A New Variational Approach Based on Proximal Deep Injection and Gradient Intensity Similarity for Spatio-spectral Image Fusion

Zhong-Cheng Wu, Ting-Zhu Huang, *Member, IEEE*, Liang-Jian Deng, *Member, IEEE*, Gemine Vivone, *Senior Member, IEEE*, Jia-Qing Miao, Jin-Fan Hu, Xi-Le Zhao, *Member, IEEE*

***Abstract*—Pansharpening is a very debated spatio-spectral fusion problem. It refers to the fusion of a high spatial resolution panchromatic (PAN) image with a lower spatial but higher spectral resolution multispectral (LRMS) image in order to obtain an image with high resolution in both the domains. In this paper, we propose a novel variational optimization-based (VO) approach to address this issue incorporating the outcome of a deep convolutional neural network (DCNN). This solution can take advantages of both the paradigms. On one hand, higher performance can be expected introducing machine learning methods based on the training by examples philosophy into VO approaches. On other hand, the combination of VO techniques with DCNNs can aid the generalization ability of these latter. In particular, we formulate a ℓ_2-based proximal deep injection term to evaluate the distance between the DCNN outcome and the desired high spatial resolution multispectral image. This represents the regularization term for our VO model. Furthermore, a new data fitting term measuring the spatial fidelity is proposed. Finally, the proposed convex VO problem is efficiently solved by exploiting the framework of the alternating direction method of multipliers, thus guaranteeing the convergence of the algorithm. Extensive experiments both on simulated and real datasets demonstrate that the proposed approach can outperform state-of-the-art spatio-spectral fusion methods, even showing a significant generalization ability. Please find the project page: https://liangjiandeng.github.io/Projects_Res/DMPIF_2020jstars.html.**

***Index Terms*—Variational Approaches, Deep Convolutional Neural Networks, Dynamic Gradient Sparsity, Gradient Intensity Similarity, Pansharpening, Image Fusion, Remote Sensing.**

I. Introduction

Multispectral (MS) remote sensing images have become widely exploited in many fields, such as environmental monitoring, agriculture and classification. However, due to constraints on the signal-to-noise ratio, many existing remote sensing sensors, *e.g.*, IKONOS, Pl*é*iades, WorldView-2 and WorldView-3, have to make a fundamental tradeoff between the spatial and spectral resolutions [3], [4], [5], [6], [7], [8]. Generally, they almost simultaneously capture two images of the same scene in pursuit of richer information, including a panchromatic (PAN) image with a higher spatial resolution but a unique spectral channel and an MS image with a lower spatial resolution but a better spectral content. The PAN/MS fusion (the so-called pansharpening) is committed to integrating the spatial details contained in the PAN image and the spectral information contained in the low spatial resolution MS (LRMS) image to reconstruct a high spatial resolution MS (HRMS) image. For the convenience of understanding intuitively, the schematic of the PAN/MS fusion, as well as several fused images, are presented in Fig. 1.

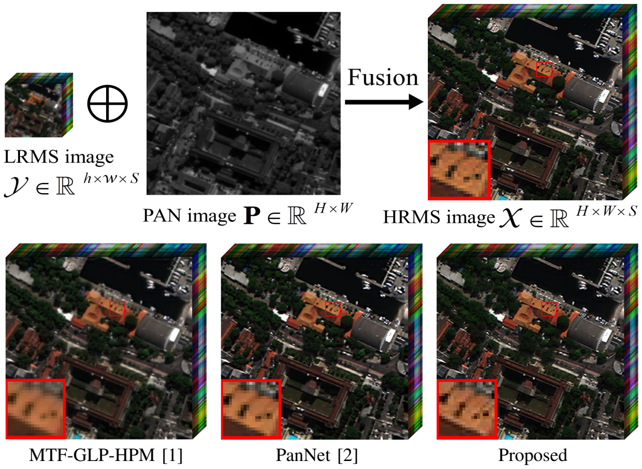

LRMS image $\mathcal{Y} \in \mathbb{R}^{h\times w\times S}$
$\oplus$
PAN image $\mathbf{P} \in \mathbb{R}^{H\times W}$
Fusion
HRMS image $\mathcal{X} \in \mathbb{R}^{H\times W\times S}$
MTF-GLP-HPM [1]
PanNet [2]
Proposed

Fig. 1. Top row: the schematic of the PAN/MS fusion on a simulated Rio data (source: WorldView-3). Bottom row: fused images by the MTF-GLP-HPM [1], the PanNet [2] and the proposed method, respectively. The red, green and blue channels of the shown images are extracted from the 5-th, the 3-rd and the 2-nd bands, respectively. Note that, the HRMS image $\mathcal{X}$ displayed in the top row is the ground-truth (GT) image.

Spatio-spectral image fusion has aroused widespread interest in academia and numerous methods have been proposed so far [9], [10], [11], [12], [13], [14], [15]. Most of them can be generally classified into four major categories [16], [17]: 1) Component substitution (CS) methods; 2) Multi-resolution analysis (MRA) methods; 3) Variational optimization (VO) approaches; and 4) Machine learning (ML) techniques.

The work is supported by National Natural Science Foundation of China grants 61772003, 61702083, and 61876203, and Key Projects of Applied Basic Research in Sichuan Province (Grant No. 2020YJ0216).

Z. -C. Wu, T. -Z. Huang, L. -J. Deng, J. -F. Hu and X. -L. Zhao are with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, China (e-mail: wuzhch97@163.com; tingzhuhuang@126.com; liangjian.deng@uestc.edu.cn; 903540917@qq.com; xlzhao122003@163.com).

G. Vivone is with the Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno, 84084 Fisciano, Italy and with the Institute of Methodologies for Environmental Analysis, CNR-IMAA, 85050 Tito Scalo, Italy (e-mails: gvivone@unisa.it; gemine.vivone@imaa.cnr.it).

J. -Q. Miao is with the School of Computer Science and Technology, University of Southwest Minzu of China, Chengdu, Sichuan, 610041, China (e-mail: mjq_011114117@163.com).

Corresponding author: Ting-Zhu Huang; Liang-Jian Deng

The CS methods are the earliest and the most widely used methods due to their extremely low computational cost [17]. The principle of such an approach is the substitution of the spatial component, which is derived by spectrally transforming the LRMS image, with the PAN image. Among them, the intensity-hue-saturation (IHS) [18], the principal component analysis (PCA) [19], the Brovey transform [20], the Gram-Schmidt (GS) spectral sharpening [21] and the partial replacement adaptive component substitution (PRACS) [22] are very representative. These methods project first the upsampled LRMS images (LMS) into a new transformation domain, then replace the projected spatial component with the PAN image and, finally, perform an inverse projection to get the HRMS image. Compared to other approaches, the CS methods have a reduced computational burden. However, they usually cause severe spectral distortion [23].

In addition to CS methods, MRA approaches are other popular fusion techniques, which inject the spatial structure information extracted from the PAN image via spatial filtering into the LMS image in order to get the HRMS image. Powerful instances belonging to this class are the smoothing filter-based intensity modulation (SFIM) [24], the modulation transfer function generalized laplacian pyramid with full resolution regression-based injection model (GLP-Reg-FS) [25], and the modulation transfer function generalized Laplacian Pyramid with high-pass modulation injection model (MTF-GLP-HPM) [1]. The fused images provided by MRA approaches mainly suffer from spatial distortion, whereas spectral information is usually well-preserved. In general, MRA and CS methods can obtain relatively satisfactory fusion results. Yet, despite the advantages, neither of them establish an explicitly relational model between the observed and the desired image, which could reduce performance on certain data [17].

Different from the CS and MRA methods, the VO methods describe an exact link among the LRMS image, the PAN image, and the ideal HRMS image based on some observations and assumptions, thus formulating an energy function. Then, the desired HRMS image can be obtained by regularizing this energy function and solving an optimization problem. For instance, in [4], Fu *et al.* consider the gradient difference of the PAN image and the HRMS image in different local patches and bands rather than global constraints incorporating the spatial preservation into the proposed variational model. Generally speaking, VO methods can theoretically produce excellent results, both spatially and spectrally, with a solid mathematical foundation [17]. Unfortunately, they usually generate many unpredictable deviations once some unreasonable assumptions are made. Besides, the misalignment among spectral bands is also a tricky issue, which can cause ghosting effects. Although some algorithms with a high registration capability have been proposed, *e.g.*, [26], the fused image is usually formed by low quality high frequency structures.

The ML methods have been proposed in recent years. In particular, deep convolutional neural networks (DCNNs) show an excellent capability for nonlinear mapping learning and feature extraction [17], [27], see, *e.g.*, [2], [28], [29], [30], [31]. These methods can perfectly compensate the deficiencies of the VO methods about nonlinear mapping and managing misalignments getting state-of-the-art performance [32]. For these methods, training the network to map information is a necessary step before the fusion step. Afterwards, the fused image can be obtained by inputting the LRMS image and PAN image into the learned network. However, they are often over-dependent on the training data [4], so that the generalization of many DCNN methods is limited by their training data, *i.e.*, they have excellent performance only on data similar to the ones in the training set. In particular, because of the network parameters are fixed once the training is finished, the accuracy of DCNN-based methods cannot be further improved [4].

