

VO+Net: An Adaptive Approach Using Variational Optimization and Deep Learning for Panchromatic Sharpening

Zhong-Cheng Wu[✉], Ting-Zhu Huang[✉], Member, IEEE, Liang-Jian Deng[✉], Member, IEEE,
Jin-Fan Hu, and Gemine Vivone[✉], Senior Member, IEEE

Abstract—Pansharpening refers to a spatio-spectral fusion of a lower spatial resolution multispectral (MS) image with a high spatial resolution panchromatic image, aiming at obtaining an image with a corresponding high resolution both in the domains. In this article, we propose a generic fusion framework that is able to weightedly combine variational optimization (VO) with deep learning (DL) for the task of pansharpening, where these crucial weights directly determining the relative contribution of DL to each pixel are estimated adaptively. This framework can benefit from both VO and DL approaches, e.g., the good modeling explanation and data generalization of a VO approach with the high accuracy of a DL technique thanks to massive data training. The proposed method can be divided into three parts: 1) for the VO modeling, a general details injection term inspired by the classical multiresolution analysis is proposed as a spatial fidelity term and a spectral fidelity employing the MS sensor’s modulation transfer functions is also incorporated; 2) for the DL injection, a weighted regularization term is designed to introduce deep learning into the variational model; and 3) the final convex optimization problem is efficiently solved by the designed alternating direction method of multipliers. Extensive experiments both at reduced and full-resolution demonstrate that the proposed method outperforms recent state-of-the-art pansharpening methods, especially showing a higher accuracy and a significant generalization ability.

Index Terms—Adaptive fusion, deep learning, image fusion, multiresolution analysis, pansharpening, remote sensing, variational models.

I. INTRODUCTION

SATELLITE systems, such as IKONOS, QuickBird, Pléiades, GeoEye, WorldView-3, and WorldView-2, capture two types of images for richer spectral and spatial

Manuscript received November 5, 2020; revised January 19, 2021; accepted February 24, 2021. This work was supported in part by NSFC under Grant 61772003, Grant 61702083, and Grant 61876203; 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.)

Zhong-Cheng Wu, Ting-Zhu Huang, Liang-Jian Deng, and Jin-Fan Hu are with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: wuzhch97@163.com; tingzhuhuang@126.com; liangjian.deng@uestc.edu.cn; 903540917@qq.com).

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

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TGRS.2021.3066425>.

Digital Object Identifier 10.1109/TGRS.2021.3066425

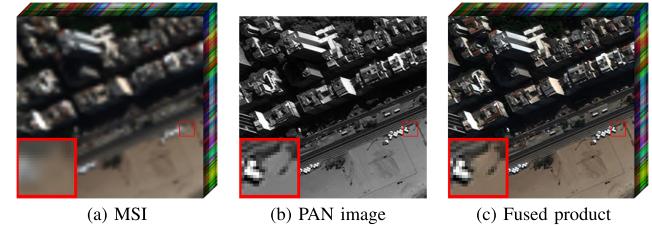


Fig. 1. Schematic illustration of pansharpening for the Rio data set at reduced resolution (source: WorldView-3). (a) MSI interpolated at PAN scale, (b) PAN image, and (c) pansharpened result obtained by our approach. The 5th, the 3rd, and the 2nd spectral channels are extracted for rendering the shown true-color images.

information. A multispectral (MS) image with a lower spatial resolution (LR-MSI) is acquired in order to retain spectral information, instead, a panchromatic (PAN) image accounts for greater spatial content. Thus, approaches for reconstructing an MSI with a higher spatial resolution (HR-MSI) driven by the simultaneously acquired PAN image are developed. In particular, the so-called pansharpening, a research line inside the class of the spatio-spectral fusion methods, is dedicated to the study of this issue. For an intuitive perspective, we present a representative example in Fig. 1.

Pansharpening plays a crucial role as a preliminary step for subsequent applications, e.g., target recognition [1] and change detection [2]. Furthermore, some commercial products, such as Bing Maps and Google Earth, exploit fused images in order to show high-resolution products. Thus, the increasing demand for pansharpened data has led to the growing number of commercial products in recent years [3] with an increment of the related academic research.

A. Related Works

Pansharpening is arousing widespread interest and a huge literature from different perspectives and methodologies can easily be found, see [4]–[6]. A general classification divides the approaches addressing the pansharpening issue into four main categories [7], [8]: 1) component substitution (CS) methods; 2) multiresolution analysis (MRA) methods; 3) variational optimization-based (VO) approaches; and 4) deep learning (DL) techniques.

The principle of CS methods is the replacement of a spatial component of the MSI, which is extracted thanks to a

spectral transformation of the LR-MSI, with the PAN image. Thus, the final product can be obtained through the relative inverse spectral transformation. Some well-known instances of methods into this class are the Brovey transform [9], the principal component analysis (PCA) [10], the partial replacement adaptive component substitution (PRACS) [11], the intensity-hue-saturation (IHS) [12], the Gram–Schmidt (GS) spectral sharpening [13], and the band dependent spatial detail (BDSD) methods [14], [15]. In general, the CS approaches are characterized by lower time-consumption, whereas they usually show severe spectral distortion [16].

Different from CS methods, the MRA approaches are based on injecting the spatial structure information extracted from the PAN image via spatial filtering into the upsampled LR-MSI. Powerful examples classified into this family are the “à-trous” wavelet transform (ATWT) [17], the smoothing filter-based intensity modulation (SFIM) [18], and generalized Laplacian Pyramid [19]. Context-based solutions have also been proposed for this class of algorithms [20], [21]. Compared with CS techniques, the pansharpened images using MRA methods mainly suffer from spatial distortion, whereas spectral contents are preserved well.

The VO approaches in pansharpening have been successfully attracted more attention for their capability of settling usual ill-posedness inverse problems [22]–[26]. These latter are committed to describing an exact link among the PAN image, LR-MSI, and the unknown HR-MSI, thus formulating an energy function. Therefore, the unknown (pansharpened) image will be get via minimizing this known function. In 2006, Ballester *et al.* [27] proposed a pioneering pansharpening method, named P+XS, based on variational regularization, mainly relying upon the assumption that the PAN image can be represented as a linear degradation of the target HR-MSI along the spectral dimension. However, the assumption is unrealistic [28] so that this method sometimes suffers from spectral distortion. To strongly avoid these drawbacks, Fu *et al.* [29] take the gradient similarity constraint between the underlying HR-MSI and the PAN image into account from the local rather than global perspective, enforcing a structural similarity and ensuring better fitting of the relation between the sensors’ responses. Besides, pansharpening methods based on unmixing [30] and dictionary learning [31] are also turned out to be effective techniques. In general, better results, both spatially and spectrally, can be theoretically produced via variational methodology with a solid mathematical foundation [8]. Unfortunately, these methods usually produce many unpredictable deviations once making some unreasonable assumptions, as in [27]. In addition, the misalignment among channels is also considered as another issue, usually resulting in artifacts on final products.

Recently, the DL class has rapidly developed for computer vision problems. Many convolutional neural networks (CNNs) based approaches have been designed for image fusion showing excellent capabilities for feature extraction and nonlinear mapping learning [8], [32]–[38]. They can perfectly compensate for the aforementioned deficiencies reflected by VO methods getting state-of-the-art performance [39]. Nevertheless, CNN-based methods are often over-dependent on the training

data [29], thus leading to a weakened generalization, i.e., these methods have excellent performance only on data similar to the ones exploited in the training set. Specifically, since the network parameters are fixed once the training is finished, the accuracy of CNN-based methods cannot be further improved [29]. Although some exceptions, see e.g. [40], where a self-supervised parameter tuning phase has been incorporated in the inference engine, have recently been implemented, the vast majority of the remaining DL-based methods still present this problem.

Moreover, hybrid approaches sharing some of the concepts of CS, MRA, VO, and DL have also been proposed, e.g., the guided filter PCA (GFPCA) [41], the additive wavelet luminance proportional (AWLP) approaches, see [42], where the first two concepts are integrated. Furthermore, Shen *et al.* [8] designed a groundbreaking architecture for remote sensing image fusion problem via combining a VO method and DL, in which HR-MSI’s gradient priors learned through a residual CNN are integrated into their image observation variational models, getting state-of-the-art results compared to traditional methods. Given the above considerations, we explored in this article, a hybrid fusion framework (detailed in Section III), which can leverage the favorable merits of both MRA and VO and DL methodologies, thus expecting to achieve higher performance on various data. Different from the work [8], we develop in this article a new strategy, which can adaptively distinguish the relative contribution of deep networks to each pixel, i.e., the weight of characteristics closer to ground-truth (GT) image will be elevated via the proposed estimator. These kinds of algorithms cannot be characterized by the behavior of any single class of the above rigid taxonomy. For instance, our scheme can simultaneously be classified as in the MRA, the VO, and the DL categories.

B. Contributions

In this article, we propose a novel “VO+Net” scheme to merge the LR-MSIs and PAN images, which can weighted combine a generic variational model and a DL technique, thus achieving higher accuracy and generalization. In particular, the designed model consists of the spectral and spatial data fitting terms and a CNN-based proximal term with pixel-wise weights that are automatically estimated (we will call this term WPDI from hereon). The spectral fidelity using the ℓ_2 norm is imposed on the LR-MSI and motivated by analyzing the degradation from an underlying HR-MSI to the LR-MSI. To capture spatial details in a proper way, the spatial counterpart is imposed by exploiting a details injection framework encouraging sharp edges of the unknown HR-MSI. Moreover, we introduce the prior knowledge contained in the CNNs into the proposed variational model through the WPDI, which is considered a regularizer. This new hybrid problem is solved by designing an alternating direction method of multipliers (ADMM)-based algorithm, which is guaranteed to efficiently converge to the global optimum. Extensive experiments both at reduced and full resolution, exploiting real data acquired by WorldView-3, WorldView-2, and QuickBird sensors, verify the

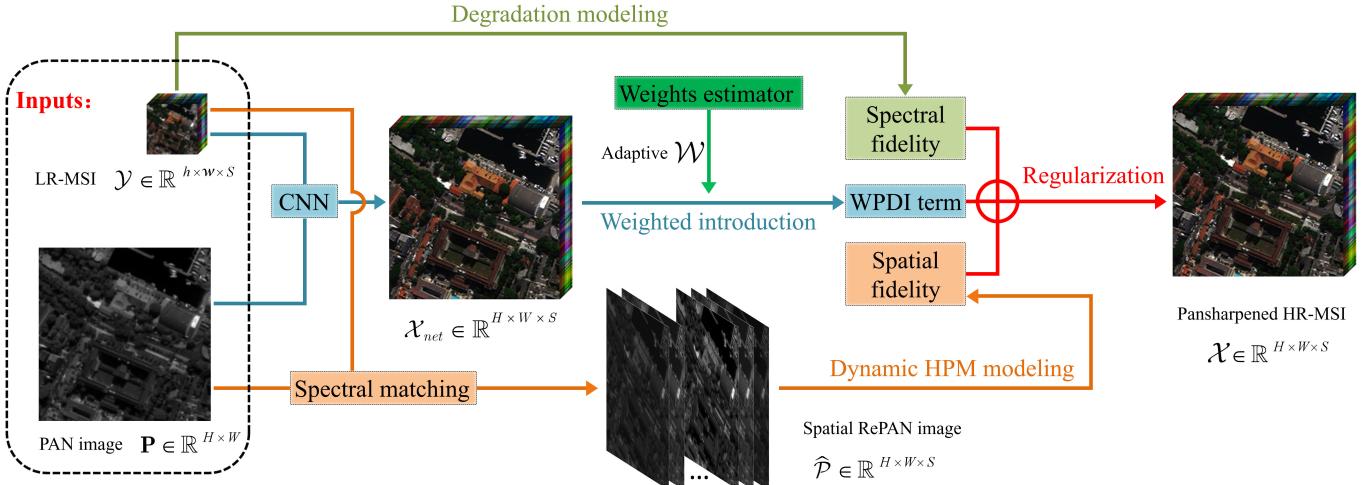


Fig. 2. Flowchart of our scheme. The spectral matching [43] denotes the pre-processing operation, see Section II-A. The acronym CNN stands for deep convolutional neural network.

superiority of the proposed framework compared to other state-of-the-art methods belonging to different categories. Finally, discussions about the parameters analysis, the ablation study, the effects of the new weighting term, the generalization ability, and the sensitivity analysis of computational load, are also provided to the readers. The flowchart of the proposed scheme is provided in Fig. 2.

