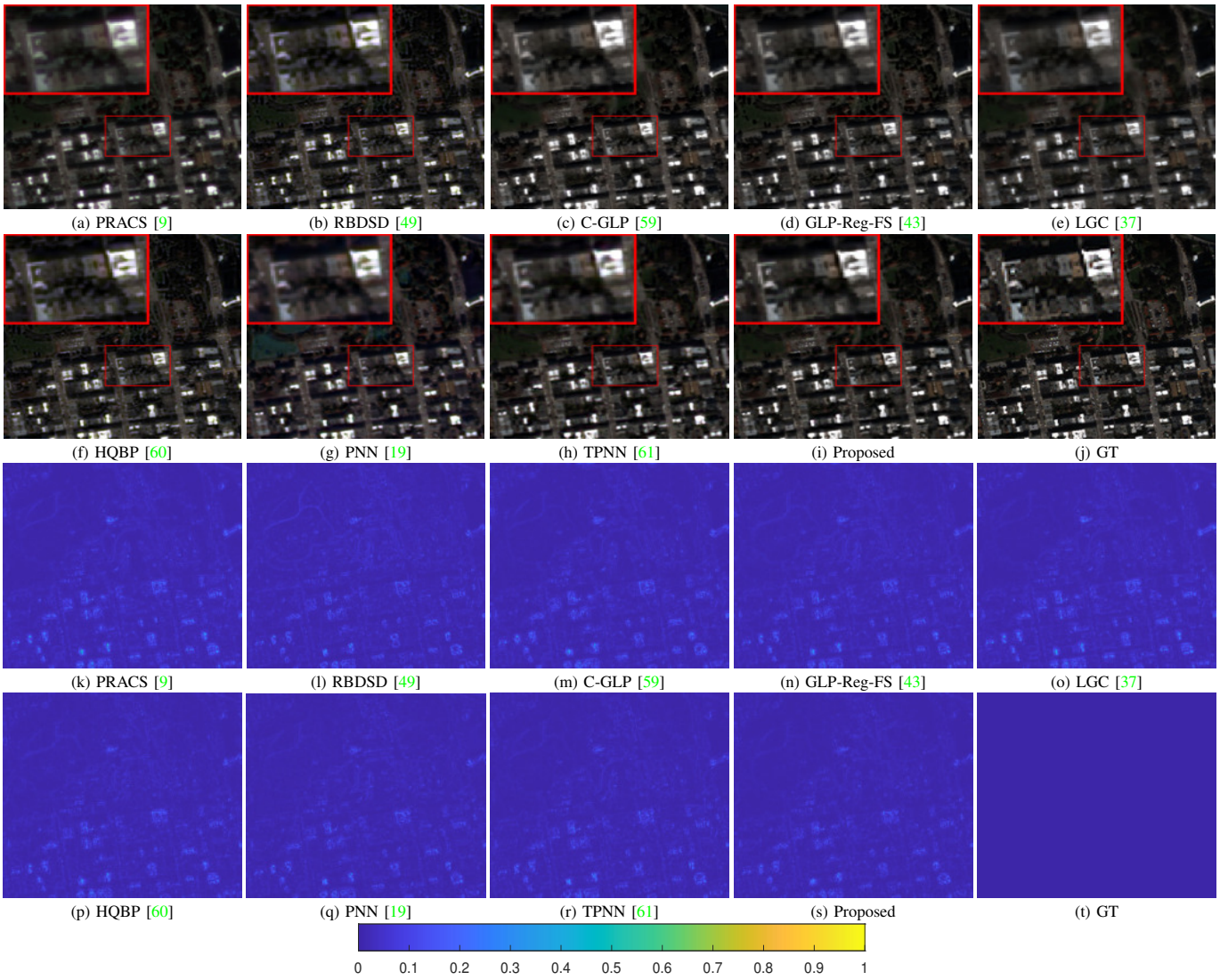


Fig. 7. Fusion results with close-ups for the compared approaches on the four-band QuickBird dataset at reduced resolution (size of the PAN image: 256 × 256). (a)-(j) Visual results in true colors of PRACS [9], RBDSD [49], C-GLP [59], GLP-Reg-FS [43], LGC [37], HQBP [60], PNN [19], TPNN [61], the proposed method, and the GT, respectively. (k)-(t) Residual images for the displayed red band.

TABLE III
QUANTITATIVE RESULTS FOR THE TEST CASE IN FIG. 7. TPNN [61] IS EXECUTED ON GPU (G). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q4	Time(s)
PRACS [9]	29.51	0.797	7.964	0.885	8.701	0.758	0.071
RBDSD [49]	30.46	0.814	8.150	0.898	7.795	0.778	**0.064**
C-GLP [59]	30.20	0.815	7.246	0.892	8.040	0.805	0.800
GLP-Reg-FS [43]	30.43	0.821	7.774	0.901	7.799	0.784	0.075
LGC [37]	29.52	0.788	7.659	0.884	8.737	0.761	9.797
HQBP [60]	30.23	0.795	8.807	0.897	7.994	0.768	13.35
PNN [19]	30.78	0.840	7.843	0.910	7.452	0.798	0.281
TPNN [61]	31.12	0.845	7.326	0.914	7.246	0.828	0.138(G)
Proposed	**31.54**	**0.852**	**6.792**	**0.922**	**6.896**	**0.849**	12.60
Ideal value	$+\infty$	**1**	**0**	**1**	**0**	**1**	-

are missing, as shown by the related close-ups in the figures. Instead, Figs. 9 (b), (e), and (h) depict some spatial artifacts in the vegetated area. Comparing the fused products with the ground-truth, we can observe that the proposed method achieves the good results. Residual images can better visualize the differences among the compared approaches, see Figs. 9 (k)-(t). It is easy to see that the edges of the houses in the scenario under test make several problems among the compared approaches. However, the proposed method still shows better results than the others, as further testified by the numerical results in Tab. V.