In this paper, we propose a novel VO approach for fusing MS and PAN images using the proximal deep injection (PDI), *i.e.*, formulating the output of a DCNN as the proximal term and integrating it into the proposed variational model for further optimization. Specifically, the proposed model consists of two data fitting terms (the spectral and the spatial ones) and a DCNN-based proximal term, namely the PDI. The spectral fidelity is imposed on the LRMS image and enables the desired image to adequately receive spectral information. The spatial counterpart is imposed by exploiting a $\ell_{2,1}$ norm encouraging dynamic gradient sparsity and group sparsity simultaneously, *i.e.* the group dynamic gradient sparsity (GDGS) [26], between the desired image and a reference PAN (RePAN) image. Unlike the original work [26] that uses as RePAN a simple replication of the PAN image along the spectral dimension, the proposed RePAN is relied upon both the LRMS and the PAN image in order to take into account of the spectral content of the spatial details. This leads to the proposed gradient intensity similarity constraints (GISC). Finally, we exploit the prior knowledge provided by a DCNN into the proposed variational model through the PDI that represents a regularization term in our framework. This new variational optimization problem is solved by designing an alternating direction method of multipliers (ADMM)-based algorithm, which is guaranteed to efficiently converge to the global minimum. Extensive experiments both on simulated and real datasets confirm the superiority of the proposed VO approach compared to other state-of-the-art methods. The flowchart of the proposed model is provided in Fig. 2.

The *contributions* of this paper are summed up as follows:

- The concept of PDI is proposed merging the outcome of a DCNN-based method with a VO approach. PDI represents a regularization term that can leverage on the output of any DCNN to improve the results in the proposed VO framework. Thus, this combination is able to complement the benefits of the DCNNs improving their generalization ability for the spatio-spectral fusion problem.
- A new spatial fidelity term is proposed exploiting a $\ell_{2,1}$ norm encouraging simultaneously dynamic gradient sparsity and group sparsity between the desired image and the properly designed RePAN image.
- A new algorithm based on ADMM is presented to solve the proposed optimization problem based on a classical spectral fidelity term and the proposed spatial fidelity and regularization terms.
- A broad experimental analysis both on simulated and real

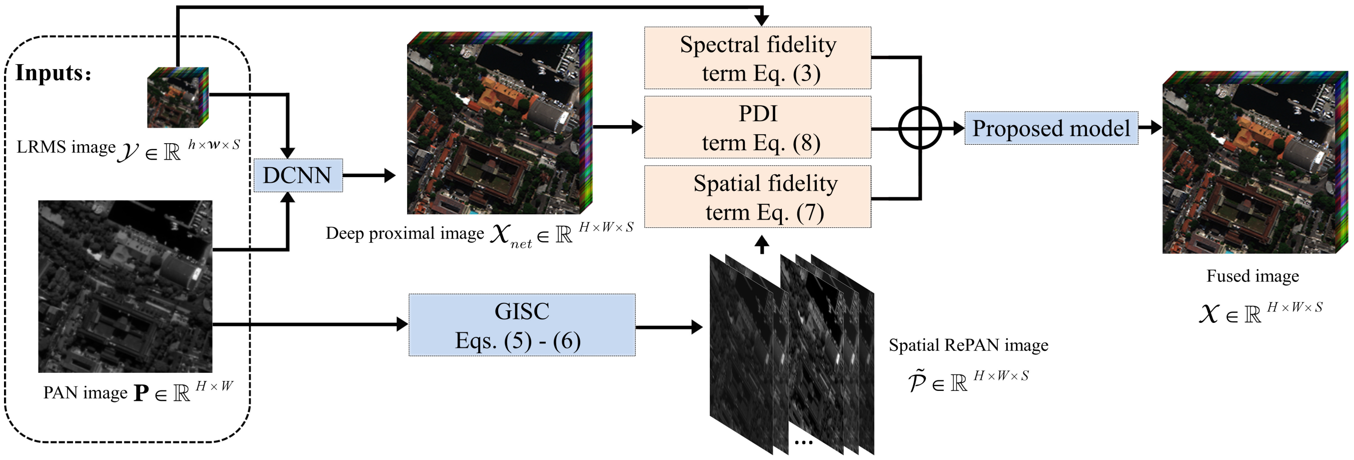

Inputs:
LRMS image $\mathcal{Y} \in \mathbb{R}^{h\times w\times S}$
PAN image $\mathbf{P} \in \mathbb{R}^{H\times W}$
DCNN
Deep proximal image $\mathcal{X}_{net} \in \mathbb{R}^{H\times W\times S}$
Spectral fidelity term Eq. (3)
PDI term Eq. (8)
Spatial fidelity term Eq. (7)
Proposed model
GISC Eqs. (5) - (6)
Spatial RePAN image $\tilde{\mathcal{P}} \in \mathbb{R}^{H\times W\times S}$
...
Fused image
$\mathcal{X} \in \mathbb{R}^{H\times W\times S}$

Fig. 2. The flowchart of our model. The GISC denotes the operation as in (5)-(6), see Sec. III-B. The acronym DCNN stands for deep convolutional neural network.

datasets is conducted to assess the performance of the proposed VO approach in different scenarios and with different acquisition devices. Furthermore, a deep robustness analysis about some crucial parameters has also been provided to the readers together with an ablation study and an assessment of the generalization abilities of the proposed approach.

The remaining of the paper is organized as follows. In Section II, the main notation and motivations will be briefly introduced. The proposed VO model is described in Section III. Afterwards, the solution of the proposed VO problem is provided in Section IV. In Section V, the experimental analysis is shown comparing the proposed method with some state-of-the-art approaches. Finally, the conclusion is drawn in Section VI.

II. Notation and Motivations

This section is devoted to the presentation of the notation and the motivations under the developing of the proposed method.

A. Notation

Before introducing the proposed model, it is necessary to state the notation used in this paper. In particular, scalars are denoted by lowercase letters, uppercase and lowercase bold letters denote matrices and vectors, respectively, and calligraphic letters denote tensor. Moreover, the main symbols and acronyms used in this paper are listed below:

- High-resolution multispectral image (HRMS): the desired HRMS image $\mathcal{X} \in \mathbb{R}^{H\times W\times S}$ with S spectral images $\mathbf{X}_i \in \mathbb{R}^{H\times W}$, $i = 1, 2, \dots, S$.
- Low-resolution multispectral image (LRMS): the observed LRMS image $\mathcal{Y} \in \mathbb{R}^{h\times w\times S}$ with S spectral images $\mathbf{Y}_i \in \mathbb{R}^{h\times w}$, $i = 1, 2, \dots S$. In particular, $h \times r = H$ and $w \times r = W$, where r denotes the scale factor between $\mathbf{X}_i$ and $\mathbf{Y}_i$,
- Panchromatic image (PAN): the observed single channel PAN image $\mathbf{P} \in \mathbb{R}^{H\times W}$.
- Reference panchromatic image (RePAN): the PAN image $\widetilde{\mathcal{P}} \in \mathbb{R}^{H\times W\times S}$ used as reference for the spatial fidelity term in the proposed model.

B. Motivations

DCNN methods have recently been attracted attention for their ability in nonlinear mapping yielding competitive results even for image fusion. However, they have also been shown some drawbacks, as analyzed in Section I, because of the huge data dependence and the use of network parameters that can be hardly adjusted on data very different from the ones shown in the training phase. Therefore, an optimization strategy needs to be developed to compensate for these shortcomings and to further improve the performance. For this reason, we propose the concept of PDI, aiming to introduce the output of a DCNN-based method into a VO framework. The PDI term establishes a relationship between this output and the desired HRMS image, thus transferring DCNNs into a variational fusion framework.

Furthermore, there is a need to complete the variational model including the data fitting terms relating the spectral and the spatial data in input with the desired and unknown HRMS data. In that model, the PDI plays the role of a regularization term avoiding the use of other additional prior information often used in the related literature, see, *e.g.*, sparse priors [33] and low-rank priors [34]. Having a look at most of the existing VO methods, the design of the spectral fidelity term is almost the same for all the techniques. Instead, a crucial choice regards the selection of the spatial fidelity data fitting term accounting for the spatial content into the PAN image. The difficulty is to model the indirect and nonlinear relationship between the PAN data and the unknown HRMS image. Hence, an accurate design of the spatial fidelity term for VO models can significantly improve the fusion outcomes, as shown in this work.

III. The Proposed Model

In this section, the adopted model is presented. The general framework for spatio-spectral fusion is as follows

$$\min_{\mathcal{X}} f_{spec}(\mathcal{X},\mathcal{Y}) + \lambda f_{spat}(\mathcal{X},\mathbf{P}) + \alpha f_{PDI}(\mathcal{X},\mathcal{X}_{net}), \quad (1)$$

where $f_{spec}(\mathcal{X},\mathcal{Y})$ and $f_{spat}(\mathcal{X},\mathbf{P})$ are the spectral and spatial fidelity terms, respectively, $f_{PDI}(\mathcal{X},\mathcal{X}_{net})$, *i.e.* the PDI term, plays the role of a regularization term, $\mathcal{X}_{net} \in \mathbb{R}^{H\times W\times S}$ represents the output of a generic DCNN method, λ is a positive regularization parameter, and α is a crucial parameter that links the VO model and the output of a DCNN. For the convenience of discussion, (1) can be rewritten in matrix form as

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

where $\mathbf{X}$, $\mathbf{X}_{net} \in \mathbb{R}^{S\times HW}$ and $\mathbf{Y} \in \mathbb{R}^{S\times hw}$ denote the mode-3 unfolding of $\mathcal{X}$, $\mathcal{X}_{net}$ and $\mathcal{Y}$, respectively. Please refer to [35] for more details about decompositions and applications of tensors.