The *contributions* of this article are summed up as follows.

- 1) A details injection framework inspired by the MRA methodology has been introduced in our scheme as a spatial fidelity term;
- 2) A proximal injection term with adaptive gains in order to combine the prior information from various CNNs with our variational model has been designed resulting in an approach with a better generalization ability and a higher accuracy than the two approaches taken alone;
- 3) An ADMM-based algorithm has been designed and developed to solve the formulated hybrid problem;
- 4) A huge experimental analysis based on real data has been conducted to assess the performance of the proposed framework against the benchmark consisting of state-of-the-art pansharpening approaches.

C. Organization

The remaining of this article is organized as follows. The notation and motivations are briefly presented in Section II. In Section III, the proposed regularization-based framework is detailed. Afterward, the algorithm we designed is provided in Section IV. In Section V, the experimental analysis is shown comparing the proposed approach with some state-of-the-art techniques belonging to different classes of pansharpening algorithms. Finally, conclusions are drawn in Section VI.

II. NOTATION AND MOTIVATIONS

A. Related Notation

Lowercase letters denote scalars, matrices and vectors are denoted by uppercase and lowercase bold letters, respectively, and calligraphic letters represent tensor. We denote by $\mathbf{1}$ the

all-one matrix, whose dimension, when not given explicitly, shall be inferred from the context. Furthermore, the main acronyms and symbols used in this article are reported below for $i = 1, 2, \dots, S$.

- 1) HR-MSI: $\mathcal{X} \in \mathbb{R}^{H \times W \times S}$ with S spectral images; $\mathbf{X}^{(i)} \in \mathbb{R}^{H \times W}$, where H and W are the number of pixels in the row and column dimensions, respectively.
- 2) LR-MSI: $\mathcal{Y} \in \mathbb{R}^{h \times w \times S}$ with S spectral images; $\mathbf{Y}^{(i)} \in \mathbb{R}^{h \times w}$, where $H = h \times r$, $W = w \times r$, and r stands for the scale factor.
- 3) PAN: the known single channel image $\mathbf{P} \in \mathbb{R}^{H \times W}$.
- 4) Reference panchromatic image (RePAN): the extended PAN image $\hat{\mathcal{P}} \in \mathbb{R}^{H \times W \times S}$ with S bands $\hat{\mathbf{P}}^{(i)} \in \mathbb{R}^{H \times W}$, used as reference for the proposed spatial fidelity term in Section III-B.

$\hat{\mathbf{P}}^{(i)}$ can be regarded as a function of LR-MSI bands and the PAN image and it is generated through some pre-processing operations, e.g., spectral matching [43]. Their mode-3 unfolding (refer to [44] for more details) can be marked by the sibling uppercase bold letters, namely, by $\mathbf{X} = \text{Unfold}(\mathcal{X})$, $\mathbf{Y} = \text{Unfold}(\mathcal{Y})$ and $\hat{\mathbf{P}} = \text{Unfold}(\hat{\mathcal{P}})$, respectively. Given $\mathbf{l} = (l_1, l_2) \in \Omega$, where Ω is the related image domain, $\mathbf{X}^{(i)}[\mathbf{l}]$ denotes the pixel value of $\mathbf{X}^{(i)}$ at location (l_1, l_2) . The other symbols will be defined along this article, where needed.

B. Motivations

Thanks to the remarkable nonlinear mapping ability of DL, CNN-based methods have recently attracted attention, yielding competitive results even for image fusion. Nevertheless, these methods have also been revealed some fatal defects, as analyzed in Section I, e.g., the enormous data dependence, owing to the use of network parameters that can hardly be adjusted on data very different from the ones shown in the training phase. As a result, an optimization strategy that can make up for these disadvantages needs to be developed to further improving the performance. For this reason, we propose the concept of WPDI, aiming to introduce the output of a CNN-based method into a VO framework. The WPDI term establishes a relation between this output and the target HR-MSI, thus transferring

the required prior information from CNN into a variational fusion framework for further optimization. In this model, the WPDI plays the generic role of a regularizer avoiding using other additional prior information often used in the related literature, see, e.g., low-rank prior [45] and sparse prior [46].

Furthermore, a variational model that includes the spectral and spatial data in input fitting terms needs to be completed. To the best of our knowledge, the spectral fidelity is designed almost in the same way for most VO techniques, even if rare innovations can be found in the related literature, e.g., [5]. Instead, a crucial choice regards the spatial fidelity data fitting term accounting for the PAN image's spatial content, considering the clear difficulty in modeling the underlying nonlinear even indirect relation. Hence, a new scheme to boost the spatial structure is developed based on the multi-resolution analysis philosophy characterized by a superior spectral fidelity.

III. PROPOSED MODEL

This section is devoted to presenting the proposed model and the related spatial, spectral, and WPDI terms.

A. Proposed VO+Net Fusion Framework

In this work, we propose a general fusion framework linking the traditional VO and the burgeoning DL techniques for pansharpening, aiming to benefit from both the VO and DL approaches, e.g., the good modeling explanation and data generalization of VO approaches as well as the high accuracy of DL techniques. In particular, this uniformed framework is built up thanks to an adaptive tensor-format weight \mathcal{W} . The optimization model is as follows:

$$\min_{\mathcal{X}} f_{\text{vo}}(\mathcal{X}, \mathcal{Y}, \mathbf{P}) + \alpha f_{\text{WPDI}}(\mathcal{X}, \mathcal{X}_{\text{net}}, \mathcal{W}) \quad (1)$$

where $f_{\text{vo}}(\mathcal{X}, \mathcal{Y}, \mathbf{P})$ is the VO model involving the underlying image \mathcal{X} and the observed images \mathcal{Y} and \mathbf{P} ; $f_{\text{WPDI}}(\mathcal{X}, \mathcal{X}_{\text{net}}, \mathcal{W})$, i.e., the WPDI term, plays the role of a prior term, α is a positive parameter to balance the VO and the WPDI terms, $\mathcal{X}_{\text{net}} \in \mathbb{R}^{H \times W \times S}$ indicates the output of a generic CNN-based method and $\mathcal{W} \in \mathbb{R}^{H \times W \times S}$ is the tensor-format weight that is viewed as a bridge connecting VO and DL in element-wise manner.

More specifically, $f_{\text{vo}}(\mathcal{X}, \mathcal{Y}, \mathbf{P})$ in (1) can be substituted by two popular data fitting terms in pansharpening, i.e., $f_{\text{spec}}(\mathcal{X}, \mathcal{Y})$ (the spectral fidelity term) and $f_{\text{spat}}(\mathcal{X}, \mathbf{P})$ (the spatial fidelity term). Afterward, the adopted pansharpening model is given by

$$\min_{\mathcal{X}} f_{\text{spec}}(\mathcal{X}, \mathcal{Y}) + \lambda f_{\text{spat}}(\mathcal{X}, \mathbf{P}) + \alpha f_{\text{WPDI}}(\mathcal{X}, \mathcal{X}_{\text{net}}, \mathcal{W}) \quad (2)$$

where λ is a positive regularization parameter.

For simplicity, the problem in (2) can be simply rewritten as

$$\min_{\mathbf{X}} f_{\text{spec}}(\mathbf{X}, \mathbf{Y}) + \lambda f_{\text{spat}}(\mathbf{X}, \mathbf{P}) + \alpha f_{\text{WPDI}}(\mathbf{X}, \mathbf{X}_{\text{net}}, \mathbf{W}) \quad (3)$$

where $\mathbf{X}, \mathbf{X}_{\text{net}}, \mathbf{W} \in \mathbb{R}^{S \times H \times W}$ and $\mathbf{Y} \in \mathbb{R}^{S \times h \times w}$ are the $\text{Unfold}(\mathcal{X})$, the $\text{Unfold}(\mathcal{X}_{\text{net}})$, the $\text{Unfold}(\mathcal{W})$ and the $\text{Unfold}(\mathcal{Y})$, respectively.

B. Spectral Fidelity Term

Many existing spatio-spectral fusion methods, see e.g. [9], [11], [14], upsample the LR-MSI to the PAN scale. Nevertheless, imprecise information could be carried using this (usually biased) interpolation procedure, thus impacting on the performance. Accordingly, the downsampled version of the underlying HR-MSI is considered in this article. We perform this operation according to the point spread function (PSF) of the MSI sensor [47], [48], aiming to design the blurring operation. Thus, we model that

$$\mathbf{Y} = \mathbf{XBS} + \xi_1 \quad (4)$$

where $\mathbf{B} \in \mathbb{R}^{HW \times HW}$ is the matrixization of blurring operators, $\mathbf{S} \in \mathbb{R}^{HW \times hw}$ denotes the decimation matrix consisting of sparse components, and ξ_1 indicates a zero-mean Gaussian noise assumed to follow a normal distribution $\mathcal{N}(0, \sigma_1^2)$ with a standard deviation σ_1 . From (4), we paradigmize a spectral fidelity term as

$$f_{\text{spec}}(\mathbf{X}, \mathbf{Y}) = \|\mathbf{XBS} - \mathbf{Y}\|_F^2 \quad (5)$$

where $\|\cdot\|_F$ is the Frobenius norm. In some previous articles, researchers used a uniform or a fixed Gaussian kernel to define \mathbf{B} without taking into account of the possible differences along the spectral domain. Indeed, kernels designed to match the MS sensor's modulation transfer functions (MTFs) are advisable [16], [49]. These usually assume a Gaussian shape with band-dependent standard deviations that are set thanks to the above-mentioned information [16], [49]. Thus, these values are also sensor-dependent invalidating the use of simple solutions based on uniform or fixed Gaussian kernels. In this work, we define \mathbf{B} via the above-mentioned prior information based on the MS sensor's MTFs for a more accurate implementation of the deblurring operation. See [47], [48], [50]–[52] for more information about the matrices \mathbf{B} and \mathbf{S} .

C. Proposed Spatial Fidelity Term

Many spatial fidelity models, see [53], prefer choosing intensity similarity or structure similarity as a constraint, forcing pixels or gradient values between the unknown HR-MSI and the PAN image to be consistent. However, the spectral accuracy of the results is often affected and tends to get worse with the increase of the channels to be fused (see e.g. approaches into the CS class). Considering the analysis drawn in Section I, an alternative technique for obtaining the critical details is constituted by the *multiresolution analysis* (MRA), which is generally achieved by linear decomposition procedures, such as wavelet [54] and Laplacian pyramids [55]. These often yield a high fidelity and consistency in rendering the spatial and spectral features in the final image retaining a good tradeoff, especially when the characteristics of the MS and PAN acquisition systems are considered [16], [56]. Besides, their main advantages also consist in temporal coherence [57] and robustness to aliasing under proper conditions [58]. Inspired by the idea under MRA approaches, we exploit the MRA framework presented in [16] and [57] defining a new spatial improved term based on the extraction first and then injection of spatial details.

The general MRA framework [16], [57] can be formulated as

$$\mathbf{X}^{(i)} = \mathbf{Y}_H^{(i)} + \mathbf{G}^{(i)} \circ \left(\widehat{\mathbf{P}}^{(i)} - \widehat{\mathbf{P}}_L^{(i)} \right) \quad (6)$$

for $i = 1, 2, \dots, S$, where $\mathbf{Y}_H^{(i)}, \mathbf{G}^{(i)} \in \mathbb{R}^{H \times W}$ denote the upsampled version of $\mathbf{Y}^{(i)}$ at the scale of the $\widehat{\mathbf{P}}^{(i)}$ and the injection coefficients, respectively, the symbol \circ represents the element-wise multiplication, and $\widehat{\mathbf{P}}_L^{(i)}$ is a low-pass version of $\widehat{\mathbf{P}}^{(i)}$ via filtering with impulse response h_i , i.e., $\widehat{\mathbf{P}}_L^{(i)} = \widehat{\mathbf{P}}^{(i)} * h_i$ [59], where $*$ indicates the convolution operation. Moreover, $\widehat{\mathbf{P}}^{(i)} - \widehat{\mathbf{P}}_L^{(i)}$ are the image details regarded as a high-pass version of $\widehat{\mathbf{P}}^{(i)}$. Two common options for defining the injection gains are $\mathbf{G}^{(i)} = \mathbf{1}$, namely the so-called *additive* injection scheme, and $\mathbf{G}^{(i)} = \mathbf{Y}_H^{(i)} \oslash \widehat{\mathbf{P}}_L^{(i)}$, which is referred to as the *multiplicative* injection scheme or *high-pass modulation* (HPM) scheme, where \oslash denotes an element-wise division. The latter, adopted in this article, can reproduce the local intensity contrast of the PAN image in the fused image [60] and turns out to be better suited than the former, being explained by its greater flexibility in configuring the local weights [16].