D. The Full Resolution Assessment

We conduct some experiments on data at full resolution to further assess the proposed method.

1) The IKONOS Sichuan Test Case: The IKONOS Sichuan dataset consists of a PAN image and a 4-bands MS image acquired by the IKONOS sensor over the Sichuan region in China. The size of the PAN image is 512×576. The results of the methods in the benchmark are displayed in Fig. 10. The proposed method is still valid for these 4-bands data at

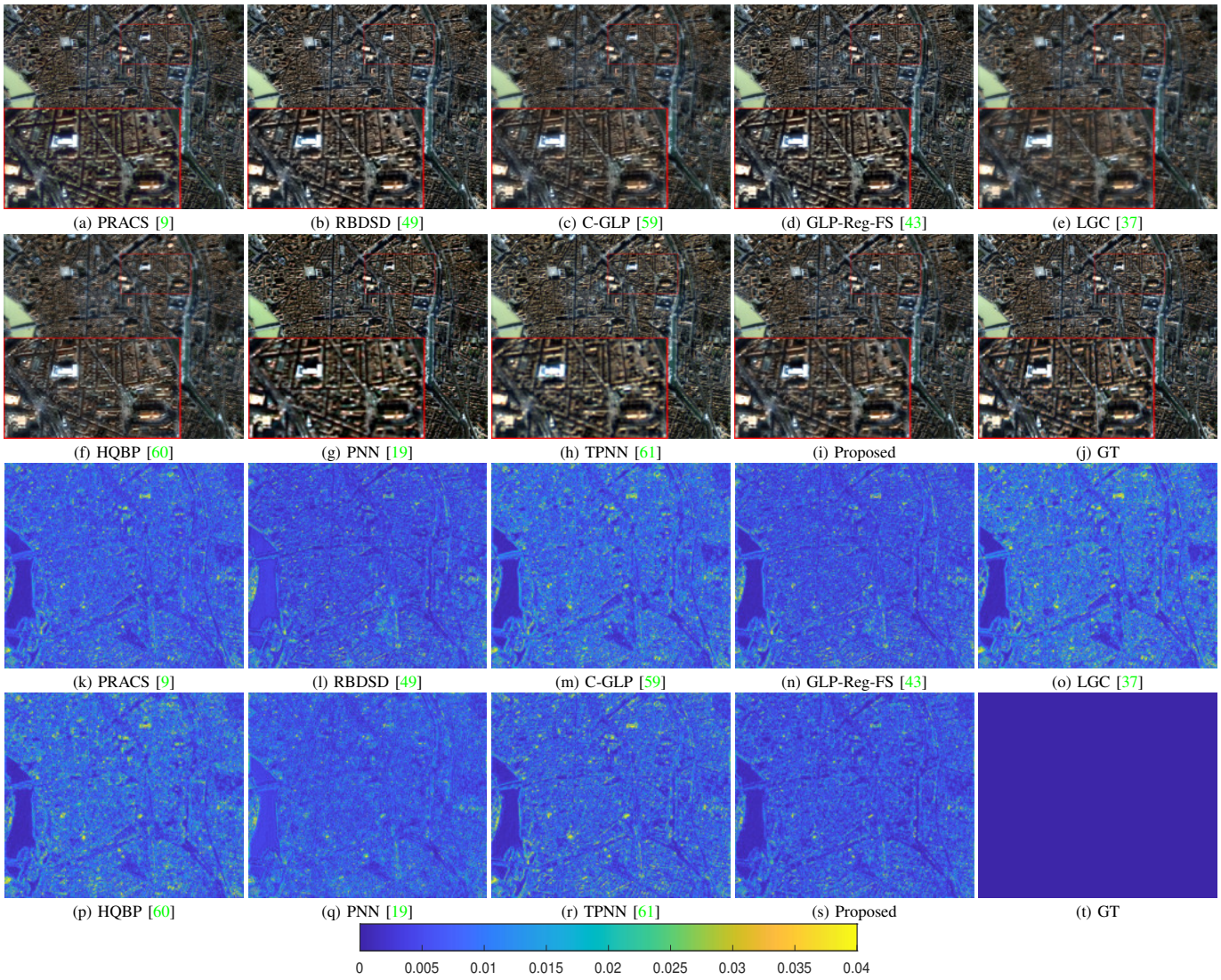


Fig. 8. Fusion results with close-ups for the compared approaches on the four-band IKONOS Toulouse dataset at reduced resolution (size of the PAN image: 512×512). (a)-(j) Visual results in true colors of PRACS [9], RBDSD [49], C-GLP [59], GLP-Reg-FS [43], LGC [37], HQBP [60], PNN [19], TPNN [61], the proposed method, and the GT, respectively. (k)-(t) Residual images for the displayed red band.

full resolution. Because of the absence of a GT image, the QNR index is calculated. As shown in Tab. VI, the proposed method can get the best overall accuracy, as measured by the QNR index.