A. The Spectral Fidelity Term

Many existing fusion methods, *e.g.* [20], [22], [36], upsample the LRMS image and extract spectral information from this upsampled image. However, inaccurate information could be introduced using this simple approach, thus impacting on the performance. Therefore, we consider in this paper the downsampled version of the unknown HRMS image. This operation is performed according to the point spread function of the spaceborne sensor [37], [38] in order to design the blurring operation. Thus, we have that

$$f_{spec}(\mathbf{X},\mathbf{Y}) = \|\mathbf{XBS}-\mathbf{Y}\|_F^2, \quad (3)$$

where $\mathbf{B} \in \mathbb{R}^{HW\times HW}$ is the convolution matrix that blurs each row of $\mathbf{X}$, $\mathbf{S} \in \mathbb{R}^{HW\times hw}$ denotes the decimation matrix and $\|\cdot\|_F$ is the Frobenius norm. In particular, $\mathbf{B}$ is a block-circulant-circulant-block matrix, where the periodic boundary conditions for the rows of $\mathbf{X}$ are satisfied. This property for $\mathbf{B}$ is at the basis of many fast deblurring algorithms [39], [40], [41]. In this work, we do not estimate the kernel $\mathbf{B}$, even in the case of real data. A classical assumption is to have filters matched with the MS sensor's modulation transfer function (MTF) to extract spatial details, which is also the way used in the work. Usually, these filters are set exploiting some prior information, as the gains at Nyquist frequency. Obviously, for some reasons (*e.g.*, aging) these values could be slightly wrong and for these reasons some recent research has been focused on this issue, see *e.g.*, [42], [43]. Anyway, the estimation of the convolutional blur $\mathbf{B}$ is out-of-scope of this paper, but it can surely deserve future developments to slightly improve the performance in particular at full resolution. More details about the matrices $\mathbf{B}$ and $\mathbf{S}$ can be found in [37], [38], [41], [44], [45].

The spectral fidelity term establishes a direct relationship between $\mathbf{X}$ and $\mathbf{Y}$. Thus, we can map the whole spectral information of $\mathbf{Y}$ in $\mathbf{X}$. Despite this, spatial structures are still missing for the ill-conditioning of $\mathbf{S}$. Thus, this term alone is far from being able to reconstruct the missing spatial information.

B. The Proposed Spatial Fidelity Term

It is well-known that the distribution of materials is locally continuous, which tends to generate piecewise smooth data. Therefore, the gradient of the desired HRMS image should be sparse and the non-zero elements correspond to the positions of the boundaries of the spatial structures. Besides, since the PAN image and the LRMS image are captured over the same scene, the boundary locations of the desired HRMS image should be theoretically unified with the ones of the PAN image and the same properties should be present across all the bands of the HRMS image. Thus, this connection can be established on the gradient domain. Chen *et al.* [26] assign the pixels with the same spatial position across all the channels of $\mathcal{X}$ into one group and constrain their sparsity using $\ell_{2,1}$ norm defined as follows

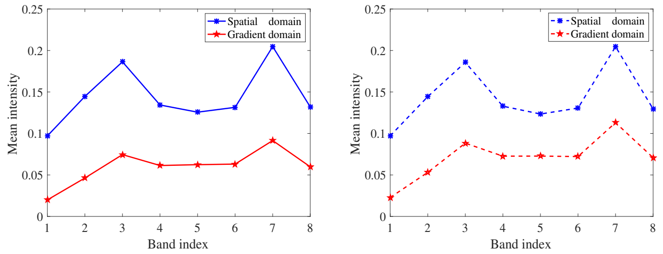
$$\begin{aligned}&\|\nabla\mathcal{X}-\nabla\mathcal{P}\|_{2,1} = \\ &= \sum\nolimits_i \sum\nolimits_j \sqrt{\sum\nolimits_k \sum\nolimits_q \left(\nabla_q \mathcal{X}_{i,j,k} - \nabla_q \mathcal{P}_{i,j,k}\right)^2}\quad ,\end{aligned} \quad (4)$$

where ∇_q, $q = 1, 2$, denote the forward finite difference operators on the first and second coordinates, respectively, and $\mathcal{P}$ indicates the expansion of $\mathbf{P}$ by duplicating it along the S bands. The $\ell_{2,1}$ norm constrains the sparsity of each band of $\mathcal{X}$ enforcing the same position of the structures as in the PAN image. Furthermore, the grouping improves the spectral correlation of the structures.

However, some shortcomings still exist when $\mathcal{P}$ is used. In view of this, we propose a novel concept of gradient intensity similarity constraints (GISC). In particular, it is worth to be pointed out that the spectral response of any material to various wavelengths is different. Hence, the gradient intensities of the desired HRMS image are not usually constant along the spectral bands, thus presenting a spectral content. However, simply replicating the PAN image over the bands does not help to support the last statement. Indeed, having a look at (4), we can note that both the position and the gradient intensity of $\mathcal{X}$ are forced to the $\mathcal{P}$ ones, thus impacting on the spatial optimization. Therefore, a pivotal improvement making the gradient intensity structures of $\mathcal{P}$ as similar as possible to the desired ones (*i.e.*, the ones of the target HRMS image) is advisable.

The variation of the gradient intensity means along the spectral bands is analyzed in Fig. 3. In particular, this figure shows the mean intensity both on the original spatial and the gradient domains on the simulated Rio dataset captured by the WorldView-3 sensor. The analysis is conducted both on the reference (high spatial resolution) MS image, see Fig. 3(a), and the LRMS image, see Fig. 3(b). From Fig. 3(a), it is clear to see that there is a positive correlation between the spatial and the gradient domain curves. Therefore, we can reasonably conclude that the desired gradient intensity structures can be reconstructed along the spectral bands by exploiting the means of the spectral bands in the spatial domain. Unfortunately, the reference (ground-truth, GT) HRMS image is unavailable

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2020.3030129, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing



(a) (b)

Fig. 3. Mean intensity on the spatial and the gradient domains for (a) the ground-truth (reference) image and (b) the corresponding LRMS image on the simulated Rio WorldView-3 dataset.

(because it is the goal of the fusion process). Thus, in Fig. 3(b), the same analysis as in Fig. 3(a) is performed on the available LRMS image. The same trend as in Fig. 3(a) can be pointed out, thus suggesting the use of the means along the spectral bands of the LRMS image $\mathcal{Y}$ to drive the reconstruction of the desired gradient intensity structures. As a result, we can adjust the spatial intensity mean of the i-th band of $\mathcal{P}$ (say it $\mathbf{P}_i$) to be equal to that of $\mathbf{Y}_i$, thus capturing the desired gradient intensity structures. The improved RePAN image $\widetilde{\mathcal{P}}$ can be obtained as

$$\widetilde{\mathcal{P}} = \psi(\mathcal{Y}, \mathcal{P}), \tag{5}$$

where ψ represents a linear definition operator defined as follows

$$\widetilde{\mathbf{P}}_i := \frac{\mathcal{M}(\mathbf{Y}_i)}{\mathcal{M}(\mathbf{P}_i)}\mathbf{P},\ i = 1, 2, \ldots, S., \tag{6}$$

where $\widetilde{\mathbf{P}}_i$ denotes the i-th band of the $\widetilde{\mathcal{P}}$ and $\mathcal{M}\cdot$ represents the mean operator.

Based on the analysis above, the GDGS term with GISC can be employed to enhance the spatial resolution of $\mathbf{X}$. Combining (5)-(6) with (4), the spatial fidelity term in matrix form can be written as

$$f_{spat}(\mathbf{X}, \widetilde{\mathbf{P}}) = \left\|\nabla\mathbf{X} - \nabla\widetilde{\mathbf{P}}\right\|_{2,1}, \tag{7}$$

where $\widetilde{\mathbf{P}} \in \mathbb{R}^{S\times HW}$ is the mode-3 unfolding of $\widetilde{\mathcal{P}}$.

The major difference between (7) and (4) proposed in [26] is the used RePAN image. In particular, the proposed RePAN image can transfer many more similar gradient intensity structures than the solution proposed in [26], thus achieving higher performance. The effects of the new proposed spatial fidelity term are assessed in Section V.

C. The PDI Term

As stated in Section II-B, we introduce the output of a DCNN-based method into the proposed VO model through the PDI term. Hence, we have that

$$f_{PDI}(\mathbf{X}, \mathbf{X}_{net}) = \|\mathbf{X} - \mathbf{X}_{net}\|_F^2 . \tag{8}$$

Therefore, the final optimization model can be formulated as

$$\min_{\mathbf{X}} \|\mathbf{XBS} - \mathbf{Y}\|_F^2 + \lambda\left\|\nabla\mathbf{X} - \nabla\widetilde{\mathbf{P}}\right\|_{2,1} + \alpha\|\mathbf{X} - \mathbf{X}_{net}\|_F^2 . \tag{9}$$

The proposed model explores the available spatial and spectral information using the two data fitting terms. Furthermore, with the PDI term viewed as a regularization, the prior information acquired by any DCNN, see Section V for details, through the training phase, is incorporating into the VO approach improving the performance.