Analogously to Section III-A, we make the substitution of \mathbf{Y}_H with \mathbf{XB} , to avoid the introduction of the LR-MSI, \mathbf{Y} , thus independently modeling the relation from the unknown HR-MSI to the PAN image. Hence, considering the matrix form, we have

$$\mathbf{G} \circ \left(\widehat{\mathbf{P}} - \widehat{\mathbf{P}}_L \right) = \mathbf{X} - \mathbf{XB} + \xi_2 \quad (7)$$

where ξ_2 is an error assumed to follow a zero-mean normal distribution $\mathcal{N}(0, \sigma_2^2)$ with a standard deviation σ_2 . This is an implicit expression about the histogram-matched PAN image, contributing to a componentwise mapping from an unknown HR-MSI to the details image closely related to the PAN image. It is worth to be remarked that by comparing with the structure-similarity models employing the hypothesis of linear combination, see [3], the spatial fidelity model in (7) can improve spatial details while preserving the spectral content to the greatest extent, benefiting from the dual characteristics of the MRA and of the element by element mapping in the high-frequency domain (using HPM injection models).

Accordingly, an MRA-based details injection model for a spatial fidelity can be obtained as follows:

$$f_{\text{spta}}(\mathbf{X}, \mathbf{P}) = \left\| \mathbf{X} - \left(\mathbf{XB} + \mathbf{G} \circ \left(\widehat{\mathbf{P}} - \widehat{\mathbf{P}}_L \right) \right) \right\|_F^2. \quad (8)$$

In this article, we adopt a HPM scheme to deal with the details extraction. Namely, we define $\mathbf{G} = \mathbf{XB} \oslash \widehat{\mathbf{P}}_L$ (directly using \mathbf{XB} instead of \mathbf{Y}_H), thus having

$$\mathbf{XB} + \mathbf{G} \circ \left(\widehat{\mathbf{P}} - \widehat{\mathbf{P}}_L \right) = \mathbf{XB} \circ \left(\widehat{\mathbf{P}} \oslash \widehat{\mathbf{P}}_L \right) \quad (9)$$

where $\widehat{\mathbf{P}}_L$ is defined exploiting MTF-based filters as in Section III-A.

Accordingly, the final spatial fidelity term relying upon an HPM scheme can be formalized as follows:

$$f_{\text{spta}}(\mathbf{X}, \mathbf{P}) = \left\| \mathbf{X} - \mathbf{XB} \circ \left(\widehat{\mathbf{P}} \oslash \widehat{\mathbf{P}}_L \right) \right\|_F^2. \quad (10)$$

The advantages of the formulation in (10) are an enhancement of the spatial details while preserving spectral content and the absence of any interpolation operation, unlike in (6).

D. WPDI Term

In [61] the following proximal deep injection (PDI) term modulated by a constant parameter α as a regularizer is introduced

$$f_{\text{PDI}}(\mathbf{X}, \mathbf{X}_{\text{net}}) = \|\mathbf{X} - \mathbf{X}_{\text{net}}\|_F^2. \quad (11)$$

By using this proximal term, it is interesting to note in [61] that the proposed technique gets always better performance than the CNN-based outcome \mathbf{X}_{net} , through controlling the parameter α . If the input data significantly differs from the training data of the CNN-based method, one could select a smaller α , even zero, to weaken the PDI term. However, from the element-wise perspective, we argue that, even in the worst case where the CNN result, \mathbf{X}_{net} , seems so far from the optimality, some pixel values in \mathbf{X}_{net} are very close to that of the GT image. This means that it is better to have a spatial-dependent weight, aiming to locally differentiate the relative contributions of \mathbf{X}_{net} . Thus, we propose a pixel-based weight to minimize the distance between the unknown HR-MSI and the CNN-based outcome \mathbf{X}_{net} , named weighted proximal deep injection (WPDI) term. Besides, a novel strategy is designed to adaptively estimate this weight. The proximal term is defined as follows:

$$f_{\text{WPDI}}(\mathbf{X}, \mathbf{X}_{\text{net}}) = \|\mathbf{W} \circ (\mathbf{X} - \mathbf{X}_{\text{net}})\|_F^2 \quad (12)$$

where $\mathbf{W} \in \mathbb{R}^{S \times HW}$ denotes a coefficient matrix defining a weight for each pixel. A better result can be expected through a set of weighting factors that are inversely proportional to the absolute error between GT image and \mathbf{X}_{net} . Thus, still considering the parameter α , we propose a new weighting procedure (absorbing α into the paradigm to simplify the mathematical formula) as follows:

$$\mathbf{W}_\alpha = \sqrt{\alpha} \mathbf{W} = \sqrt{\alpha} w f(\Delta) \quad (13)$$

where α is a balance parameter, $w f$ is strictly defined by an element-wise monotone decreasing function, and Δ denotes the absolute residual matrix. In Section V, we define $w f$ as

$$w f(\Delta) \stackrel{\text{def}}{=} \sqrt{\mathbf{1} - \Delta} \quad (14)$$

in which $\mathbf{1} - \Delta$ is selected as a correction to $\mathbf{W} = \mathbf{1}$ (i.e., Unweighted), and $\sqrt{(\cdot)}$ plays a role to offset the quadratic effect, as in sub-solution (34) of Section IV. Note that, outliers exceeding 1 in Δ are forced to be 1. This definition suggests that the undetermined \mathbf{W} has a simple dependence on the reconstructed accuracy of the network or, more precisely, on the dispersion of the reconstructed residuals. The degree of dispersion is measured by the standard deviation of Δ , i.e., $\text{STD}(\Delta)$, abbreviated as STDR from hereon. The procedure in (13) is committed to distinguishing the effectiveness of each element, thus referring to as *differential weighting based on the dispersion of residuals* (DWDR). The higher the STDR, the more significant the effect of the DWDR. Section V-E experimentally demonstrates the validity of the

proposed weighting scheme. However, the reference GT image is unavailable, which leads to the next estimation step based on (7).

E. Estimating Δ

Combining (7) with (9), we have that

$$\begin{aligned}\mathbf{X} &= \mathbf{G} \circ (\widehat{\mathbf{P}} - \widehat{\mathbf{P}}_L) + \mathbf{X}\mathbf{B} - \xi_2 \\ &= \mathbf{X}\mathbf{B} \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L) - \xi_2'.\end{aligned}\quad (15)$$

Thus, we can redefine the expression above using \mathbf{X}_{net} and the GT image, say it \mathbf{X}_{gt} , as follows:

$$\begin{aligned}\mathbf{X}_{\text{net}} &= \mathbf{X}_{\text{net}}\mathbf{B} \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L) - \xi_2' \\ \mathbf{X}_{gt} &= \mathbf{X}_{gt}\mathbf{B} \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L) - \xi_2''\end{aligned}\quad (16)$$

where ξ_2' and ξ_2'' are the two reconstruction errors assumed to be equivalent when inducing. Then, the absolute residual matrix can be estimated as

$$\begin{aligned}\Delta &= \Gamma_+(\mathbf{X}_{\text{net}} - \mathbf{X}_{gt}) \\ &= \Gamma_+((\mathbf{X}_{\text{net}}\mathbf{B} - \mathbf{X}_{gt}\mathbf{B}) \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L) - (\xi_2' - \xi_2'')) \\ &\approx \Gamma_+((\mathbf{X}_{\text{net}}\mathbf{B} - \mathbf{Y}_H) \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L))\end{aligned}\quad (17)$$

where Γ_+ indicates the componentwise projection operator on the set of nonnegative real numbers, that is,

$$\Gamma_+(\mathbf{X})[I] \stackrel{\text{def}}{=} \begin{cases} -\mathbf{X}[I], & \text{If } \mathbf{X}[I] < 0 \\ \mathbf{X}[I], & \text{If } \mathbf{X}[I] \geq 0. \end{cases}\quad (18)$$

Accordingly, the final optimization model consisting of three energy terms is expressed as

$$\begin{aligned}\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{X}\mathbf{B}\mathbf{S} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{X} - \mathbf{X}\mathbf{B} \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L)\|_F^2 \\ + \|\mathbf{W}_\alpha \circ (\mathbf{X} - \mathbf{X}_{\text{net}})\|_F^2.\end{aligned}\quad (19)$$

The proposed scheme explores the obtainable spatial details and spectral contents via the first two fitting terms for data in the input. Furthermore, with the WPDI term treated as a regularizer, the prior knowledge provided by any CNN is incorporated into the proposed VO model to optimize further.

Although the objective function (19) is convex and differentiable, it is inadvisable to directly generate the derivative because of a heavy computational and storage burden. In the next section, a new ADMM-based algorithm will be explored to pursue an iterative solution.

IV. PROPOSED ALGORITHM

This section applies an ADMM methodology [62] designed for structured convex optimization problems to the proposed model, aiming to separate the problem in three independent subproblems, all having a closed-form solution. By introducing two auxiliary variables \mathbf{U} and \mathbf{V} , we rewrite (19) as the following equivalent constrained problem:

$$\begin{aligned}\min_{\mathbf{X}, \mathbf{U}, \mathbf{V}} \frac{1}{2} \|\mathbf{U}\mathbf{S} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{X} - \mathbf{U} \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L)\|_F^2 \\ + \|\mathbf{W}_\alpha \circ \mathbf{V} - \mathbf{W}_\alpha \circ \mathbf{X}_{\text{net}}\|_F^2 \\ \text{s.t. } \mathbf{U} = \mathbf{X}\mathbf{B}, \quad \mathbf{V} = \mathbf{X}.\end{aligned}\quad (20)$$

The augmented Lagrangian function of this constrained minimization problem (20) can be expressed as

$$\begin{aligned}\mathcal{L}_{\eta_1, \eta_2}(\mathbf{X}, \mathbf{U}, \mathbf{V}, \Lambda_1, \Lambda_2) \\ = \frac{1}{2} \|\mathbf{U}\mathbf{S} - \mathbf{Y}\|_F^2 \\ + \lambda \left\| \mathbf{X} - \mathbf{U} \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L) \right\|_F^2 + \|\mathbf{W}_\alpha \circ \mathbf{V} - \mathbf{W}_\alpha \circ \mathbf{X}_{\text{net}}\|_F^2 \\ + \frac{\eta_1}{2} \left\| \mathbf{X} - \mathbf{U} + \frac{\Lambda_1}{\eta_1} \right\|_F^2 + \frac{\eta_2}{2} \left\| \mathbf{X} - \mathbf{V} + \frac{\Lambda_2}{\eta_2} \right\|_F^2 + \text{const}\end{aligned}\quad (21)$$

where Λ_1, Λ_2 denote the Lagrange multipliers, $\eta_1, \eta_2 > 0$ are two penalty parameters, and const represents a generic constant. Afterward, (21) can be alternatively and iteratively solved via updating the following simpler subproblems.