2) The WorldView-2 Test Case: The WorldView-2 dataset is a full resolution set of data consisting of a PAN and an 8-bands MS data acquired by the WorldView-2 sensor. The MS image is acquired in the visible near-infrared with a spatial resolution of 1.85 m and the PAN image (depicted in Fig. 11) has a spatial resolution of 0.46 m. The size of the PAN image is 256×256. The fusion results are displayed in Fig. 11. The proposed method gets better results than the others. Indeed, by comparing it with the PAN image, the spatial details are well preserved. About the quantitative results, our approach gets the best values for all the considered metrics, as shown in Tab. VII.

TABLE IV
QUANTITATIVE RESULTS FOR THE TEST CASE IN FIG. 8. TPNN [61] IS EXECUTED ON GPU (G). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q4	Time(s)
PRACS [9]	38.98	0.935	3.476	0.922	3.833	0.851	0.347
RBDSD [49]	<u>40.08</u>	<u>0.954</u>	**2.598**	0.930	<u>2.988</u>	<u>0.909</u>	0.234
C-GLP [59]	38.14	0.924	3.275	0.896	3.976	0.818	2.815
GLP-Reg-FS [43]	39.66	0.950	2.848	0.924	3.177	0.896	<u>0.182</u>
LGC [37]	37.69	0.916	3.375	0.899	4.261	0.784	47.67
HQBP [60]	38.18	0.925	3.486	0.896	3.989	0.835	41.96
PNN [19]	39.50	0.953	3.204	<u>0.934</u>	3.357	**0.912**	0.443
TPNN [61]	38.10	0.929	3.390	0.887	4.054	0.842	**0.090(G)**
Proposed	**40.47**	**0.956**	<u>2.589</u>	**0.937**	**2.964**	0.907	47.27
Ideal value	**+∞**	**1**	**0**	**1**	**0**	**1**	**-**