The objective function (9) is convex but non-smooth, which fails to generate the derivative with respect to $\mathbf{X}$ as directly as in [37], [44]. Therefore, we propose in the next section a new ADMM-based algorithm to solve it in an efficient way.

IV. The Proposed Algorithm

The proposed model can be solved by exploiting the ADMM framework [46], [47], [48], [49], which separates the Frobenius norm and $\ell_{2,1}$ norm into two independent subproblems that have closed-form and iterative solutions, respectively. In our model, we can rewrite (9) as an equivalent constrained problem through introducing the auxiliary variable $\mathbf{W} = \mathbf{X} - \widetilde{\mathbf{P}}$, namely, we have that

$$\begin{aligned}\min_{\mathbf{X},\mathbf{W}} \|\mathbf{XBS} - \mathbf{Y}\|_F^2 + \lambda\|\nabla\mathbf{W}\|_{2,1} + \alpha\|\mathbf{X} - \mathbf{X}_{net}\|_F^2 \\ s.t.\quad \mathbf{W} = \mathbf{X} - \widetilde{\mathbf{P}}.\end{aligned} \tag{10}$$

The augmented Lagrangian function of the constrained model (10) can be expressed as

$$\begin{aligned}\mathcal{L}_\eta(\mathbf{X}, \mathbf{W}, \mathbf{\Theta}) = \|\mathbf{XBS} - \mathbf{Y}\|_F^2 + \lambda\|\nabla\mathbf{W}\|_{2,1} + \alpha\|\mathbf{X} - \mathbf{X}_{net}\|_F^2 \\ + \frac{\eta}{2}\left\|\mathbf{X} - \widetilde{\mathbf{P}} - \mathbf{W} + \frac{\mathbf{\Theta}}{\eta}\right\|_F^2 + const,\end{aligned} \tag{11}$$

where $\mathbf{\Theta}$ denotes the Lagrange multiplier, $\eta > 0$ is a penalty parameter, and *const* represents a generic constant. Afterwards, the minimization problem (11) can be solved iteratively and alternatively via updating the following three simpler subproblems:

1) The $\mathbf{X}$-subproblem can be accurately updated by solving the following function:

$$\begin{aligned}\mathbf{X}^{k+1} = \arg\min_{\mathbf{X}}\ \|\mathbf{XBS} - \mathbf{Y}\|_F^2 + \alpha\|\mathbf{X} - \mathbf{X}_{net}\|_F^2 \\ + \frac{\eta}{2}\left\|\mathbf{X} - \widetilde{\mathbf{P}} - \mathbf{W}^k + \frac{\mathbf{\Theta}^k}{\eta}\right\|_F^2,\end{aligned} \tag{12}$$

which is a least squares problem and has the following closed-form solution,

$$\mathbf{X}^{k+1} = \mathbf{Z}^k\mathbf{H}^{-1}, \tag{13}$$

where

$$\mathbf{Z}^k = \eta\widetilde{\mathbf{P}} + \eta\mathbf{W}^k - \mathbf{\Theta}^k + 2\mathbf{Y}(\mathbf{BS})^T + 2\alpha\mathbf{X}_{net}, \tag{14}$$

$$\mathbf{H} = 2\mathbf{BS}(\mathbf{BS})^T + (2\alpha + \eta)\mathbf{I}. \tag{15}$$

and $\cdot^T$ is the transpose operator.
It is worth to be remarked that the matrix $\mathbf{BS}(\mathbf{BS})^T$ is semi-positive, and the identity matrix $\mathbf{I}$ is positive, hence, the matrix $\mathbf{H}$ is invertible.

2) The $\mathbf{W}$-subproblem can be updated by minimizing the following problem:

$$\mathbf{W}^{k+1} = \arg\min_{\mathbf{W}} \ \frac{\eta}{2}\left\|\mathbf{X}^{k+1} - \widetilde{\mathbf{P}} - \mathbf{W} + \frac{\mathbf{\Theta}^k}{\eta}\right\|_F^2 + \lambda\left\|\nabla\mathbf{W}\right\|_{2,1}. \tag{16}$$

Defining the following intermediate variable

$$\mathbf{T}^k = \mathbf{X}^{k+1} - \widetilde{\mathbf{P}} + \frac{\mathbf{\Theta}^k}{\eta}, \tag{17}$$

the problem (16) can be rewritten as

$$\mathbf{W}^{k+1} = \arg\min_{\mathbf{W}} \ \frac{\eta}{2}\left\|\mathbf{W} - \mathbf{T}^k\right\|_F^2 + \lambda\left\|\nabla\mathbf{W}\right\|_{2,1}. \tag{18}$$

The $\mathbf{W}$-subproblem (18) is well-known as the vectorial total variation (VTV) denoising problem [50], [51]. Different from the $\mathbf{X}$-subproblem, it has no closed-form solution. For the purpose of solving this subproblem more conveniently, the $\mathbf{W}$, $\mathbf{T}^k$ and $\mathbf{\Theta}^k$ are now reshaped to the tensors $\mathcal{W}$, $\mathcal{T}^k$ and $\mathit{\Theta}^k \in \mathbb{R}^{H\times W\times S}$, respectively, and the $\mathcal{W}$-subproblem can be solved by the VTV denoising algorithm [51] accelerated by the FISTA framework [52]. We follow the previous work [52] using similar symbols without causing confusion.

Assuming three third-order tensors $\mathcal{R} \in \mathbb{R}^{(H-1)\times W\times S}$ and $\mathcal{S} \in \mathbb{R}^{H\times(W-1)\times S}$ and $\mathcal{N} \in \mathbb{R}^{H\times W\times S}$, a linear operator $\mathit{\Gamma}$ with respect to $\mathcal{R}$, $\mathcal{S}$ is defined as

$$\mathit{\Gamma}(\mathcal{R},\mathcal{S})_{i,j,k} = \mathcal{R}_{i,j,k} - \mathcal{R}_{i-1,j,k} + \mathcal{S}_{i,j,k} - \mathcal{S}_{i,j-1,k}\ , \tag{19}$$

where $i = 1,2,\ldots,H$, $j = 1,2,\ldots,W$ and $k = 1,2,\ldots,S$. In particular, all the variables defined in (19) have a zero padding boundary, *e.g.*, $\mathcal{R}_{0,j,k} = 0$ and $\mathcal{R}_{H,j,k} = 0$. The inverse linear operation corresponding to $\mathit{\Gamma}$ is defined as

$$\mathit{\Gamma}^T(\mathcal{N}) := (\mathcal{R},\mathcal{S})\ , \tag{20}$$

where

$$\mathcal{R}_{i,j,k} = \mathcal{N}_{i,j,k} - \mathcal{N}_{i+1,j,k},\ \mathcal{S}_{i,j,k} = \mathcal{N}_{i,j,k} - \mathcal{N}_{i,j+1,k}. \tag{21}$$

In addition, the projection operator $\mathbb{P}$ utilized to force $\sum_{k=1}^{S}\left(\mathcal{R}_{i,j,k}^2 + \mathcal{S}_{i,j,k}^2\right) \le 1$, $|\mathcal{R}_{i,W,k}| \le 1$ and $|\mathcal{S}_{H,j,k}| \le 1$ is employed to constrain the fused image to be in a given set. More details can be found in [53]. Based on the above notations and definitions, we can summarize the solution of the $\mathcal{W}$-subproblem in Algorithm 1. Note that, we need to re-unfold the solution $\mathcal{W}^{k+1}$ into a matrix $\mathbf{W}^{k+1}$ along the third dimension after executing Algorithm 1.