A. Updating \mathbf{X}

By fixing $\mathbf{U}, \mathbf{V}, \Lambda_1, \Lambda_2$, then, the \mathbf{X} -subproblem is converted to

$$\begin{aligned}\arg \min_{\mathbf{X}} \lambda \left\| \mathbf{X} - \mathbf{U} \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L) \right\|_F^2 + \frac{\eta_1}{2} \left\| \mathbf{X}\mathbf{B} - \mathbf{U} + \frac{\Lambda_1}{\eta_1} \right\|_F^2 \\ + \frac{\eta_2}{2} \left\| \mathbf{X} - \mathbf{V} + \frac{\Lambda_2}{\eta_2} \right\|_F^2\end{aligned}\quad (22)$$

which is a simple least squares problem and has a closed-form solution under the condition of a periodic boundary, as

$$\mathbf{X} := \mathcal{F}^{-1} \left(\frac{\mathbf{M} + \mathbf{N}}{(2\lambda + \eta_2)\mathbf{1} + \eta_1 \mathcal{F}(\mathbf{B}) \circ \mathcal{F}(\mathbf{B})^\dagger} \right)\quad (23)$$

with

$$\mathbf{M} = 2\lambda \mathcal{F}(\mathbf{U} \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L)) + \eta_1 \mathcal{F}(\mathbf{U}) \circ \mathcal{F}(\mathbf{B})^\dagger\quad (24)$$

$$\mathbf{N} = -\mathcal{F}(\Lambda_1) \circ \mathcal{F}(\mathbf{B})^\dagger + \eta_2 \mathcal{F}(\mathbf{V}) - \mathcal{F}(\Lambda_2)\quad (25)$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ are the fast Fourier transform (FFT) and its inverse operators, respectively, the superscript \dagger denotes the complex conjugate, and the division is componentwise, as well. Exploiting MTF-based filters (generating different blurring kernels for each spectral band), this solution can be rearranged as a single channel-wise expression, appearing as

$$\mathbf{X}^{(i)} := \mathcal{F}^{-1} \left(\frac{\mathbf{M}^{(i)} + \mathbf{N}^{(i)}}{(2\lambda + \eta_2)\mathbf{1} + \eta_1 \mathbf{O}^{(i)} \circ (\mathbf{O}^{(i)})^\dagger} \right)\quad (26)$$

for $i = 1, 2, \dots, S$, in which

$$\mathbf{M}^{(i)} = 2\lambda \mathcal{F}(\mathbf{U}^{(i)} \circ (\widehat{\mathbf{P}}^{(i)} \otimes \widehat{\mathbf{P}}_L^{(i)})) + \eta_1 \mathcal{F}(\mathbf{U}^{(i)}) \circ (\mathbf{O}^{(i)})^\dagger\quad (27)$$

$$\mathbf{N}^{(i)} = -\mathcal{F}(\Lambda_1^{(i)}) \circ (\mathbf{O}^{(i)})^\dagger + \eta_2 \mathcal{F}(\mathbf{V}^{(i)}) - \mathcal{F}(\Lambda_2^{(i)})\quad (28)$$

where $\mathbf{O}^{(i)} \in \mathbb{R}^{H \times W}$ denotes the optical transfer function (OTF) obtained by applying the FFT to $h_i \in \mathbb{R}^{r \times r}$ matching the MS sensor's MTF at the i th spectral channel.

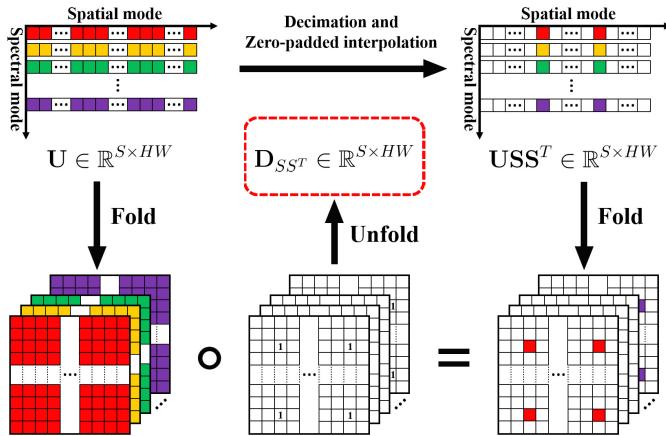


Fig. 3. Graphical illustration of (30) when $r = 4$. White squares with a solid line and a blank content are zero values.

B. Updating \mathbf{U}

Similarly, the variable \mathbf{U} can be updated by solving the following function:

$$\begin{aligned} \arg \min_{\mathbf{U}} \frac{1}{2} \|\mathbf{U}\mathbf{S} - \mathbf{Y}\|_F^2 + \lambda \left\| \mathbf{X} - \mathbf{U} \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L) \right\|_F^2 \\ + \frac{\eta_1}{2} \left\| \mathbf{X}\mathbf{B} - \mathbf{U} + \frac{\Lambda_1}{\eta_1} \right\|_F^2. \end{aligned} \quad (29)$$

Before going into the details, we provide to the readers a definition of a novel concept, saying it *decimation mask*.

Definition 4.1 (Decimation Mask): Given a scale factor r , the decimation mask is defined as a sparse matrix $\mathbf{K} \in \mathbb{R}^{r \times r}$, whose entry is 1 only in one position determined by the decimation scheme, e.g., $\mathbf{K}[3, 3] = 1$ when $r = 4$ (as in our article) whereas $\mathbf{K}[i] = 0$ for all the other cases.

Afterward, the following equation holds, also interpreted as in Fig. 3

$$\mathbf{U}\mathbf{S}^T = \mathbf{U} \circ \mathbf{D}_{SS^T} \quad (30)$$

with

$$\mathbf{D}_{SS^T} = [\text{vec}(\mathbf{K} \otimes \mathbf{1}), \text{vec}(\mathbf{K} \otimes \mathbf{1}), \dots, \text{vec}(\mathbf{K} \otimes \mathbf{1})]^T \quad (31)$$

where the symbol \otimes denotes the Kronecker product, $\text{vec}(\cdot)$ implies the vectorization operator, and $\mathbf{1} \in \mathbb{R}^{h \times w}$. Based on (30) and (31), the solution of the least squares problem (29) can be denoted as

$$\mathbf{U} := \frac{\mathbf{Y}\mathbf{S}^T + \eta_1 \mathbf{X}\mathbf{B} + 2\lambda \mathbf{X} \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L) + \Lambda_1}{\mathbf{D}_{SS^T} + \eta_1 \mathbf{1} + 2\lambda (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L) \circ (\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L)}. \quad (32)$$

Note that the element-wise division is needed, as well.

C. Updating \mathbf{V}

Different from the above-mentioned procedures, the \mathbf{V} -subproblem is simpler and has a straightforward solution according to

$$\arg \min_{\mathbf{V}} \|\mathbf{W}_\alpha \circ \mathbf{V} - \mathbf{W}_\alpha \circ \mathbf{X}_{\text{net}}\|_F^2 + \frac{\eta_2}{2} \left\| \mathbf{X} - \mathbf{V} + \frac{\Lambda_2}{\eta_2} \right\|_F^2 \quad (33)$$

Algorithm 1 ADMM-Based Solver for the Proposed Pan-sharpening Model (19)

Input: LR-MSI \mathbf{Y} , PAN image \mathbf{P} , \mathbf{X}_{net} , λ , α , η_1 , η_2 , r , k_{mit} .
Initialization: $\mathbf{X}^0 = \Psi(\mathbf{Y}, r)$, $\mathbf{U}^0 = \mathbf{V}^0 = \Lambda_1^0 = \Lambda_2^0 = \mathbf{0}$

1: **while** $k < k_{\text{mit}}$ and $\text{relcha} > \varepsilon$ **do**
2: Update \mathbf{X} via (26) - (28).
3: Update \mathbf{U} via (31) - (32).
4: Update \mathbf{V} via (34).
5: Update Lagrange multiplier Λ_1 via (35).
6: Update Lagrange multiplier Λ_2 via (36).
7: **end while**
Output: Fused HR-MSI \mathbf{X}

correspondingly, the closed-form solution can be directly given by

$$\mathbf{V} := \frac{2\mathbf{W}_\alpha \circ \mathbf{W}_\alpha \circ \mathbf{X}_{\text{net}} + \eta_2 \mathbf{X} + \Lambda_2}{2\mathbf{W}_\alpha \circ \mathbf{W}_\alpha + \eta_2 \mathbf{1}}. \quad (34)$$

Likewise, the division is element-wise, as well.

D. Updating Λ_1 , Λ_2

According to the ADMM framework, multipliers Λ_1 , Λ_2 can be updated via

$$\Lambda_1 := \Lambda_1 + \eta_1 (\mathbf{X}\mathbf{B} - \mathbf{U}) \quad (35)$$

and

$$\Lambda_2 := \Lambda_2 + \eta_2 (\mathbf{X} - \mathbf{V}). \quad (36)$$

The algorithm considers the termination criterion as that the relative change (relcha) between two successive pansharpened results less than a tolerance value, ε , that is,

$$\text{relcha} = \|\mathbf{X}^{k+1} - \mathbf{X}^k\|_F / \|\mathbf{X}^k\|_F < \varepsilon. \quad (37)$$

The detailed algorithm solving the problem (19) is summarized in Algorithm 1, in which, k_{mit} is the maximum iteration of execution, and Ψ indicates the bicubic interpolation for accelerating convergence. The convergence of the designed iterative scheme is guaranteed [63].

V. EXPERIMENTAL RESULTS

This section is developed to compare the proposed scheme with some state-of-the-art approaches using several data sets collected by different sensors. Two types of validation, i.e., qualitative and quantitative, are conducted at reduced and full resolution, respectively. Reduced resolution data are obtained according to Wald's protocol [60] under the scale invariance assumption, i.e., filtering with filters designed to match the MS sensor's MTFs and decimation [16], [49]. All the experiments are implemented in MATLAB on a computer of 16Gb RAM and Intel(R) Core(TM) i7-5960X CPU at 3.00 GHz.

For quantitative evaluation at reduced resolution, some popular indexes exploiting a reference image are adopted, including the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [64], which are widely used to evaluate the similarity between two images in the image processing

literature; the spectral angle mapper (SAM) index [65] and spatial correlation coefficient (SCC) [66], which are utilized for assessing the spectral and spatial distortions, respectively; the erreur relative globale adimensionnelle de synthèse (ERGAS) index [67] and the Q2ⁿ index (Q4 and Q8 for 4 and 8 bands fused outcomes, respectively) [68], [69], which are employed to measure the global quality. The ideal values for SSIM, SCC, and Q2ⁿ are 1, for SAM (degree, °) and ERGAS are 0, whereas +∞ for PSNR. The *quality with no reference* (QNR) [70] index is applied to assess the performance at full resolution. It consists of a spectral distortion index D_λ and a spatial distortion index D_s . In particular, the highest QNR (i.e., 1) is obtained when both D_λ and D_s are 0 indicating the best-fused image quality. Furthermore, scale factors for all the experiments are 4, i.e., $r = 4$, the tolerance value is preset to $\varepsilon = 2 \times 10^{-5}$, and k_{mit} is empirically fixed to 200. The optimum parameters are set to provide the best compromise between the spectral and spatial distortions, measured by the SAM and the SCC, respectively.

For a broader comparative experimental analysis, some CNN approaches generating the \mathcal{X}_{net} are needed. Many state-of-the-art approaches for pansharpening can be found, see [33]–[36]. In our experiments, the PanNet proposed by Yang *et al.* [33] and the DiCNN proposed by He *et al.* [36] are selected to determinate the WPDI term. Then, these WPDI terms are employed to form our final schemes, called Ours_{PanNet} and Ours_{DiCNN}, respectively.

A. Data Sets

The satellite data sets, some of them freely available online,¹ are exploited to examine the effectiveness of the proposed pansharpening method, including: 1) the Tripoli and the Rio data sets both acquired by the WorldView-3 sensor including an MS image with 8 spectral bands and a PAN image with a spatial resolution of 0.3 m; 2) the Stockholm and the WashingtonDC data sets captured by the WorldView-2 satellite characterized by an MS image with 8 spectral bands and a PAN image at a spatial resolution of 0.4 m; and 3) the Indianapolis data set collected by the QuickBird sensor included an MS image with four spectral bands and a PAN image with a 0.61-m resolution at spatial.

B. Benchmark

We compare the proposed algorithm with some recent state-of-the-art techniques belonging to the CS, MRA, VO, and DL families. In greater details, the following algorithms are considered: EXP [55], PCA [10], GIHS [12], GSA [71], SFIM [18], MTF-GLP-HPM [72], GLP-Reg-FS [59], 18'TIP [3], 19'CVPR [29], PanNet [33], and DiCNN [36]. It is rather remarkable that the source codes of competitors are available at either of the website² or the authors' homepages. For fairness, we fine-tune the parameters of the 18'TIP and the 19'CVPR to get their best performance.

It is worth noting that the PanNet and the DiCNN are with consistent parameter configurations as that in [33] and [36],

respectively, and both of them are mainly re-trained on two different training data sets. Following Wald's protocol [37], the training data sets are simulated, as follows.

- 1) Training data from WorldView-3: 8806 {PAN, MS, GT} training samples with the size of 64×64 , $16 \times 16 \times 8$, and $64 \times 64 \times 8$, respectively, are contained.
- 2) Training data from QuickBird: 20685 {PAN, MS, GT} image pairs with the size of 64×64 , $16 \times 16 \times 4$, and $64 \times 64 \times 4$, respectively, are encompassed.

The PanNet and DiCNN trained based on the former are employed for the WorldView-3 experiments and WorldView-2 (verifying generalization ability), and trained to exploit the latter for the QuickBird experiments. Note that, all testing samples employed in Section V are spatially disjoint from the adopted training patches.

C. Qualitative Comparison

A true or pseudo color representation is selected for the qualitative analysis of the pansharpened results.