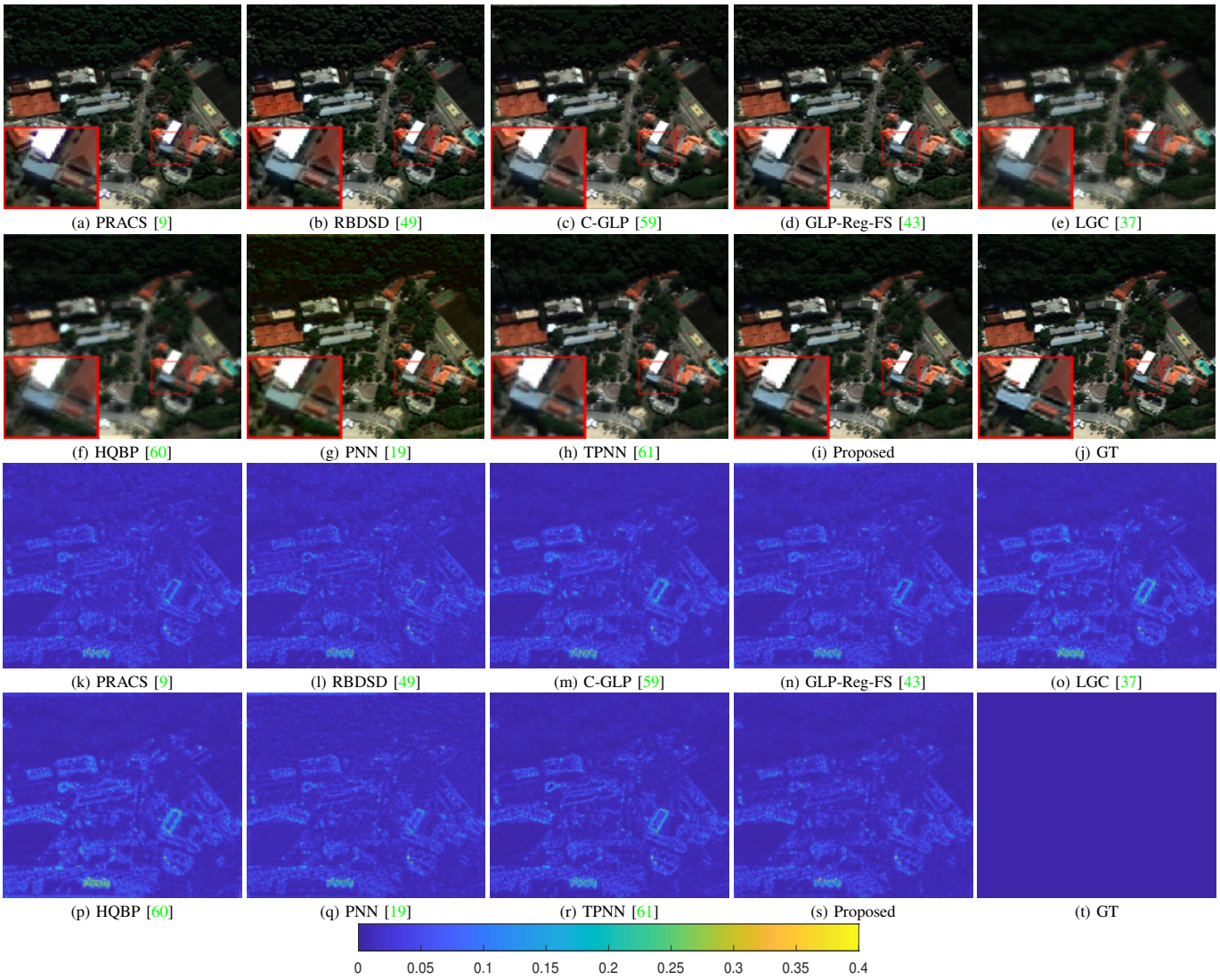


Fig. 9. Fusion results with close-ups for the compared approaches on the eight-band WorldView-3 dataset at reduced resolution (size of the PAN image: 256 × 256). (a)-(j) Visual results in true colors of PRACS [9], RBDSD [49], C-GLP [59], GLP-Reg-FS [43], LGC [37], HQBP [60], PNN [19], TPNN [61], the proposed method, and the GT, respectively. (k)-(t) Residual images for the displayed red band.

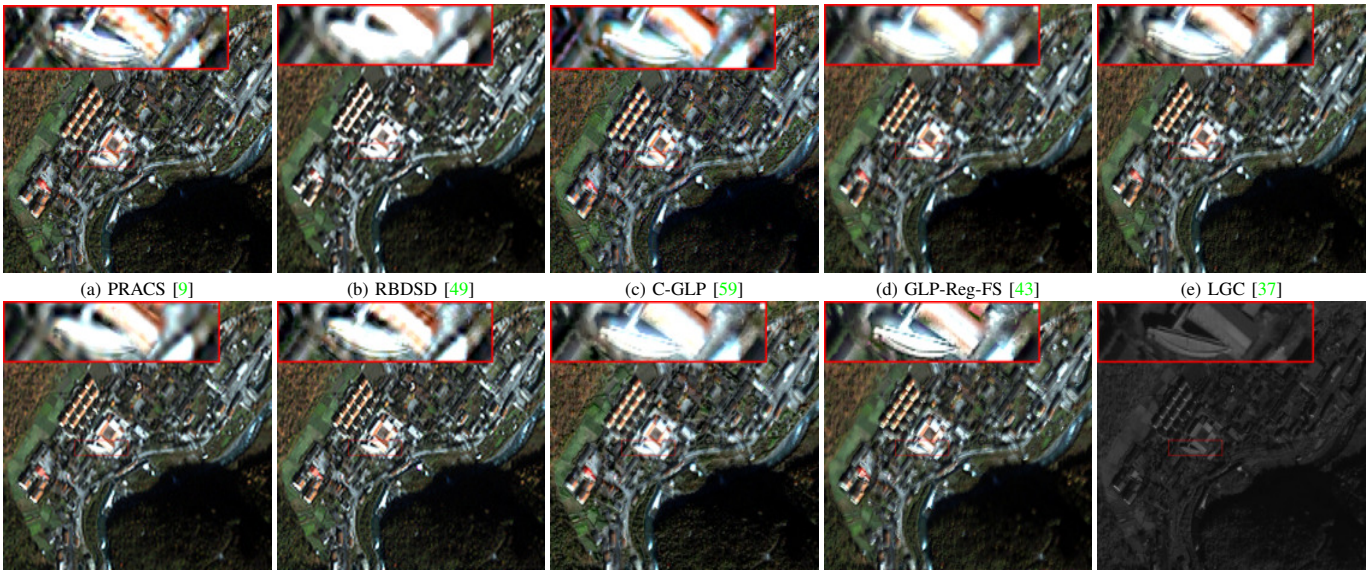
TABLE V
QUANTITATIVE RESULTS FOR THE TEST CASE IN FIG.9. TPNN [61] IS EXECUTED ON GPU (G). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q8	Time(s)
PRACS [9]	31.55	0.837	8.547	0.927	5.677	**0.929**	0.157
RBDSD [49]	32.24	0.858	8.540	0.936	5.214	0.907	**0.103**
C-GLP [59]	32.16	0.861	7.624	0.938	5.230	0.865	1.207
GLP-Reg-FS [43]	31.89	0.845	8.385	0.932	5.479	0.876	0.132
LGC [37]	31.24	0.845	7.082	0.930	5.713	0.832	17.24
HQBP [60]	30.79	0.808	7.924	0.918	6.117	0.819	24.47
PNN [19]	30.80	0.876	9.184	0.941	5.546	0.849	0.190
TPNN [61]	32.50	0.885	**6.520**	0.945	4.852	0.872	0.139(G)
Proposed	**33.30**	**0.896**	6.538	**0.951**	**4.558**	0.899	28.97
Ideal value	$+\infty$	**1**	**0**	**1**	**0**	**1**	-