3) According to the ADMM framework, the Lagrangian multiplier $\mathbf{\Theta}$ can be updated by

$$\mathbf{\Theta}^{k+1} = \mathbf{\Theta}^k + \eta\left(\mathbf{X}^{k+1} - \widetilde{\mathbf{P}} - \mathbf{W}^{k+1}\right). \tag{22}$$

Algorithm 1 Algorithm for updating $\mathcal{W}$

```
Input: $(\mathcal{U}^1,\mathcal{V}^1) = (\mathcal{R}^0,\mathcal{S}^0) = (\mathbf{0}_{(H-1)\times W\times S}, \mathbf{0}_{H\times(W-1)\times S})$,
    $\eta$, $\lambda$, $\mathcal{T}^k = \mathcal{X}^{k+1} - \widetilde{\mathcal{P}} + \frac{\mathit{\Theta}^k}{\eta}$, $t^1 = 1$.
for $p = 1$ to $maxitertion$ do
    $(\mathcal{R}^p,\mathcal{S}^p) = \mathbb{P}\left[(\mathcal{U}^p,\mathcal{V}^p) + \frac{\eta}{8\lambda}\mathit{\Gamma}^T\left(\mathcal{T}^k - \frac{\lambda}{\eta}\mathit{\Gamma}(\mathcal{U}^p,\mathcal{V}^p)\right)\right]$
    $t^{p+1} = \frac{1+\sqrt{1+4(t^p)^2}}{2}$
    $(\mathcal{U}^{p+1},\mathcal{V}^{p+1}) = (\mathcal{R}^p,\mathcal{S}^p) + \frac{t^p-1}{t^{p+1}}(\mathcal{R}^p - \mathcal{R}^{p-1}, \mathcal{S}^p - \mathcal{S}^{p-1})$
end for
Output: $\mathcal{W}^{k+1} = \mathcal{T}^k - \frac{\lambda}{\eta}\mathit{\Gamma}(\mathcal{R}^p,\mathcal{S}^p)$
```

Algorithm 2 The ADMM-based algorithm for the proposed model (9).

```
Input: LRMS image $\mathbf{Y}$, PAN image $\mathbf{P}$, proximal image $\mathbf{X}_{net}$,
    $\lambda$, $\alpha$, $\eta$, $r$, $p_{mit}$, $k_{mit}$.
Initialization: $\mathbf{X}^0 = \Phi(\mathbf{Y}, r), \mathbf{W}^0 = \mathbf{\Theta}^0 = \mathbf{0}$
while $relcha > \varepsilon$ and $k < k_{mit}$ do
    Generate $\widetilde{\mathbf{P}}$ via (5) - (6).
    Update $\mathbf{X}$ via (13) - (15).
    Update $\mathbf{W}$ via Algorithm 1.
    Update Lagrange multiplier $\mathbf{\Theta}$ via (22).
end while
Output: Fused HRMS image $\mathbf{X}$
```

The stopping criterion of the proposed algorithm is based on the relative change ($relcha$) between two successive fused images. In particular, this latter should be less than a tolerance value, ε, *i.e.*,

$$relcha = \left\|\mathbf{X}^{k+1} - \mathbf{X}^k\right\|_F / \left\|\mathbf{X}^k\right\|_F < \varepsilon. \tag{23}$$

The proposed algorithm to solve the problem in (9) is summarized in Algorithm 2. The convergence of the proposed iterative approach is guaranteed [46]. In Algorithm 2, p_{mit} and k_{mit} are the maximum iterations of the inner and the outer layers, and Φ indicates the upsampling operation using bicubic interpolation.

V. Experimental Results

This section is devoted to the comparison between the proposed method and some state-of-the-art approaches using several datasets acquired by different sensors. A true color representation is selected for the qualitative analysis of the fused results. Furthermore, all the methods are run in MATLAB (R2016a) on a computer of 16Gb RAM and Intel(R) Core(TM) i5-4590 CPU: @3.30 GHz.

In order to assess the quality of the different methods, some popular quality indexes are adopted. In particular, the dimensionless global error in synthesis (ERGAS) index [54], the spectral angle mapper (SAM) index [55], the Q2^n (Q8 for 8-band datasets and Q4 for 4-band datasets) index [56], the peak signal-to-noise ratio (PSNR), and the structural similarity index (SSIM) [57] have been selected. According to the statistics in [16], the first three metrics are the most widely used in pansharpening studies. However, even the PSNR and SSIM are widely used to evaluate the similarity between two images in the image processing literature. The ideal values for SAM and ERGAS are 0, for Q2^n and SSIM are 1, whereas $+\infty$ for PSNR. Furthermore, the scale factors for all the datasets are 4, *i.e.*, $r = 4$ and the tolerance value is set to $\varepsilon = 2 \times 10^{-4}$. A deeper discussion on the parameters tuning can be found in Section V-E.

In order to have a more compact experimental analysis, we need to fix a specific DCNN approach that generates the $\mathcal{X}_{net}$. Many state-of-the-art DCNN-based methods for pansharpening can be found, *e.g.* [2], [28], [29], [30]. In our experiments, the PanNet proposed by Yang *et al.* [2] is selected to determinate the PDI term. More DCNN options are discussed in Section V-E.

A. Datasets

In our experiments, different datasets are exploited to demonstrate the superiority of the proposed approach. Some datasets are freely available at the website [1]. The exploited datasets are: i) the Rio and the Tripoli datasets both captured by the WorldView-3 sensor with an MS image with 8 spectral bands and a PAN image at spatial resolution of 0.3 m, ii) the WashingtonDC and the Stockholm datasets acquired by the WorldView-2 sensor including an MS image with 8 spectral bands and a PAN image with a spatial resolution of 0.4 m, iii) the Toulouse dataset acquired by the IKONOS sensor over the city of Toulouse (France) with an MS image consisting of 4 spectral bands and a PAN image at spatial resolution of 1.0 m, and iv) the *Pléiades2 dataset* acquired by the Pl$\acute{e}$iades sensor with an MS image consisting of 4 spectral bands and a PAN with a spatial resolution of 0.5 m.

B. Benchmark

The following state-of-the-art methods are used for comparison:

- EXP: MS image interpolation, using a polynomial kernel with 23 coefficients [58].
- MTF-GLP-HPM: modulation transfer function - generalized laplacian pyramid with high pass modulation injection model [1].
- GLP-Reg-FS: modulation transfer function - generalized laplacian pyramid and a new full resolution regression-based injection model [25].
- CVPR19: variational pansharpening approach with local gradient constraint [4].
- DiCNN: pansharpening method via detail injection-based convolutional neural networks [30].
- PanNet: a DCNN method for pansharpening [2].

It is worth to be remarked that the source codes of the approaches into the benchmark are available either at the website [2] or the authors' homepages. In order to have a fair comparison, we adjust the parameters of these methods to get their best performance.

C. The Reduced Resolution Assessment

Reduced resolution (simulated) data are obtained according to Wald's protocol [59], *i.e.* by filtering with filters designed to match the MS sensor's MTFs and decimation [23], [60]. We consider the degraded PAN image as the input PAN image, similarly, the original LRMS and its degraded version play the role of the reference (ground-truth) HRMS image and the input LRMS image, respectively.

1) The Reduced Resolution Rio Dataset: In this experiment, λ, α, and η are empirically set to 0.011, 0.50, and 0.1, respectively. Fig. 4 shows the visual results and the related mean absolute error (MAE) maps for all the methods. In Fig. 4, we clearly observe that our approach achieves excellent performance. The traditional methods, *i.e.*, MTF-GLP-HPM, GLP-Reg-FS, and CVPR19, preserve the spectral information, but loosing many spatial details. Although DiCNN and PanNet achieve excellent results, our approach shows its superiority obtaining a better residual map. This analysis is corroborated by the calculation of the quality indexes reported in Table I-(a). Our method clearly obtains the best results for all the quality indexes.

2) The Reduced Resolution Tripoli Dataset: In this experiment, the reduced resolution (simulated) Tripoli dataset is employed to further verify the superiority of our model on WorldView-3 data, the same parameters as for the reduced resolution Rio experiment are exploited. Table I-(b) summarizes the quantitative results showing the best performance of the proposed approach whatever the quality index. The average running time is also reported in Table I. The MTF-GLP-HPM and the GLP-Reg-FS are classical approaches, thus showing a limited computational burden. Even the PanNet and the DiCNN, belonging to the ML class, have a reduced computational effort during the test phase. However, our approach compared to another one in the same (VO) class, *i.e.*, the CVPR19, shows a reduced running time. Thus, this analysis, computed on a patch of 256×256 pixels, demonstrates that the computational burden of the proposed approach can be considered acceptable for addressing a real image fusion problem.

3) The Reduced Resolution WashingtonDC Dataset: In this experiment, λ, α and η are empirically set to 0.003, 0.43, and 0.1, respectively. Fig. 5 shows the visual performance and the corresponding MAE maps. It is clear that the proposed method preserves details in a better way than the other methods, which have evident spatial blurring. From the MAE maps, we can find that only our method restores the vertical stripe structure. In Table II-(a), the quality indexes are calculated. Our method still outperforms the other approaches into the benchmark.

4) The Reduced Resolution Stockholm Dataset: In this experiment, the set of the parameters is slightly adjusted. In particular, λ and α are equal to 0.012 and $= 0.44$, respectively. Instead, η is the same as in the previous test case. The performance of all the methods are reported in Table II-(b). Our method shows its clear superiority with respect to all the other techniques, in particular, even compared to the PanNet and the DiCNN. Finally, the computational analysis in Table II is in line with the previously obtained results.