1) *Reduced Resolution Assessment*: As aforementioned, we give the comparative results at reduced resolution, as shown via an RGB rendering in Figs. 4–6. Inspecting Figs. 4 and 6, only the results of PanNet, DiCNN, and ours (including Ours_{PanNet} and Ours_{DiCNN}) can clearly retain details. Having a look at the error maps depicted in Figs. 4 and 6, we can clearly see that the PanNet, the DiCNN, and the proposed methods obtain the best performance both from a spectral and a spatial points of view compared to the GT image.

The performance in Fig. 5 differs from the ones in Fig. 4. Specifically, the PanNet and the DiCNN achieve poorer visual appearance with respect to many traditional methods, e.g., MTF-GLP-HPM [72] and GLP-Reg-FS [59], on the reduced resolution WashingtonDC data set, since the two networks are both trained on WorldView-3 data set (i.e., because of an unsatisfactory network generalization). In contrast, the results obtained by our framework show high visual quality as depicted in the related close-ups and having a look at the difference maps.

2) *Full Resolution Assessment*: Then, we corroborate the proposed models on full resolution images in order to further support the obtained results at reduced resolution. For these experiments, the images used for testing are extracted from the full resolution Rio and Stockholm data sets. All the visual inspections are displayed in Figs. 7 and 8. It is worth noting that the near-infrared (NIR) channel is selected to highlight the differences in the vegetated areas. As shown in Figs. 7 and 8, many results performed at full resolution are in line with those obtained on the reduced resolution Washington DC data set and our methods always obtain the details closest to the PAN image regarded as a reference for the spatial enhancement.

From the close-ups in the two experiments, we clearly observe that the highways and cars, which are blurred in the EXP images but recognizable in the PAN images, are super-resolved by our algorithms and without both visible artifacts and spectral distortion. Among the compared approaches, the GIHS often provides outcomes with severe spectral distortion, clearly visible to an easy inspection of the fused

¹<http://www.digitalglobe.com/samples?search=Imagery>

²<http://openremotesensing.net/kb/codes/pansharpening/>

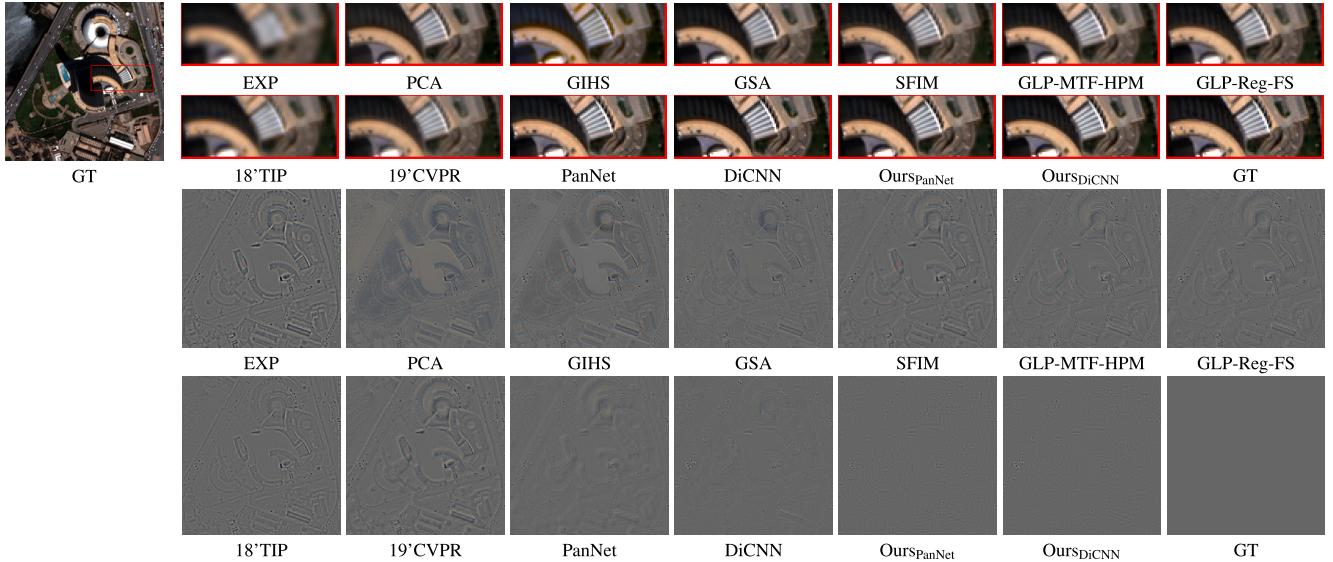


Fig. 4. Visual results on the reduced resolution Tripoli data set (source: WorldView-3). The scale of the PAN image is 256×256 . Top united row: the visual inspection of the GT image and the close-ups of images fused by EXP, PCA, GIHS, GSA, SFIM, MTF-GLP-HPM, GLP-Reg-FS, 18'TIP, 19'CVPR, PanNet, DiCNN, Ours_{PanNet}, Ours_{DiCNN}, and of the GT image, respectively. Bottom united row: the corresponding residual maps using the GT image as reference. For a better visualization, 0.4 has been added to all channels.

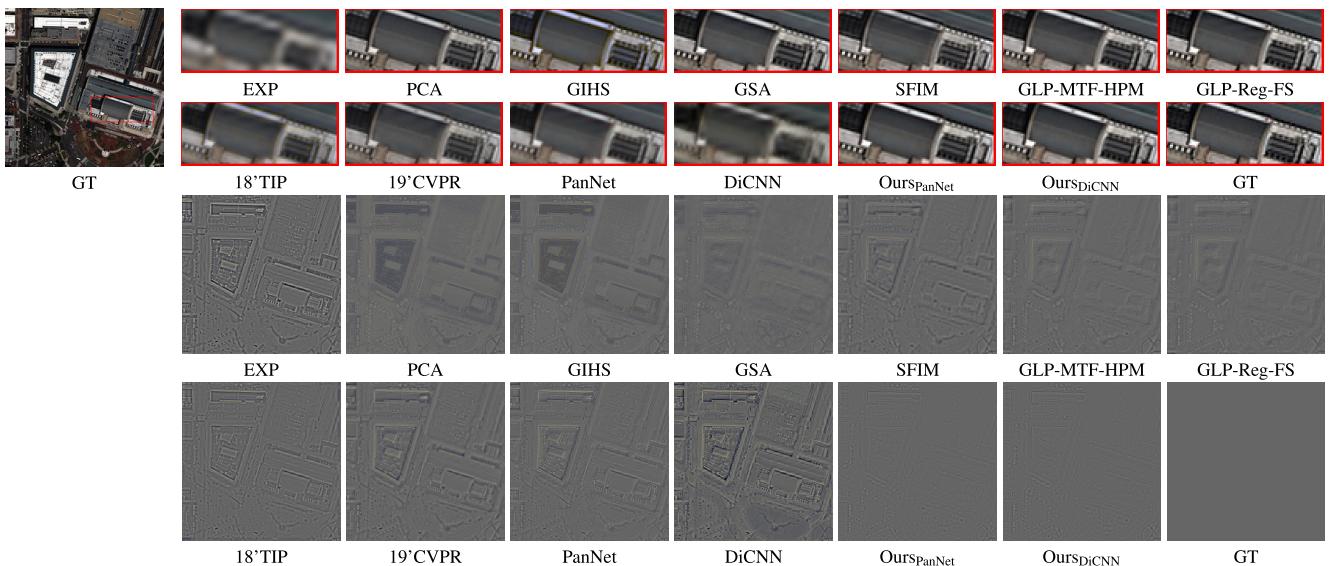


Fig. 5. Visual results on the reduced resolution Washington DC data set (source: WorldView-2). The scale of the PAN image is 256×256 . Top united row: the visual inspection of the GT image and the close-ups of images fused by EXP, PCA, GIHS, GSA, SFIM, MTF-GLP-HPM, GLP-Reg-FS, 18'TIP, 19'CVPR, PanNet, DiCNN, Ours_{PanNet}, Ours_{DiCNN}, and of the GT image, respectively. Bottom united row: the corresponding residual maps using the GT image as reference. For a better visualization, 0.4 has been added to all channels.

products. Thus, the superiority of the proposed framework is corroborated at full resolution.

D. Quantitative Comparison

To further assess the performance of the proposed method, we provide the quantitative comparisons across 75, 75, 5, 55, and 55 images with a corresponding PAN size on the five data sets in Figs. 4–8. These images are extracted in order to cover all the possible features in a scene, e.g., vegetation, water, buildings, and so forth. The statistical results for all the metrics and the execution times (in seconds, s) are shown in Tables I–III, IV-(a), IV-(b). We can clearly observe that our models consistently achieve much better average values (except for \mathbf{D}_λ) than the other methods on both reduced and

full-resolution data sets, implying that our framework can get competitive results. Moreover, our schemes can also achieve comparable efficiency with respect to variation methods.

To sum up, both qualitative and quantitative experiments are conducted indicating that the proposed scheme can get very high performance at reduced and full resolution, with respect to the benchmark. In particular, for the reduced resolution Tripoli and Indianapolis data sets, although the PanNet and the DiCNN methods also achieve excellent performance, our schemes improve them. Instead, for the reduced resolution WashingtonDC and the full resolution Rio and Stockholm data sets, where the effects of these networks are relatively lower, state-of-the-art results are still obtained by our framework. It is noteworthy that better results can probably be produced

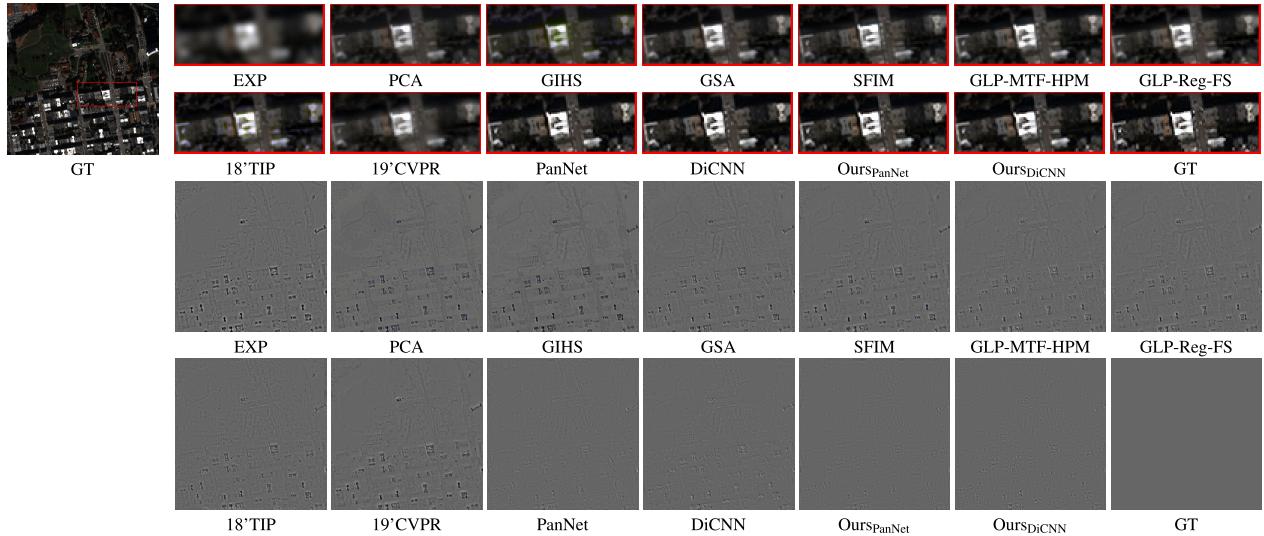


Fig. 6. Visual results on the reduced resolution Indianapolis data set (source: QuickBird). (Top united row) Visual inspection of the GT image and the close-ups of images fused by EXP, PCA, GIHS, GSA, SFIM, MTF-GLP-HPM, GLP-Reg-FS, 18'TIP, 19'CVPR, PanNet, DiCNN, Ours_{PanNet}, Ours_{DiCNN}, and of the GT image, respectively. (Bottom united row) Corresponding residual maps using the GT image as reference. For a better visualization, 0.4 has been added to all channels.