E. Discussions

1) Analysis of Parameters: In this section, we analyze the parameters used in the proposed approach. Although there are many parameters, some similarities between the spatial dimensions of the TV regularization can be found. We can assume $\beta_1 = \beta_2$ and $\eta_2 = \eta_3$ for fine-tuning the parameters in an easier way. For a better visualization, the metrics are normalized by $(metric - \texttt{Mean}(metric))/\texttt{Std}(metric)$, where $\texttt{Mean}(\cdot)$ denotes the averaging operation and $\texttt{Std}(\cdot)$ denotes the standard deviation operation. We only adjust one parameter at a time, thus having all the others fixed. The test is performed by using the Pléiades dataset. As shown in Fig. 12, the parameters λ and η_1 are more sensitive than the others. Thus, we can fine-tune first the parameters λ and η_1, then adjusting the others. The parameters can be quickly fixed in this way, even if we have 8 parameters in our model. Considering the best point to balance all the different metrics, we selected the following set of parameters to be used in our experimental analysis: $\lambda = 5 \times 10^{-5}$, $\eta_1 = 1 \times 10^{-4}$, $\beta_1 = \beta_2 = 1 \times 10^{-7}$, $\eta_2 = \eta_3 = 1 \times 10^{-3}$, $\beta_3 = 1 \times 10^{-5}$, and $\eta_4 = 5 \times 10^{-8}$.

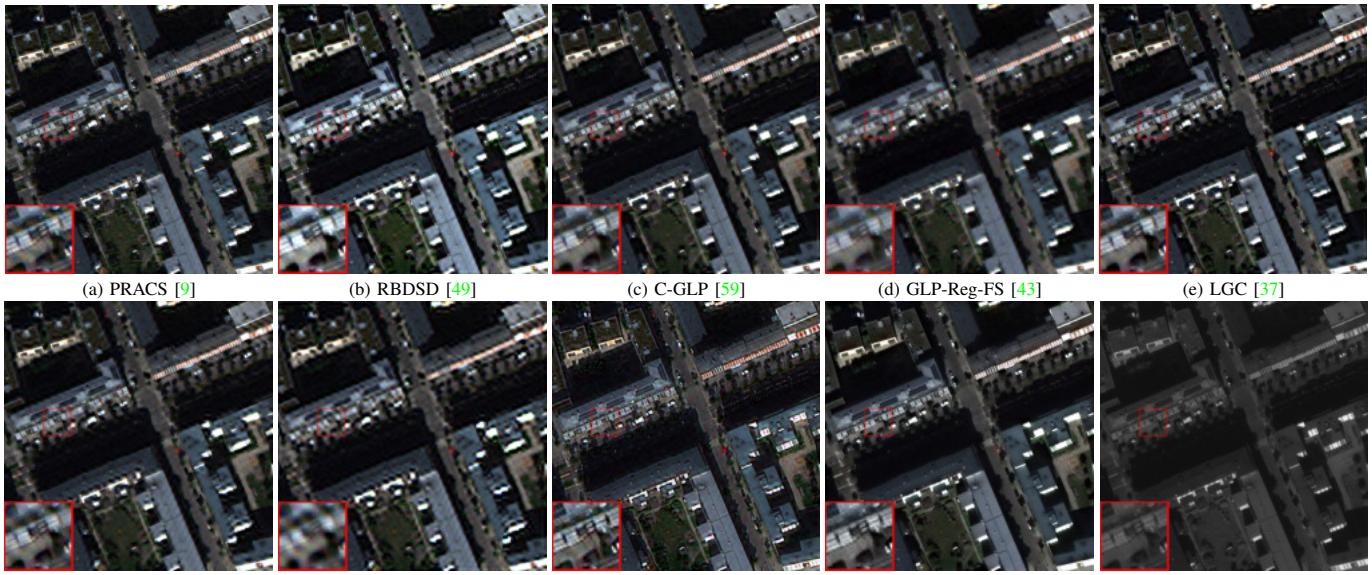
2) The Ablation Study: The proposed model consists of the CDIF, the spectral fidelity term, and the conventional TV



(a) PRACS [9] (b) RBDSD [49] (c) C-GLP [59] (d) GLP-Reg-FS [43] (e) LGC [37]

(f) HQBP [60] (g) PNN [19] (h) TPNN [61] (i) Proposed (j) PAN

Fig. 10. Fusion results with close-ups for the compared approaches on the four-band IKONOS Sichuan dataset at full resolution (size of the PAN image: 512 × 576). (a)-(j) Visual results in true colors of PRACS [9], RBDSD [49], C-GLP [59], GLP-Reg-FS [43], LGC [37], HQBP [60], PNN [19], TPNN [61], the proposed method, and the PAN image, respectively.