Finally, the MAEs calculated for each spectral bands of the fused products obtained by the methods into the benchmark are depicted in Fig. 6 for the four simulated test cases. The horizontal lines indicate the average along the spectral

[1] http://www.digitalglobe.com/samples?search=Imagery

[2] http://openremotesensing.net/kb/codes/pansharpening/

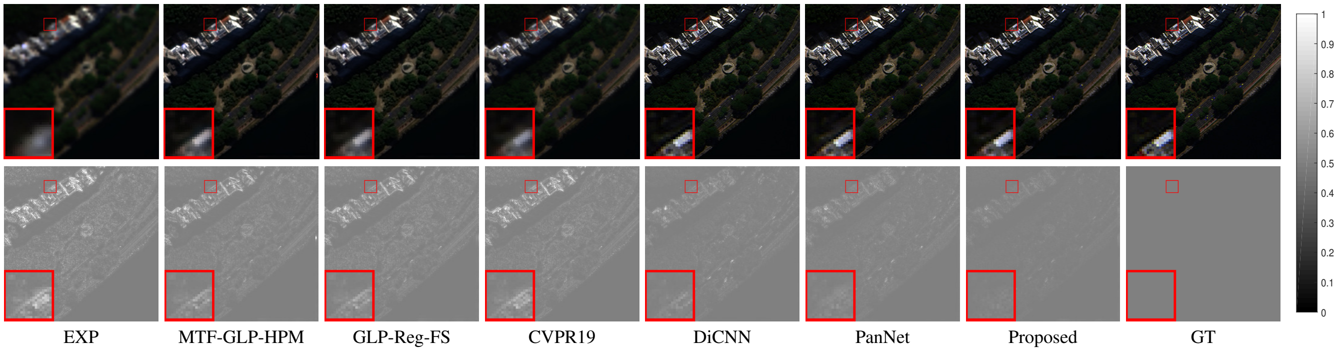


Fig. 4. The fusion results on the reduced resolution (simulated) Rio dataset (source: WorldView-3). Top row: the visual performance of the EXP, MTF-GLP-HPM, GLP-Reg-FS, CVPR19, DiCNN, PanNet, Proposed method, and the ground-truth (GT) image, respectively. Bottom row: the corresponding MAE maps using the GT image as reference. For a better visualization, we doubled the intensities of the MAE maps and added 0.5.

TABLE I
QUALITY METRICS FOR ALL THE COMPARED APPROACHES ON THE REDUCED RESOLUTION (SIMULATED) RIO AND TRIPOLI DATASETS, RESPECTIVELY. (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	*(a) Rio*					*(b) Tripoli*					Average time(s)
	ERGAS	*SAM*	*Q8*	*PSNR*	*SSIM*	*ERGAS*	*SAM*	*Q8*	*PSNR*	*SSIM*	
EXP	8.9481	6.3662	0.6775	29.3264	0.7872	4.8097	4.1088	0.8173	27.7309	0.7104	**0.03**
MTF-GLP-HPM	6.3748	5.7967	0.7996	31.0209	0.8867	2.9489	3.9725	0.9301	31.9819	0.8876	0.29
GLP-Reg-FS	6.6462	6.2046	0.7867	31.7901	0.8746	2.9339	3.8925	0.9312	31.9857	0.8900	0.38
CVPR19	7.0125	5.6664	0.7681	31.4024	0.8701	3.5465	3.8256	0.9017	30.4052	0.8515	20.26
DiCNN	3.8738	4.2262	0.8600	36.4007	0.9526	2.0734	3.1568	0.9649	34.8528	0.9390	0.22
PanNet	3.8588	4.4518	0.8643	36.3447	0.9529	2.1272	3.1887	0.9648	34.5557	0.9367	0.41
Proposed	**3.6266**	**4.0229**	**0.8718**	**36.7143**	**0.9556**	**1.9491**	**3.0014**	**0.9694**	**35.1881**	**0.9413**	4.22
Ideal value	**0**	**0**	**1**	**$+\infty$**	**1**	**0**	**0**	**1**	**$+\infty$**	**1**	-

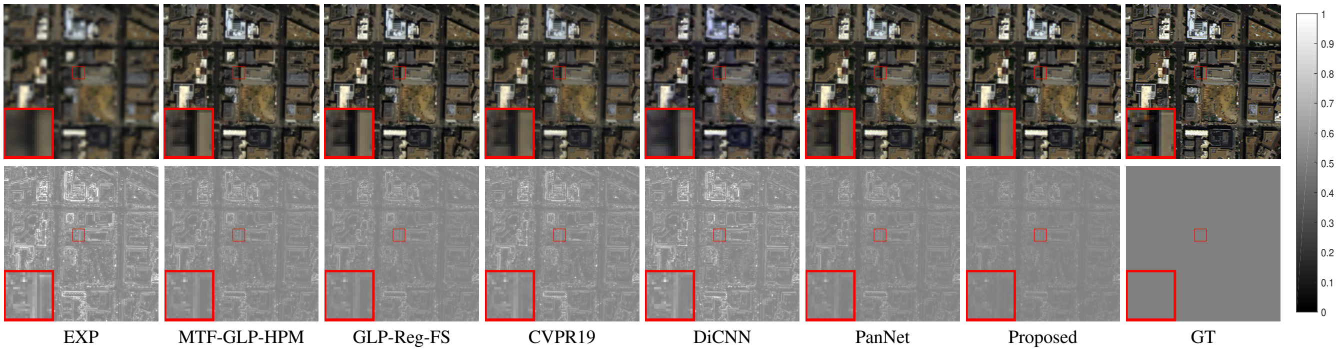


Fig. 5. The fusion results on the reduced resolution (simulated) WashingtonDC dataset (source: WorldView-2). Top row: the visual performance of the EXP, MTF-GLP-HPM, GLP-Reg-FS, CVPR19, DiCNN, PanNet, Proposed method, and the ground-truth (GT) image, respectively. Bottom row: the corresponding MAE maps using the GT image as reference. For a better visualization, we doubled the intensities of the MAE maps and added 0.5.

dimension of the MAE values for each method. The lower the value, the better the performance. For all the test cases, it is clear to see that the proposed approach gets the best results, *i.e.* the smallest average of the band-dependent MAEs indicating a fused image close to the reference (GT) image.

D. The Full Resolution Assessment

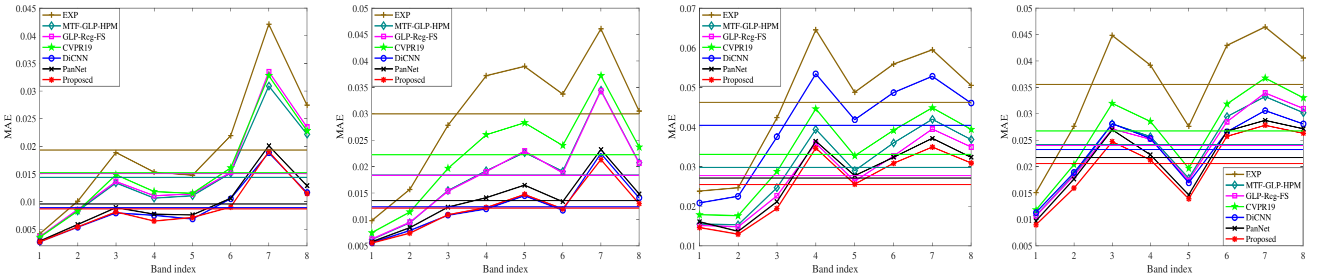
To corroborate the results obtained at reduced resolution, the proposed model is tested on real dataset at full resolution, *i.e.* without degrading the original MS and PAN data according to Wald's protocol. For these experiments, the full resolution Tripoli and Stockholm datasets are employed. Unlike reduced resolution (simulated) experiments, where a reference (GT) image is available, in this case, no reference can be exploited to assess the performance. Fortunately, metrics without reference have been proposed in the related literature. In particular, the *quality with no reference (QNR)* [61] consisting of a spectral distortion index D_λ and a spatial distortion index D_s is often adopted to this aim.

1) The Full Resolution Tripoli Dataset: In this experiment, we use the same parameters as for the reduced resolution Rio experiment, *i.e.*, $\lambda = 0.011$, $\alpha = 0.50$ and $\eta = 0.1$. Fig. 7 shows the fusion results and the MAE maps comparing the degraded version (based on the MTF filtering [60]) of the fused image with the original LRMS image in order to have a quick look about the consistency of the fusion products. In Fig. 7, we can note that the proposed model achieves similar details compared to the PAN ones. Furthermore, from the