TABLE I

QUANTITATIVE RESULTS FOR ALL THE COMPARED METHODS ON 75 IMAGES FROM THE REDUCED RESOLUTION TRIPOLI DATA SET
(SOURCE: WORLDVIEW-3). ALL THE CNN METHODS ARE EXECUTED ON GPU (G), WHEREAS THE OTHER APPROACHES
USE THE CPU (C). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM°	SCC	ERGAS	Q8	Runtime(s)
EXP [55]	26.358 ± 1.870	0.641 ± 0.054	4.313 ± 0.694	0.860 ± 0.037	5.648 ± 1.086	0.722 ± 0.069	0.049(C)
PCA [10]	29.170 ± 1.591	0.857 ± 0.034	4.445 ± 0.658	0.949 ± 0.024	4.057 ± 0.737	0.865 ± 0.058	0.040(C)
GIHS [12]	28.436 ± 1.725	0.836 ± 0.033	5.099 ± 0.850	0.928 ± 0.021	4.332 ± 0.813	0.850 ± 0.050	0.003(C)
GSA [71]	30.701 ± 1.512	0.885 ± 0.029	3.955 ± 0.607	0.952 ± 0.020	3.382 ± 0.590	0.903 ± 0.055	0.096(C)
SFIM [18]	29.548 ± 1.778	0.846 ± 0.032	4.173 ± 0.640	0.939 ± 0.020	3.896 ± 0.760	0.874 ± 0.050	<u>0.029(C)</u>
MTF-GLP-HPM [72]	30.625 ± 1.627	0.875 ± 0.030	4.088 ± 0.611	0.949 ± 0.019	3.430 ± 0.627	0.902 ± 0.050	0.335(C)
GLP-Reg-FS [59]	31.026 ± 1.370	0.891 ± 0.029	3.969 ± 0.585	0.952 ± 0.019	3.251 ± 0.516	0.911 ± 0.055	0.350(C)
18'TIP [3]	29.165 ± 1.640	0.804 ± 0.025	4.270 ± 0.708	0.931 ± 0.019	4.168 ± 0.740	0.862 ± 0.053	94.030(C)
19'CVPR [29]	29.098 ± 1.625	0.830 ± 0.026	4.004 ± 0.635	0.934 ± 0.018	4.088 ± 0.708	0.857 ± 0.056	29.092(C)
PanNet [33]	33.730 ± 1.296	0.940 ± 0.035	3.299 ± 0.499	0.972 ± 0.018	2.356 ± 0.396	0.949 ± 0.053	1.225(G)
DiCNN [36]	34.473 ± 1.287	0.944 ± 0.035	3.285 ± 0.511	0.975 ± 0.019	2.173 ± 0.396	0.953 ± 0.052	1.245(G)
Ours _{PanNet}	34.980 ± 1.354	0.947 ± 0.035	3.068 ± 0.477	0.978 ± 0.018	2.043 ± 0.389	<u>0.957 ± 0.051</u>	24.772(C)
Ours _{DiCNN}	35.030 ± 1.348	0.947 ± 0.034	3.165 ± 0.504	0.978 ± 0.017	2.045 ± 0.392	0.958 ± 0.049	24.856(C)
Ideal value	$+\infty$	1	0	1	0	1	-

TABLE II

QUANTITATIVE RESULTS FOR ALL THE COMPARED APPROACHES ON 75 IMAGES FROM THE REDUCED RESOLUTION WASHINGTON DC DATA SET (SOURCE: WORLDVIEW-2). ALL THE CNN METHODS ARE EXECUTED ON GPU (G), WHEREAS THE OTHER APPROACHES USE THE CPU (C). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM°	SCC	ERGAS	Q8	Runtime(s)
EXP [55]	25.298 ± 1.991	0.572 ± 0.090	7.118 ± 1.123	0.816 ± 0.042	7.162 ± 0.826	0.653 ± 0.060	0.050(C)
PCA [10]	27.769 ± 2.282	0.796 ± 0.114	8.394 ± 2.820	0.892 ± 0.163	5.595 ± 1.737	0.772 ± 0.176	0.044(C)
GIHS [12]	28.025 ± 1.724	0.816 ± 0.040	7.873 ± 1.169	0.928 ± 0.025	5.140 ± 0.680	0.817 ± 0.071	0.003(C)
GSA [71]	31.215 ± 1.488	0.876 ± 0.038	6.763 ± 1.410	0.955 ± 0.014	3.729 ± 0.569	0.889 ± 0.075	0.098(C)
SFIM [18]	29.649 ± 1.855	0.850 ± 0.034	6.930 ± 1.361	0.938 ± 0.022	5.413 ± 6.446	0.864 ± 0.067	<u>0.029(C)</u>
MTF-GLP-HPM [72]	30.567 ± 1.552	0.870 ± 0.032	6.870 ± 1.395	0.946 ± 0.016	4.469 ± 2.221	0.884 ± 0.070	0.335(C)
GLP-Reg-FS [59]	31.004 ± 1.483	0.872 ± 0.040	7.004 ± 1.535	0.949 ± 0.016	3.838 ± 0.617	0.887 ± 0.076	0.349(C)
18'TIP [3]	28.436 ± 1.721	0.774 ± 0.038	6.986 ± 1.107	0.918 ± 0.018	5.175 ± 0.557	0.825 ± 0.067	95.459(C)
19'CVPR [29]	29.105 ± 1.776	0.830 ± 0.031	6.370 ± 1.080	0.936 ± 0.013	4.610 ± 0.513	0.842 ± 0.076	29.181(C)
PanNet [33]	29.183 ± 1.811	0.825 ± 0.031	6.152 ± 0.968	0.930 ± 0.016	4.579 ± 0.515	0.855 ± 0.077	1.225(G)
DiCNN [36]	25.846 ± 1.221	0.702 ± 0.052	7.675 ± 0.947	0.865 ± 0.031	6.712 ± 0.635	0.754 ± 0.073	1.245(G)
Ours _{PanNet}	32.127 ± 1.618	0.894 ± 0.026	6.104 ± 1.205	0.961 ± 0.011	3.322 ± 0.473	0.911 ± 0.071	25.082(C)
Ours _{DiCNN}	32.204 ± 1.604	0.897 ± 0.025	6.057 ± 1.169	0.962 ± 0.011	3.290 ± 0.470	0.914 ± 0.071	24.940(C)
Ideal value	$+\infty$	1	0	1	0	1	-

TABLE III

QUANTITATIVE RESULTS FOR ALL THE COMPARED APPROACHES ON 5 IMAGES FROM THE REDUCED RESOLUTION INDIANAPOLIS DATA SET (SOURCE: QUICKBIRD). ALL THE CNN METHODS ARE EXECUTED ON GPU (G), WHEREAS THE OTHER APPROACHES USE THE CPU (C). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM°	SCC	ERGAS	Q4	Runtime(s)
EXP [55]	27.725 ± 0.290	0.700 ± 0.007	8.234 ± 0.058	0.808 ± 0.002	10.635 ± 0.207	0.655 ± 0.004	<u>0.017(C)</u>
PCA [10]	29.153 ± 0.223	0.786 ± 0.003	8.827 ± 0.183	0.884 ± 0.006	9.017 ± 0.110	0.716 ± 0.004	<u>0.036(C)</u>
GIHS [12]	28.802 ± 0.222	0.771 ± 0.002	9.183 ± 0.074	0.877 ± 0.002	9.338 ± 0.119	0.682 ± 0.005	0.003(C)
GSA [71]	30.243 ± 0.197	0.815 ± 0.002	8.019 ± 0.064	0.899 ± 0.001	7.949 ± 0.078	0.772 ± 0.004	<u>0.053(C)</u>
SFIM [18]	29.989 ± 0.249	0.811 ± 0.004	7.799 ± 0.041	0.895 ± 0.001	8.195 ± 0.130	0.772 ± 0.002	<u>0.017(C)</u>
MTF-GLP-HPM [72]	30.905 ± 0.215	0.831 ± 0.002	7.590 ± 0.044	0.911 ± 0.001	7.374 ± 0.094	0.799 ± 0.003	<u>0.144(C)</u>
GLP-Reg-FS [59]	30.338 ± 0.178	0.818 ± 0.002	7.872 ± 0.057	0.899 ± 0.002	7.844 ± 0.066	0.777 ± 0.004	<u>0.146(C)</u>
18'TIP [3]	30.179 ± 0.220	0.791 ± 0.001	8.923 ± 0.052	0.896 ± 0.001	7.988 ± 0.106	0.762 ± 0.006	<u>39.296(C)</u>
19'CVPR [29]	29.508 ± 0.269	0.786 ± 0.004	7.747 ± 0.034	0.883 ± 0.001	8.704 ± 0.150	0.757 ± 0.002	<u>8.421(C)</u>
PanNet [33]	<u>34.718 ± 0.115</u>	<u>0.917 ± 0.001</u>	5.388 ± 0.047	<u>0.963 ± 0.001</u>	<u>4.763 ± 0.047</u>	<u>0.897 ± 0.004</u>	<u>0.831(G)</u>
DiCNN [36]	33.420 ± 0.194	0.902 ± 0.001	5.459 ± 0.039	0.950 ± 0.001	5.517 ± 0.060	0.886 ± 0.004	<u>0.852(G)</u>
Ours _{PanNet}	35.093 ± 0.106	0.919 ± 0.001	5.307 ± 0.051	0.965 ± 0.001	4.620 ± 0.042	0.906 ± 0.004	<u>5.920(C)</u>
Ours _{DiCNN}	33.986 ± 0.183	0.906 ± 0.001	5.383 ± 0.032	0.955 ± 0.001	5.247 ± 0.050	0.897 ± 0.003	<u>5.837(C)</u>
Ideal value	$+\infty$	1	0	1	0	1	-

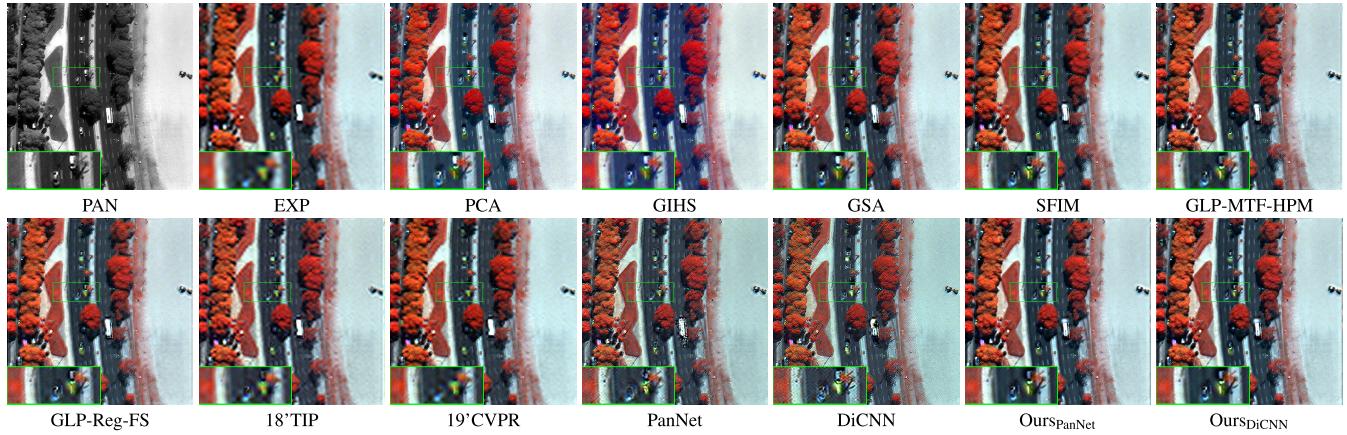


Fig. 7. Pseudo color (NIR(7th), G(3rd), and B(2nd) bands as R, G, and B channels) compositions for the full resolution Rio data set (source: WorldView-3). The scale of the PAN image is 400×400 . Close-ups are exhibited in the down-left corners, zooming in to see the details.