(a) PRACS [9] (b) RBDSD [49] (c) C-GLP [59] (d) GLP-Reg-FS [43] (e) LGC [37]

(f) HQBP [60] (g) PNN [19] (h) TPNN [61] (i) Proposed (j) PAN

Fig. 11. Fusion results with close-ups for the compared approaches on the eight-band WorldView-2 dataset at full resolution (size of the PAN image: 256 × 256). (a)-(j) Visual results in true colors of PRACS [9], RBDSD [49], C-GLP [59], GLP-Reg-FS [43], LGC [37], HQBP [60], PNN [19], TPNN [61], the proposed method, and the PAN image, respectively.

regularization. To analyze the role of these three different parts, the following three sub-models are considered:
Sub-model-I:

$$\min_{\mathbf{X}} \ \|\nabla\mathbf{X} - \mathbf{G}\circ\nabla\mathbf{P}\|_F^2 + \beta_1\|\nabla_1\mathbf{X}\|_1^2 + \beta_2\|\nabla_2\mathbf{X}\|_1^2 + \beta_3\|\nabla_3\mathbf{X}\|_1^2, \tag{28}$$

Sub-model-II:

$$\min_{\mathbf{X}} \ \|\mathbf{XBS} - \mathbf{Y}\|_F^2 + \beta_1\|\nabla_1\mathbf{X}\|_1^2 + \beta_2\|\nabla_2\mathbf{X}\|_1^2 + \beta_3\|\nabla_3\mathbf{X}\|_1^2, \tag{29}$$

Sub-model-III:

$$\min_{\mathbf{X}} \ \|\mathbf{XBS} - \mathbf{Y}\|_F^2 + \lambda\|\nabla\mathbf{X} - \mathbf{G}\circ\nabla\mathbf{P}\|_F^2. \tag{30}$$

We compared the three sub-models with the proposed model on the Pléiades dataset. As shown in Tab. VIII, all the three terms are necessary to get the highest performance and the importance of the two fidelity terms (i.e., the CDIF and the spectral fidelity term) is evident.

3) Analysis About the Context-aware Strategy: In this paper, we proposed the CDIF to extract spatial information in an effective way. Specifically, we estimated different adaptive coefficients, $\mathbf{G}$, for different context regions. As shown by

TABLE VI
QUANTITATIVE RESULTS FOR THE TEST CASE IN FIG. 10. TPNN [61] IS EXECUTED ON GPU (G). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	D_λ	D_s	QNR	Time(s)
PRACS [9]	0.029	0.249	0.730	0.419
RBDSD [49]	0.086	0.396	0.552	0.285
C-GLP [59]	0.032	0.076	0.895	3.948
GLP-Reg-FS [43]	0.096	0.274	0.656	0.184
LGC [37]	**0.008**	0.157	0.837	60.57
HQBP [60]	0.063	0.106	0.838	68.77
PNN [19]	0.094	0.267	0.663	0.332
TPNN [61]	0.012	0.026	0.963	**0.125(G)**
Proposed	0.020	**0.010**	**0.970**	54.76
Ideal value	**0**	**0**	**1**	-

TABLE VII
QUANTITATIVE RESULTS FOR THE TEST CASE IN FIG. 11. TPNN [61] IS EXECUTED ON GPU (G). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	D_λ	D_s	QNR	Time(s)
PRACS [9]	0.012	0.091	0.898	0.395
RBDSD [49]	0.028	0.041	0.932	0.325
C-GLP [59]	0.025	0.090	0.887	1.386
GLP-Reg-FS [43]	0.070	0.104	0.833	0.180
LGC [37]	0.016	0.060	0.925	53.49
HQBP [60]	0.031	0.019	0.951	65.89
PNN [19]	0.121	0.047	0.839	0.323
TPNN [61]	0.041	0.029	0.931	**0.164(G)**
Proposed	**0.011**	**0.014**	**0.975**	72.80
Ideal value	**0**	**0**	**1**	-

TABLE VIII
QUANTITATIVE RESULTS FOR THE ABLATION STUDY ON THE PLÉIADES DATASET. (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q4	Time(s)
Sub-model-I	15.53	0.192	15.730	0.831	22.351	0.208	13.35
Sub-model-II	21.20	0.519	34.867	0.655	38.372	0.596	12.80
Sub-model-III	37.30	**0.966**	2.814	0.980	2.045	0.977	**1.911**
Proposed	**37.35**	0.965	**2.884**	**0.981**	**2.040**	**0.978**	13.44
Ideal value	$+\infty$	**1**	**0**	**1**	**0**	**1**	-

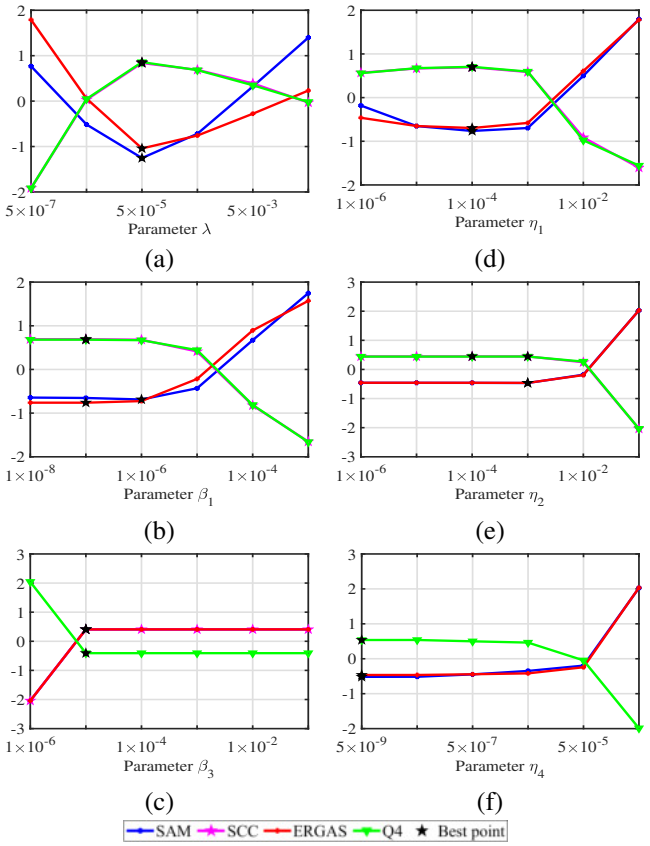


Fig. 12. Robustness analysis (using the SAM, the SCC, the ERGAS, the Q4 as metrics) for the parameters λ, β_1, β_3, η_1, η_2, and η_4 on the Pléiades dataset.