TABLE II

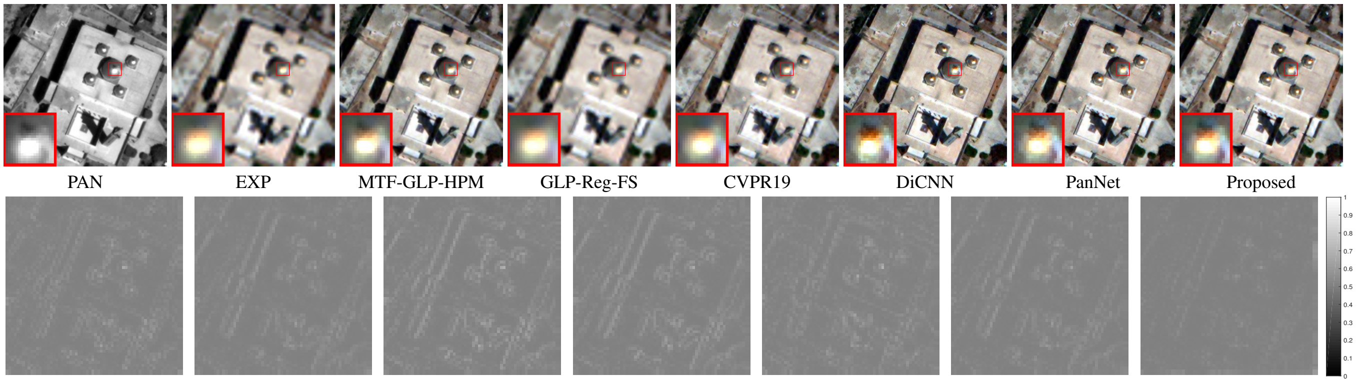
QUALITY METRICS FOR ALL THE COMPARED APPROACHES ON THE REDUCED RESOLUTION (SIMULATED) WASHINGTONDC AND THE STOCKHOLM DATASETS, RESPECTIVELY. (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	*(a) WashingtonDC*					*(b) Stockholm*					Average
	ERGAS	*SAM*	*Q8*	*PSNR*	*SSIM*	*ERGAS*	*SAM*	*Q8*	*PSNR*	*SSIM*	time(s)
EXP	6.7945	6.6693	0.7305	24.2532	0.5367	10.5392	7.6555	0.6536	24.8813	0.6398	**0.03**
MTF-GLP-HPM	4.4810	6.1534	0.8950	27.9970	0.8059	7.3327	6.8871	0.8339	27.8446	0.8338	0.23
GLP-Reg-FS	4.2127	5.8836	0.9090	28.5064	0.8321	7.3569	7.0824	0.8399	28.2198	0.8392	0.26
CVPR19	4.9324	5.9979	0.8651	27.1235	0.7688	8.1130	6.9643	0.7939	27.2945	0.7987	20.17
DiCNN	5.9310	7.0829	0.8200	25.4925	0.6852	6.7927	6.9590	0.8586	28.6846	0.8608	0.22
PanNet	4.0688	5.4871	0.9138	28.7873	0.8415	6.4703	6.4899	0.8657	29.1310	0.8700	0.41
Proposed	**3.8175**	**5.2298**	**0.9260**	**29.4152**	**0.8562**	**6.2653**	**6.2571**	**0.9062**	**29.6430**	**0.8776**	4.08
Ideal value	**0**	**0**	**1**	**$+\infty$**	**1**	**0**	**0**	**1**	**$+\infty$**	**1**	-



(a) Rio experiment (b) Tripoli experiment (c) WashingtonDC experiment (d) Stockholm experiment

Fig. 6. The spectral MAE curves for all the methods on the reduced resolution (simulated) datasets: (a) Rio dataset, (b) Tripoli dataset, (c) WashingtonDC dataset, and (d) Stockholm dataset. The horizontal lines represent the average value of the MAEs calculated for all the spectral bands. The lower the value, the better the performance.



PAN EXP MTF-GLP-HPM GLP-Reg-FS CVPR19 DiCNN PanNet Proposed

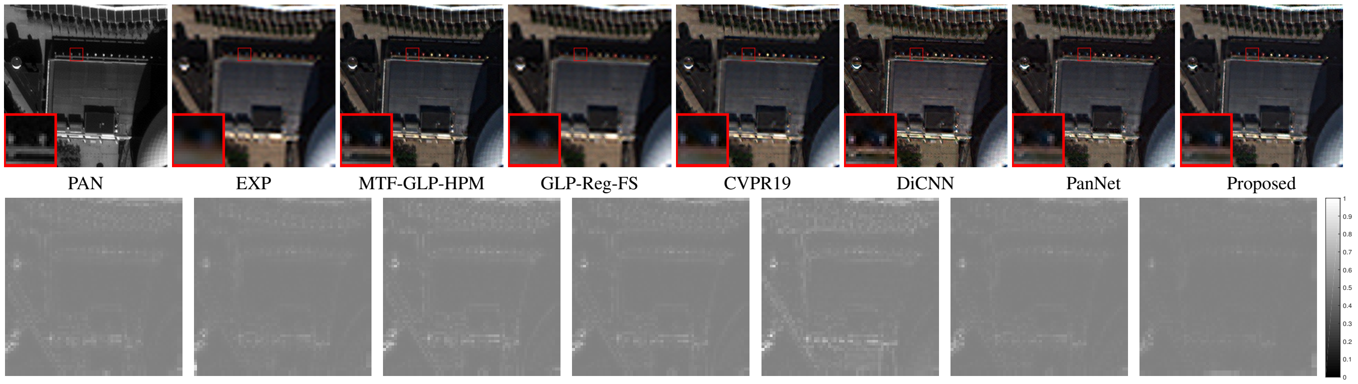
(a) EXP (b) MTF-GLP-HPM (c) GLP-Reg-FS (d) CVPR19 (e) DiCNN (f) PanNet (g) Proposed

Fig. 7. The fusion results on the full resolution Tripoli dataset (source: WorldView-3). Top row: the PAN image, the visual performance of all the methods: (a) EXP, (b) MTF-GLP-HPM, (c) GLP-Reg-FS, (d) CVPR19, (e) DiCNN, (f) PanNet and (g) Proposed method. Bottom row: the MAE maps between the degraded fusion outcome (by filtering with MTF-based filters and decimation) and the LRMS image. For a better visualization, we doubled the intensities of the MAE maps and added 0.5.

MAEs in Figs. 7(a)-(g), we can note a good consistency of our approach with respect to the compared methods. The metrics D_λ, D_s and QNR are reported in Table III-(a). We can remark that our method yields the best performance in terms of D_s and QNR metrics making an excellent trade-off between the spectral fidelity and the spatial enhancement.

2) The Full Resolution WashingtonDC Dataset: In this experiment, we slightly tune the parameters λ, α and η. Thus, the used values are $\lambda = 0.025$, $\alpha = 0.57$ and $\eta = 0.1$. The visual performance and corresponding MAE maps are depicted in Fig. 8. Our method together with the MTF-GLP-HPM obtain the lowest values in the related MAE maps, thus getting generally good performance from a spectral point of view. The CVPR19 and the GLP-Reg-FS outcomes are highly distorted from a spatial point of view. The DiCNN and PanNet outcomes suffer from both spatial and spectral distortions. The quality indexes at full resolution are reported in Tab III-(b). The GLP-Reg-FS gets the best D_λ index, whereas the proposed method has the best performance in terms of D_s and the overall QNR index. This means that the proposed approach has comprehensive advantages with respect to the benchmark.

In summary, the broad experiments on datasets both at reduced resolution (simulated) and at full resolution show that the proposed method gets very high performance, both qualitatively and quantitatively, with respect to the benchmark. In particular, for the Rio and the Tripoli datasets, although the



(a) EXP (b) MTF-GLP-HPM (c) GLP-Reg-FS (d) CVPR19 (e) DiCNN (f) PanNet (g) Proposed

Fig. 8. The fusion results on the full resolution Stockholm dataset (source: WorldView-2). Top row: the PAN image, the visual performance of all the methods: (a) EXP, (b) MTF-GLP-HPM, (c) GLP-Reg-FS, (d) CVPR19, (e) DiCNN, (f) PanNet and (g) Proposed method. Bottom row: the MAE maps between the degraded fusion outcome (by filtering with MTF-based filters and decimation) and the LRMS image. For a better visualization, we doubled the intensities of the MAE maps and added 0.5.

TABLE III
QUALITY METRICS ON THE FULL RESOLUTION (A) TRIPOLI AND (B) STOCKHOLM DATASETS. (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	*(a) Real Tripoli*			*(b) Real Stockholm*			Average
	D_λ	D_s	QNR	D_λ	D_s	QNR	time(s)
EXP	**0.0008**	0.0332	0.9660	0.0042	0.0902	0.9060	**0.03**
MTF-GLP-HPM	0.0130	0.0194	0.9678	0.0213	0.0216	0.9576	0.26
GLP-Reg-FS	**0.0008**	0.0320	0.9673	**0.0018**	0.0582	0.9401	0.25
CVPR19	0.0031	0.0230	0.9740	0.0103	0.0266	0.9634	20.10
DiCNN	0.0065	0.0167	0.9770	0.0283	0.0364	0.9363	0.22
PanNet	0.0039	0.0084	0.9878	0.0186	0.0155	0.9661	0.41
Proposed	0.0037	**0.0048**	**0.9915**	0.0135	**0.0060**	**0.9806**	4.93
Ideal value	0	0	1	0	0	1	-

PanNet method achieves very high performance, the proposed approach is still able to improve them. Instead for the WashingtonDC and the Stockholm datasets, where the performance of the PanNet is not outstanding, high performance of the proposed framework is still obtained. It is worth to be remarked that the fine-tuning of the parameters of the proposed method on the WorldView-3 datasets can lead to better results, but a fixed set of parameters has been selected in order to show the robustness of the proposed method.

E. Discussion

This section is devoted to some discussions about the proposed approach, such as the parameter tuning and the ablation study. For the sake of brevity, all the discussions (except for the generalization analysis) are related to the reduced resolution Rio dataset exploiting the main quality metrics, *i.e.*, the EGRAS, the SAM, and the overall quality index Q8.

1) The Parameter Tuning: Five parameters are related to our approach, *i.e.*, λ, α, η, p_{mit} and k_{mit}. In particular, k_{mit} is used to prevent the algorithm from iterating endlessly, thus it can be appropriately set to a high number. Fig. 9 depicts the performance varying the other parameters. All the parameters are fixed except for the parameter that we want to analyze. Figs. 9(a)-(b) show that only slight changes can happen varying both λ and α demonstrating the robustness of our method with respect to these parameters. Fig. 9(c) shows that the proposed method is almost insensitive to the penalty parameter η in the adopted range of values $10^{-2} \leq \eta \leq 10^{1}$. Finally, Fig. 9(d) depicts a sharply convergence by increasing the number of the inner iterations p. However, the higher the number of iterations, the higher computational burden. For this reason, we set p_{mit} to 10 in our study considering the balance between computation and performance.