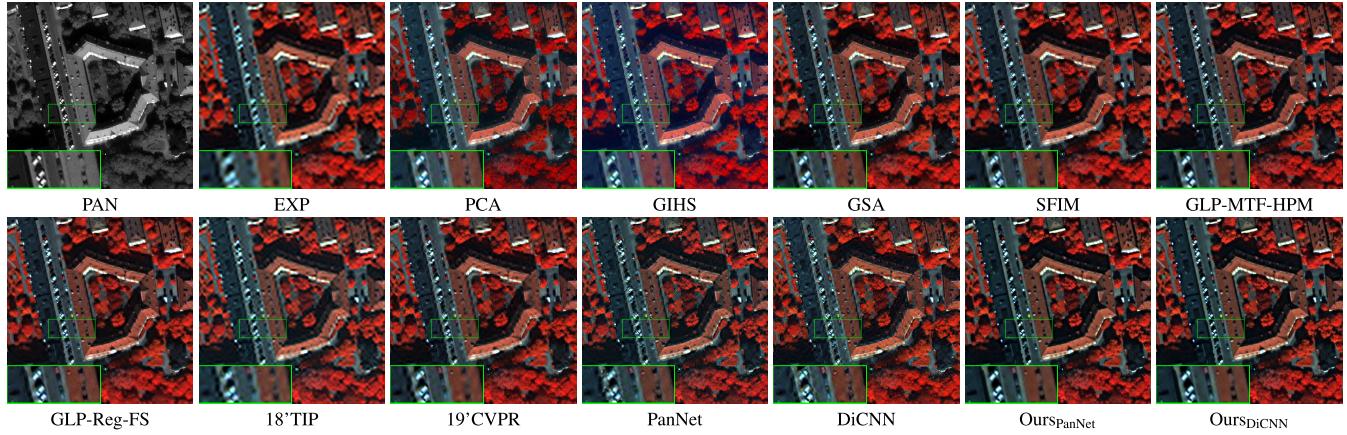


Fig. 8. False color [NIR(7th), G(3rd), and B(2nd) bands as R, G, and B channels] compositions for the full resolution Stockholm data set (source: WorldView-2). The scale of the PAN image is 400×400 . Close-ups are exhibited in the down-left corners, zooming in to see the details.

via fine-tuning the parameters of our schemes for different experiments on the same data set, but a fixed set of parameters is adopted for showing the algorithm robustness by varying the test cases.

E. Discussions

In this section, some discussions about the proposed scheme, see e.g., the parameters analysis and the ablation study, are given. All the discussions are about the proposed

TABLE IV

QUANTITATIVE RESULTS FOR ALL THE COMPARED METHODS ON (A) 55 IMAGES FROM THE FULL RESOLUTION RIO DATA SET (SOURCE: WORLDVIEW-3). (B) 55 IMAGES FROM THE FULL RESOLUTION STOCKHOLM DATA SET (SOURCE: WORLDVIEW-2), RESPECTIVELY. ALL THE CNN METHODS ARE EXECUTED ON GPU (G), WHEREAS THE OTHER APPROACHES USE CPU (C). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	(a) Full Resolution Rio Dataset			(b) Full Resolution Stockholm Dataset			Runtime(s)
	D _λ	D _s	QNR	D _λ	D _s	QNR	
EXP [55]	0.004 ± 0.002	0.111 ± 0.021	0.886 ± 0.021	0.004 ± 0.005	0.052 ± 0.020	0.944 ± 0.021	<u>0.076</u> (C)
PCA [10]	0.019 ± 0.013	0.075 ± 0.036	0.908 ± 0.041	0.036 ± 0.029	0.124 ± 0.040	0.845 ± 0.059	0.105(C)
GIHS [12]	0.058 ± 0.052	0.068 ± 0.040	0.879 ± 0.083	0.045 ± 0.020	0.090 ± 0.021	0.870 ± 0.034	0.007 (C)
GSA [71]	0.018 ± 0.009	0.048 ± 0.014	0.934 ± 0.021	0.027 ± 0.012	0.069 ± 0.014	0.907 ± 0.024	0.195(C)
SFIM [18]	0.024 ± 0.011	0.034 ± 0.016	0.943 ± 0.023	0.032 ± 0.011	0.058 ± 0.013	0.912 ± 0.021	0.096(C)
MTF-GLP-HPM [72]	0.033 ± 0.013	0.039 ± 0.014	0.929 ± 0.024	0.045 ± 0.014	0.054 ± 0.016	0.904 ± 0.026	0.416(C)
GLP-Reg-FS [59]	0.029 ± 0.011	0.037 ± 0.013	0.934 ± 0.020	0.049 ± 0.017	0.064 ± 0.016	0.890 ± 0.030	0.408(C)
18'TIP [3]	0.030 ± 0.017	0.067 ± 0.019	0.905 ± 0.026	0.021 ± 0.006	0.022 ± 0.009	0.958 ± 0.014	234.343(C)
19'CVPR [29]	<u>0.016 ± 0.006</u>	0.048 ± 0.013	0.937 ± 0.017	<u>0.016 ± 0.004</u>	0.039 ± 0.011	0.946 ± 0.012	87.237(C)
PanNet [33]	0.043 ± 0.019	0.044 ± 0.022	0.914 ± 0.037	0.029 ± 0.007	0.019 ± 0.009	0.953 ± 0.014	1.227(G)
DiCNN [36]	0.029 ± 0.013	0.041 ± 0.017	0.931 ± 0.024	0.031 ± 0.012	0.057 ± 0.022	0.914 ± 0.031	1.316(G)
Ours _{PanNet}	0.026 ± 0.013	<u>0.024 ± 0.012</u>	0.951 ± 0.023	0.020 ± 0.005	0.012 ± 0.006	0.968 ± 0.007	69.610(C)
Ours _{DiCNN}	0.021 ± 0.009	0.021 ± 0.010	0.959 ± 0.017	0.021 ± 0.004	0.014 ± 0.006	0.965 ± 0.007	73.621(C)
Ideal value	0	0	1	0	0	1	-

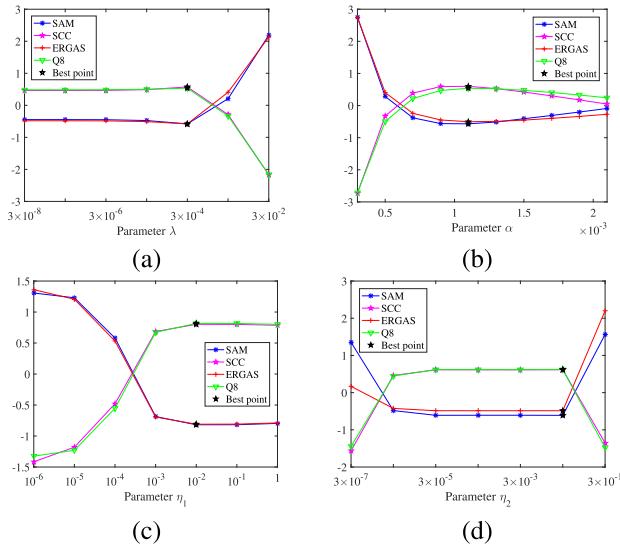


Fig. 9. SAM, SCC, ERGAS, Q8 curves for: the regularization parameters (a) λ and (b) α , the penalty parameters (c) η_1 and (d) η_2 on a reduced resolution Tripoli data (source: WorldView-3). The best points are pointed out with a black star. Note that, for better distinguishing the slight difference and reflecting the changing trend, we process the obtained indexes by $(\text{index} - \text{Mean(index)}) / \text{Std(index)}$, where $\text{Mean}(\cdot)$ and $\text{Std}(\cdot)$ represent the mean and standard deviation operations, respectively. Besides, the mean and standard deviation for SAM, SCC, ERGAS and Q8 are: (a) 0.30783 ± 0.2293 ; 0.9847 ± 0.0041 ; 2.0113 ± 0.2824 ; 0.9670 ± 0.0091 ; (b) 2.9458 ± 0.0185 ; 0.9870 ± 0.0003 ; 1.8529 ± 0.0350 ; 0.9718 ± 0.0007 ; (c) 3.0243 ± 0.1015 ; 0.9858 ± 0.0017 ; 1.9360 ± 0.1164 ; 0.9695 ± 0.0030 ; (d) 3.0390 ± 0.1583 ; 0.9854 ± 0.0028 ; 2.1918 ± 0.6798 ; 0.9682 ± 0.0060 .

Ours_{PanNet} approach using the reduced resolution Tripoli data (as well as other images if needed) for brevity.

1) *Parameters Analysis:* Four parameters are involved in the proposed model, i.e., λ , α , η_1 , η_2 . Fig. 9 depicts the performance varying these parameters on a reduced resolution Tripoli data. We fix all the parameters except for the one to be analyzed, aiming to make the parameter selection step simpler. It is clear that the combination of $\lambda = 3 \times 10^{-4}$, $\alpha = 1.1 \times 10^{-3}$, $\eta_1 = 10^{-2}$ and $\eta_2 = 3 \times 10^{-2}$ is the best choice whatever the adopted quality metric. Figs. 9(a)–(d) only show slight changes

TABLE V

QUANTITATIVE RESULTS FOR THE ABLATION EXPERIMENT USING THE REDUCED RESOLUTION TRIPOLI DATA (SOURCE: WORLDVIEW-3). (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	PSNR	SSIM	SAM [°]	SCC	ERGAS	Q8
PanNet [33]	34.655	0.937	3.189	0.983	2.127	0.965
Submodel-I	33.588	0.912	3.384	0.979	2.453	0.955
Submodel-II	<u>35.815</u>	<u>0.944</u>	<u>2.968</u>	0.987	<u>1.863</u>	0.972
Submodel-III	35.095	0.943	3.158	0.985	2.022	0.968
Ours _{PanNet}	35.976	0.945	2.935	0.987	1.834	0.972
Ideal value	$+\infty$	1	0	1	0	1

for all the considered indexes, thus confirming the robustness of our scheme related to its parameters. The same policy can be applied to other remote sensing data sets in order to determine the corresponding optimal parameter configuration.

2) *Ablation Study:* We conduct an ablation study on our model for a deeper insights, and then, three sub-models are generated as follows:

Submodel-I

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{X} \mathbf{B} \mathbf{S} - \mathbf{Y}\|_F^2 + \lambda \left\| \mathbf{X} - \mathbf{X} \mathbf{B} \circ \left(\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L \right) \right\|_F^2. \quad (38)$$

Submodel-II

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{X} \mathbf{B} \mathbf{S} - \mathbf{Y}\|_F^2 + \left\| \mathbf{W}_\alpha \circ (\mathbf{X} - \mathbf{X}_{\text{net}}) \right\|_F^2. \quad (39)$$

Submodel-III

$$\min_{\mathbf{X}} \frac{1}{2} \left\| \mathbf{X} - \mathbf{X} \mathbf{B} \circ \left(\widehat{\mathbf{P}} \otimes \widehat{\mathbf{P}}_L \right) \right\|_F^2 + \left\| \mathbf{W}_\alpha \circ (\mathbf{X} - \mathbf{X}_{\text{net}}) \right\|_F^2. \quad (40)$$

We perform the tests corresponding to Fig. 4, again, using these models with optimal parameters. The quantitative results are reported in Table V. We can observe that the models involving the WPDI term (i.e., Submodels-II and -III) get better performance than the ones of the PanNet, which is used to determinate the WPDI term. Furthermore, the fusion performance of the Submodel-II is comparable to that of the Ours_{PanNet} (see Table V) indicating the dominant role of the

TABLE VI
QUANTITATIVE RESULTS FOR THE DWDR EXPERIMENTS. (BOLD: BEST; UNDERLINE: SECOND BEST)

Dataset	Method	STDR	PSNR	SSIM	SAM°	SCC	ERGAS	Q8
Reduced resolution Tripoli	PanNet [33]	0.0086	34.655	0.937	3.189	0.983	2.127	0.965
	Ours _{PanNet} (Unweighted)		35.930	0.943	2.949	0.984	1.846	0.970
	Ours _{PanNet} (Weighted)		35.976	0.945	2.935	0.987	1.834	0.972
Reduced resolution WashingtonDC	PanNet [33]	0.0147	28.301	0.830	5.118	0.962	3.861	0.904
	Ours _{PanNet} (Unweighted)		32.775	0.925	4.706	0.984	2.353	0.964
	Ours _{PanNet} (Weighted)		32.969	0.927	4.673	0.987	2.304	0.967
Ideal value			+∞	1	0	1	0	1

TABLE VII

QUANTITATIVE RESULTS OBTAINED ON FOUR IMAGES EXTRACTED FROM THE REDUCED RESOLUTION TOULOUSE DATA SET
(SOURCE: IKONOS) AND 16 IMAGES EXTRACTED FROM THE REDUCED RESOLUTION PlÉIADES2 DATA SET
(SOURCE: PlÉIADES). (BOLD: BEST; UNDERLINE: SECOND BEST)

Dataset	Sensor	Method	PSNR	SSIM	SAM°	SCC	ERGAS	Q4
Toulouse	IKONOS	PanNet [33]	34.625 ± 0.592	0.873 ± 0.005	5.542 ± 0.352	0.791 ± 0.006	5.786 ± 0.290	0.723 ± 0.007
		Ours _{PanNet}	41.683 ± 0.089	0.956 ± 0.005	3.394 ± 0.275	0.932 ± 0.010	3.202 ± 0.224	0.897 ± 0.012
Pléades2	Pléades	PanNet [33]	27.953 ± 1.173	0.853 ± 0.019	4.983 ± 0.679	0.907 ± 0.018	5.442 ± 0.741	0.874 ± 0.020
		Ours _{PanNet}	38.061 ± 1.181	0.957 ± 0.008	2.876 ± 0.223	0.985 ± 0.002	2.142 ± 0.206	0.979 ± 0.003
Ideal value			+∞	1	0	1	0	1

spectral fidelity term. Moreover, all the three terms in our model contribute to the final results, none of them can be removed without decreasing performance. Finally, it is noteworthy that the quality metrics of Submodel-I (the variational model without the WPDI term) are aligned to the ones of the PanNet, thus overcoming the performance of the other traditional methods in the benchmark.