TABLE IX
QUANTITATIVE RESULTS CONSIDERING THREE DIFFERENT STRATEGIES TO ESTIMATE THE ADAPTIVE COEFFICIENTS, $\mathbf{G}$, ON THE PLÉIADES DATASET. (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q4	Time(s)
Direct	20.953	0.636	10.551	0.512	12.076	0.581	12.77
Local	34.795	0.937	3.056	0.953	2.580	0.963	**12.42**
Proposed	**37.353**	**0.965**	**2.884**	**0.981**	**2.040**	**0.978**	13.44
Ideal value	$+\infty$	**1**	**0**	**1**	**0**	**1**	-

(7), a direct way to estimate the adaptive coefficients is pixel-by-pixel. Thus, the adaptive coefficients, $\mathbf{G}$, can be calculated as:

$$\mathbf{G} = \nabla(\mathbf{XB}) \oslash \nabla(\overline{\mathbf{P}}\mathbf{B}), \tag{31}$$

where $\oslash$ denotes the element-wise division. A second way is by exploiting local patches. Thus, we can split the original data in patches estimating the adaptive coefficients, $\mathbf{G}$, for each patch.

Thus, we compare the CDIF solution with the two other methods on the Pléiades dataset, again. The quantitative results are summarized in Tab. IX. It is easy to see that the CDIF method shows the best performance.

4) Analysis About the Number of Clusters: The number of clusters in a context-aware strategy using a k-means clustering algorithm should be determined. Thus, we discuss, in this section, about the influence on our approach of the changing of the number of clusters. As shown in Fig. 13, the results are quite stable by changing the number of clusters for the Pléiades test case. Instead, better results can be obtained by increasing the number of clusters for the QuickBird and the WorldView-3 test cases. However, the better performance is paid by an increment of the computational burden. Thus, a good balance between the fusion accuracy and the computational burden can be found selecting a number of clusters equal to 5.

VI. CONCLUSIONS

In this paper, we focused our attention on the CDIF with adaptive coefficients estimation. More specifically, we applied a clustering method to divide the pixels of an image in different context-based regions. Afterwards, adaptive coefficients are estimated by using a regression-based method for each region.

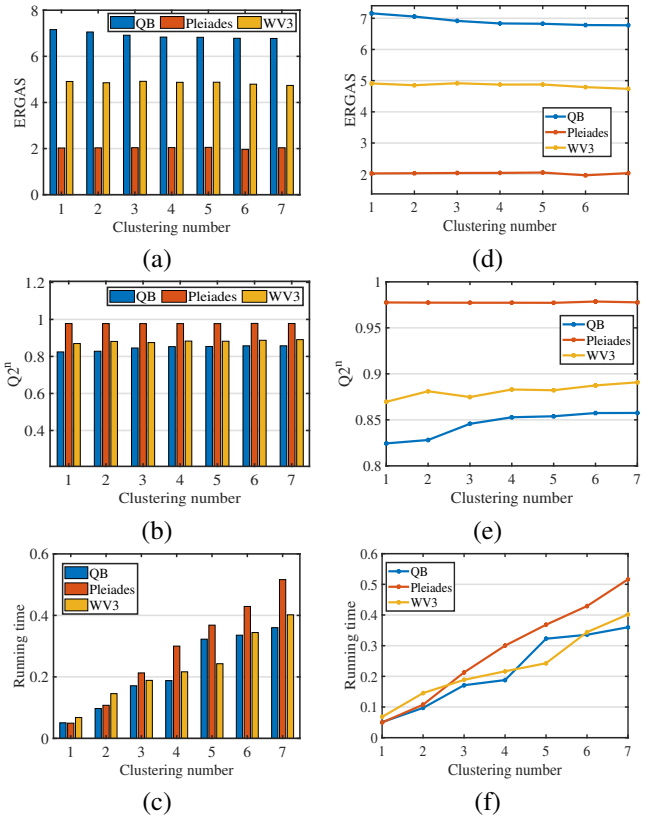


Fig. 13. Robustness analysis by changing the number of clusters. The ERGAS metric is used in (a) and (d). Instead, the $Q2^n$ is calculated in (b) and (e). Finally, the running times are shown in (c) and (f). In the figures, WV3 and QB stand for WorldView-3 and QuickBird, respectively.

In addition, the CDIF is combined with a conventional fidelity term and a total variation regularization to formulate a novel variational pansharpening model that is solved by designing an algorithm based on the ADMM framework. The experiments conducted both at reduced resolution and at full resolution demonstrated the superiority of the proposed method. Furthermore, some discussions have also been presented to the readers showing how to set the parameters of our approach, its robustness with respect to the changing of these latter, an ablation study, and the robustness with respect to the selection of a number of clusters in the proposed methodology. Future developments go towards the use of fidelity terms exploring various domains (rather than just the gradient domain) to characterize the relationship between PAN and HRMS images. Besides, concise and effective regularizers can be developed to significantly enhance the capabilities of fidelity terms.

REFERENCES

[1] G. Vivone, M. Dalla Mura, A. Garzelli, R. Restaino, G. Scarpa, M. O. Ulfarsson, L. Alparone, and J. Chanussot, "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, 2020.

[2] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, 2014.

[3] L. Loncan, L. B. De Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Z. Liao, G. A. Licciardi, M. Simoes, et al., "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, 2015.

[4] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, 2016.

[5] 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.

[6] P Kwarteng, "Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 1, pp. 339–348, 1989.

[7] 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.

[8] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," Jan. 4 2000, US Patent 6,011,875.

[9] 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, 2010.

[10] 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, 2007.

[11] 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, 2000.

[12] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 07, pp. 674–693, 1989.

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

[14] 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, 1992.

[15] 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, 2017.

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

[17] Y. Yang, H. Y. Lu, S. Y. Huang, and W. Tu, "Pansharpening based on joint-guided detail extraction," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 389–401, 2020.

[18] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang, "A new pan-sharpening method with deep neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1037–1041, 2015.

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

[20] K. Zhang, W. M. Zuo, S. H. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2808–2817.

[21] J. F. Yang, X. Y. Fu, Y. W. Hu, Y. Huang, X. H. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 1753–1761.

[22] L. He, Y. Z. Rao, J. Li, J. Chanussot, A. Plaza, J. W. Zhu, and B. Li, "Pansharpening via detail injection based convolutional neural networks," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1188–1204, 2019.

[23] H. Shen, M. Jiang, J. Li, Q. Yuan, Y. Wei, and L. Zhang, "Spatial–spectral fusion by combining deep learning and variational model," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6169–6181, 2019.

[24] 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, 2021.

[25] 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.*, pp. 1–15, 2021.

[26] X. Y. Fu, W. Wang, Y. Huang, X. H. Ding, and J. Paisley, "Deep multiscale detail networks for multiband spectral image sharpening,"

IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 5, pp. 2090–2104, 2020.

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

[28] X. Y. Cao, X. Y. Fu, C. Xu, and D. Y. Meng, "Deep spatial-spectral global reasoning network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–14, 2021.

[29] Z. R. Jin, L. J. Deng, T. J. Zhang, and X. X. Jin, "BAM: Bilateral activation mechanism for image fusion," in *Proc. ACM Int. Conf. Multimedia*, 2021, pp. 4315–4323.

[30] Y. D. Wang, L. J. Deng, T. J. Zhang, and X. Wu, "SSconv: Explicit spectral-to-spatial convolution for pansharpening," in *Proc. ACM Int. Conf. Multimedia*, 2021, pp. 4472–4480.

[31] D. F. Hong, L. R. 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, 2021.

[32] D. F. Hong, Z. Han, J. Yao, L. R. Gao, B. Zhang, A. Plaza, and J. Chanussot, "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–1, 2021.

[33] D. F. Hong, L. R. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, 2020.

[34] X. Meng, H. Shen, H. Li, L.i 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, 2019.

[35] 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, 2006.

[36] L. J. Deng, G. Vivone, W. H. Guo, M. Dalla Mura, and J. Chanussot, "A variational pansharpening approach based on reproducible kernel Hilbert space and Heaviside function," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4330–4344, 2018.

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

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

[39] L. J. Deng, M. Y. 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, 2019.

[40] Y. Y. Jiang, X. H. Ding, D. L. Zeng, Y. Huang, and J. Paisley, "Pan-sharpening with a Hyper-Laplacian penalty," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 540–548.

[41] Z. C. Wu, T. Z. Huang, L. J. Deng, G. Vivone, J. Q. Miao, J. F. Hu, and X. L. Zhao, "A new variational approach based on proximal deep injection and gradient intensity similarity for spatio-spectral image fusion," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6277–6290, 2020.

[42] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[43] 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, 2018.

[44] 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.*, pp. 1–16, 2021.

[45] R. W. Dian, S. T. Li, A. J. Guo, and L. Y. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, 2018.

[46] 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, 2002.

[47] B. Aiazzi, S. Baronti, M. Selva, and L. Alparone, "Bi-cubic interpolation for shift-free pan-sharpening," *ISPRS-J. Photogramm. Remote Sens.*, vol. 86, pp. 65–76, 2013.

[48] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 930–934, 2013.

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

[50] 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. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4381–4398, 2020.

[51] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[52] 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, 2004.

[53] R. H. 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*, 1992, vol. 1, pp. 147–149.

[54] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez, "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, 2005.

[55] 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, 2007.

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

[57] 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, 2004.

[58] 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, 2008.

[59] 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, 2020.

[60] T. T. Wang, F. M. Fang, F. Li, and G. X. Zhang, "High-quality Bayesian pansharpening," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 227–239, 2018.

[61] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, 2018.