3) *Weighting Effect*: In order to assess the improvement in using the proposed DWDR, we show in Table VI all the quality metrics on the reduced resolution Tripoli and Washington DC data sets, respectively. The performance of both weighted and unweighted (i.e., $\mathbf{W}_\alpha = \sqrt{\alpha} \mathbf{1}$) models are reported in Table VI. The advantages of using the proposed schemes based on the DWDR are clear. Besides, a better improvement (i.e., in the case of the WashingtonDC data set) is achieved thanks to a higher STDR. This further verifies our initial cue. Thus, the rationale of using a WPDI term relied upon DWDR is experimentally proved.

4) *Generalization Ability*: The Washington DC and Stockholm data sets collected by the WorldView-2 satellite are used to feeding the proposed Ours_{PanNet} model showing a higher generalization ability in a previously reported analysis. To further evaluate the generalization ability, we perform additional experiments using exactly the same parameters on both 4 IKONOS data and 16 Pléades data. In these tests, the PanNet is trained on the 2nd (blue), the 3rd (green), the 5th (red), and the 7th (near-infrared) bands of a WorldView-3 data set, thus considering the spectral responses of both the IKONOS and the Pléades sensors and the WorldView-3 ones. The visualization for two examples is shown in Fig. 10 and the mean indicators are depicted in Table VII. It is extremely obvious that our scheme significantly exceeds the PanNet in all the cases and for all the quality metrics. Thus, the generalization ability of the proposed framework is clearly confirmed compared to a method relied upon a training by example philosophy as the PanNet.

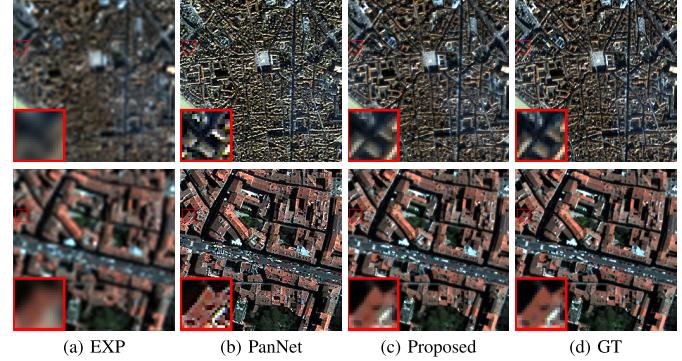


Fig. 10. (Top row) Visual results for the Toulouse data set at reduced resolution (source: IKONOS). (Bottom row) Visual results for the Pléades2 data set at reduced resolution (source: Pléades).

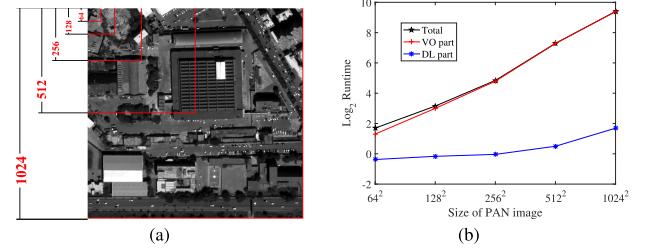


Fig. 11. Computational analysis on a full resolution Tripoli data (source: WorldView-3). (a) PAN image cropped into multiple scales. (b) Log₂ runtimes with different scale of the PAN image.

5) *Computational Burden*: All the computational analysis of our schemes, i.e., Ours_{PanNet} and Ours_{DiCNN}, are validated on the optimization model (19) in Section III. However, a complete calculation process should also include another non-negligible load, i.e., generating \mathcal{X}_{net} in advance exploiting the GPU (indicating only the testing time). In this case, the total computational load is naturally divided into two parts depending on VO and DL, respectively. Thus, we performed a computational analysis on the experimental data size using

TABLE VIII

COMPUTATION LOAD OF THE SUBPROBLEMS OF ALGORITHM 1 (TAKING FOR INSTANCE THE OURSPANNET SCHEME) USING THE FULL RESOLUTION TRIPOLI DATA (SOURCE: WORLDVIEW-3). THE PAN IMAGE IS PRESENTED IN FIG. 11 (A) WITH THE SCALE OF 256×256

Sub-problem	X	U	V	Λ_1	Λ_2
Run-time (s)	5.043	2.184	0.114	20.323	0.076

Ours_{PanNet} scheme, as displayed in Fig. 11. Furthermore, we further explored a deeper perspective on the computational burden of Algorithm 1 focusing on each subproblem, as reported in Table VIII. It is obvious that sub-problem V, which is related to the deep prior, takes up very little computational load of Algorithm 1, instead, multiplier Λ_1 occupies the main load, owing to the call of the MTF filter (filter size of 41×41^3).

VI. CONCLUSION

In this article, we have proposed a powerful weighted fusion scheme that combines VO with DL techniques to address the pansharpening problem, expecting to obtain higher accuracy, better modeling interpretability, and generalization ability. The novelty of the given model mainly included two aspects: 1) the use of the HPM scheme for spatial fidelity, resulting in a new details-injection term; and 2) the introduction of an adaptively weighted regularizer for considering the outcome of a DL method as a prior linking VO and DL classes. An efficient ADMM-based algorithm has been developed to pursue a global optimum solution. Both the reduced and full-resolution experiments are analyzed on several acknowledged data sets, qualitatively and quantitatively, and have indicated the superiority of our framework compared to a benchmark consisting of several state-of-the-art methods belonging to the four classes of pansharpening algorithms. Finally, the robustness with respect to the involved parameters, a generalization ability, an ablation study, the evaluation of the effects of using the proposed DWDR, and computational analysis have been pointed out in the discussions section of this article.

REFERENCES

- [1] A. Mohammadzadeh, A. Tavakoli, and M. J. Valadan Zoej, "Road extraction based on fuzzy logic and mathematical morphology from pan-sharpened ikonos images," *Photogramm. Rec.*, vol. 21, no. 113, pp. 44–60, Mar. 2006.
- [2] C. M. Souza, L. Firestone, L. M. Silva, and D. A. Roberts, "Mapping forest degradation in the eastern amazon from SPOT 4 through spectral mixture models," *Remote Sens. Environ.*, vol. 87, no. 4, pp. 494–506, Nov. 2003.
- [3] 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.
- [4] P. Chavez, S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data—Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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, Aug. 2019.
- [9] 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, Aug. 1987.
- [10] P. S. Chavez, Jr., 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.
- [11] 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.
- [12] W. J. Carper, T. M. Lillesand, and R. 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, Apr. 1990.
- [13] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6011875, Jan. 4, 2000.
- [14] 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.
- [15] 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.
- [16] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] R. Restaino, M. D. Mura, G. Vivone, and J. Chanussot, "Context-adaptive pansharpening based on image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 753–766, Feb. 2017.
- [21] 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.
- [22] 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, Jul. 2020.
- [23] H. Shen, X. Meng, and L. Zhang, "An integrated framework for the spatio-temporal-spectral fusion of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7135–7148, Dec. 2016.
- [24] J. Liu, "A rudin-oshier-fatemi model-based pansharpening approach using RKHS and AHF representation," *East Asian J. Appl. Math.*, vol. 9, no. 1, pp. 13–27, Jun. 2019.
- [25] 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.
- [26] G. Vivone, P. Addesso, R. Restaino, M. D. Mura, and J. Chanussot, "Pansharpening based on deconvolution for multiband filter estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 540–553, Jan. 2019.
- [27] 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.
- [28] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.

³For details about the MTF filter, see <http://openremotesensing.net/knowledgebase/a-critical-comparison-among-pansharpening-algorithms/>

- [29] 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. 10265–10274.
- [30] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.
- [31] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.
- [32] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3929–3938.
- [33] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5449–5457.
- [34] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang, "A new pan-sharpening method with deep neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1037–1041, May 2015.
- [35] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, p. 594, Jul. 2016.
- [36] L. He et al., "Pansharpening via detail injection based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1188–1204, Apr. 2019.
- [37] L.-J. Deng, G. Vivone, C. Jin, and J. Chanussot, "Detail injection-based deep convolutional neural networks for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, early access, Oct. 27, 2020, doi: [10.1109/TGRS.2020.3031366](https://doi.org/10.1109/TGRS.2020.3031366).
- [38] 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.
- [39] 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.
- [40] 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.
- [41] W. Liao et al., "Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2984–2996, Jun. 2015.
- [42] G. Vivone, L. Alparone, A. Garzelli, and S. Lolli, "Fast reproducible pansharpening based on instrument and acquisition modeling: AWLP revisited," *Remote Sens.*, vol. 11, no. 19, p. 2315, Oct. 2019.
- [43] L. Alparone, A. Garzelli, and G. Vivone, "Intersensor statistical matching for pansharpening: Theoretical issues and practical solutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4682–4695, Aug. 2017.
- [44] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009.
- [45] P. Liu, L. Xiao, and T. Li, "A variational pan-sharpening method based on spatial fractional-order geometry and spectral-spatial low-rank priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1788–1802, Mar. 2018.
- [46] Y. Jiang, X. Ding, D. Zeng, Y. Huang, and J. Paisley, "Pan-sharpening with a hyper-Laplacian penalty," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 540–548.
- [47] 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, Nov. 2018.
- [48] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.
- [49] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.
- [50] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.
- [51] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1529–1543, May 2020.
- [52] W. Dong et al., "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.
- [53] C. Chen, Y. Li, W. Liu, and J. Huang, "SIRF: Simultaneous satellite image registration and fusion in a unified framework," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4213–4224, Nov. 2015.
- [54] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniques—An introduction, review and comparison," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 4, pp. 249–263, Sep. 2007.
- [55] 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, Jan. 2002.
- [56] R. Restaino, G. Vivone, M. D. Mura, and J. Chanussot, "Fusion of multispectral and panchromatic images based on morphological operators," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2882–2895, Jun. 2016.
- [57] L. Loncan et al., "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [58] S. Baronti, B. Aiazzi, M. Selva, A. Garzelli, and L. Alparone, "A theoretical analysis of the effects of aliasing and misregistration on pansharpened imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 446–453, Jun. 2011.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [63] R. Glowinski and P. L. Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. Philadelphia, PA, USA: SIAM, 1989.
- [64] 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.
- [65] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, vol. 1, 1992, pp. 147–149.
- [66] 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.
- [67] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.
- [68] 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.
- [69] 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.
- [70] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.
- [71] 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.
- [72] 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. 22nd Digit. Avionics Syst. Conf.*, May 2003, pp. 90–94.



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

His research interests include multispectral image fusion and spectral super-resolution.



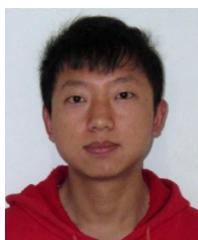
Jin-Fan Hu received the B.S. degree in computational mathematics from the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2019, where he is pursuing the M.S. degree in mathematics with the School of Mathematical Sciences.



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 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 (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 the Case Western Reserve University, Cleveland, OH, USA. In 2017, he was a postdoc with Hong Kong Baptist University (HKBU). In addition, he also stayed at the Isaac Newton Institute for Cambridge University and HKBU for short visits. He is a Research Fellow with the School of Mathematical Sciences, UESTC. His research interests include the use of partial differential equations (PDE), optimization modeling, and deep learning to address several tasks in image processing, and computer vision, e.g., resolution enhancement and restoration.



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 a Researcher with the National Research Council, Rome, Italy. His main research interests focus on statistical signal processing, detection of remotely sensed images, data fusion, and tracking algorithms.