2) The Ablation Study: In order to have deeper insights about our approach, we conducted an ablation study on the proposed model generating the following three corresponding sub-models,

$$\min_{\mathbf{X}} \left\| \nabla \widetilde{\mathbf{P}} - \nabla \mathbf{X} \right\|_{2,1} + \alpha \left\| \mathbf{X}_{net} - \mathbf{X} \right\|_F^2 , \tag{24}$$

$$\min_{\mathbf{X}} \left\| \mathbf{XBS} - \mathbf{Y} \right\|_F^2 + \alpha \left\| \mathbf{X}_{net} - \mathbf{X} \right\|_F^2 , \tag{25}$$

and

$$\min_{\mathbf{X}} \left\| \mathbf{XBS} - \mathbf{Y} \right\|_F^2 + \lambda \left\| \nabla \widetilde{\mathbf{P}} - \nabla \mathbf{X} \right\|_{2,1} . \tag{26}$$

Based on these models, we design the algorithm, again, performing the tests using optimal parameters for a fair comparison. The performance are reported in Table IV. We can observe that the model (25) is the second best approach (after the proposed one), even obtaining optimal spectral features. Furthermore, the fusion performance of the model (26) is competitive to that of the traditional methods in Tab I-(a). Finally, the advantages in using the PDI term can be also assessed by having a look at the performance of the PanNet, which is used to determinate the PDI term of the proposed approach.

3) The Generalization Analysis: In order to assess the ability of generalization of the proposed approach, we performed additional experiments using both an IKONOS dataset and a Pléiades dataset. In these tests, the PanNet is trained on the 2-nd (blue), the 3-rd (green), the 5-th (red) and the 7-th (near infra-red) bands of a WorldView-3 dataset, thus considering the spectral responses of both the IKONOS and the Pléiades sensors and the WorldView-3 ones. The fusion results are depicted in Fig. 10 and the quantitative outcomes are reported in Table V. It is clear to show that the proposed method achieves the best performance in all the cases and by using all the indexes. Thus, the generalization ability of the proposed approach is clearly demonstrated with respect to an approach based on a training by example philosophy as the PanNet (*i.e.*, a DCNN).

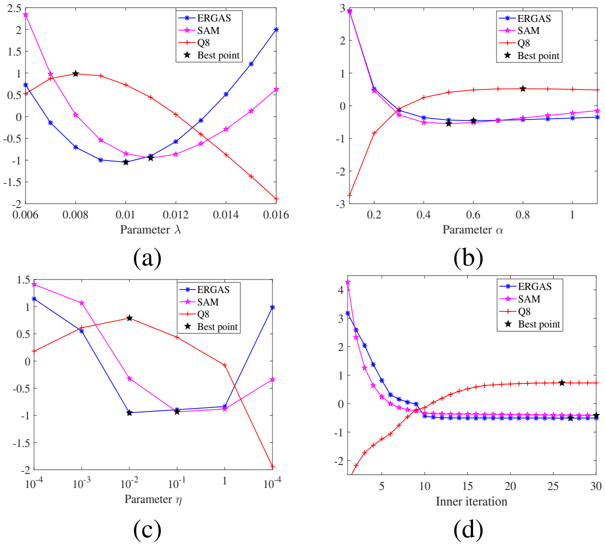
4) The GISC Effect: In order to assess the improvement in using the proposed GISC, we show in Fig. 12 the curves over the outer iterations of the quality indexes ERGAS, SAM, and Q8 using $\widetilde{\mathcal{P}}$ and $\mathcal{P}$, *i.e.*, with and without GISC, respectively. The advantages in using the proposed approach based on $\widetilde{\mathcal{P}}$ are clear. Thus, the rationale of using the proposed term relied upon $\widetilde{\mathcal{P}}$ is experimentally proved.

5) The $\mathbf{X}_{net}$ Generalization: In this section, the output of the PanNet has been used to feed the PDI term showing high performance. However, all the fused results coming from ML approaches can be theoretically used in the proposed VO framework. This is based on the idea that in the output of an ML-based approach the prior knowledge learned during the training phase is present. Thus, this latter can be exploited by the proposed approach. In order to corroborate the last statement, the proposed VO framework is also used with the output of the DiCNN method. We call this fusion result *DiCNN-based* from hereon. The quantitative results are reported in Table VI showing that both the approaches based on two different PDI (regularization) terms achieve higher performance than the DCNN-based ones. Furthermore, the two outcomes obtained by feeding our framework with different prior information have comparable performance. This proves the possibility to generalize with respect to various proximal DCNNs.

6) The Algorithm Relation: It is worth to be noted that the sub-problem of $\mathbf{X}$ in Section IV can be encoded based on diagonalizing matrix $\mathbf{B}$ and exploiting the characteristics of $\mathbf{S}$ for higher computational efficiency [62], [63], [64]. In [41], Wei *et al.* proposed a new methodology, named FUSE [3], via analytically and efficiently solving a Sylvester equation. Starting from this consideration, a comparison with the approach in [41] should be performed both from a performance and a computational points of view. Fig. 11 and Tab VII report the results. As can be seen from the pan-sharpened products and the quality indexes, the fused image provided by FUSE demonstrates a better quality than almost all the traditional methods. However, some flaws, e.g. a clear spectral distortion which may be caused by a biased estimate of the spectral response function, are evident. Therefore, our scheme is still very promising.

TABLE IV
QUALITY METRICS OF THE DIFFERENT MODELS ON THE REDUCED RESOLUTION (SIMULATED) RIO DATASET. (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	*ERGAS*	*SAM*	*Q8*
Model (24)	3.8396	4.3530	0.8661
Model (25)	3.7174	4.2823	0.8708
Model (26)	6.5851	5.6281	0.7945
PanNet	3.8588	4.4518	0.8643
Proposed	**3.6266**	**4.0229**	**0.8716**
Ideal value	**0**	**0**	**1**



(a)

(b)

(c)

(d)

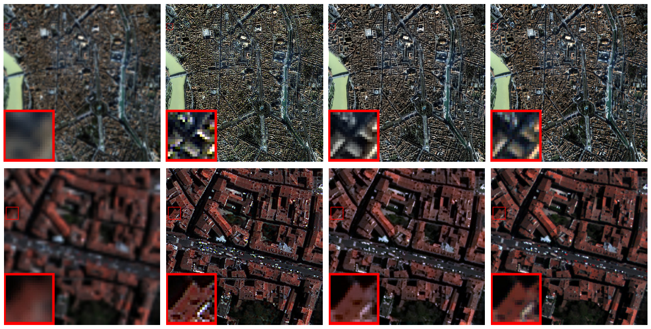
Fig. 9. The ERGAS, SAM, Q8 curves for: the regularization parameters (a) λ and (b) α, (c) the penalty parameter η and (d) the inner iteration p_{mit}. The best points are pointed out with a black star. Note that for better comparisons, we process the obtained indexes by $(index - \mathcal{M}(index))/\mathcal{S}(index)$, where $\mathcal{M}(\cdot)$ and $\mathcal{S}(\cdot)$ represent the mean and standard deviation operations, respectively. Besides, the real mean and standard deviation for ERGAS, SAM and Q8 are (a) 3.6479 ± 0.0156; 4.0454 ± 0.0131; 0.8714 ± 0.0021; (b) 3.7763 ± 0.3106; 4.1580 ± 0.2309; 0.8667 ± 0.0124; (c) 3.6747 ± 0.0392; 4.1331 ± 0.1090; 0.8706 ± 0.0027; (d) 3.6400 ± 0.0008; 4.0346 ± 0.0094; 0.8720 ± 0.0004; respectively.

VI. Conclusions

In this paper, we have proposed a variational model requiring a regularization based on the proposed PDI term in order to address the pansharpening problem. In particular, the novelty of our model is that we make a link between VO approaches and DCNN methods, thanks to the formulation of the PDI term. Furthermore, the use of MTF-based filters and a new RePAN image exploiting GISC have been introduced in this paper to improve the spatial fidelity data fitting term. An efficient ADMM-based algorithm has been developed to solve the proposed problem guaranteeing the convergence to a global optimum. A broad experimental analysis conducted on several acknowledged datasets both at reduced resolution and at full resolution has demonstrated the superiority of the proposed approach with respect to a benchmark consisting of several state-of-the-art approaches. In particular, the robustness with respect to the selection of the algorithm's parameters and the generalization abilities of the proposed VO framework have been pointed out together with a detailed ablation study.

Future developments go towards the extension of the proposed approach to the hyperspectral image superresolution problem and the development of techniques to determinate regularization parameters, thus reducing the complexity of the parameters setting phase and, at the same time, improving the performance in real practical cases.

[3] http://dobigeon.perso.enseeiht.fr/publis.html



(a) EXP (b) PanNet (c) Proposed (d) GT

Fig. 10. Top row: the fusion results on the reduced resolution (simulated) Toulouse dataset (source: IKONOS). Bottom row: the fusion results on the reduced resolution (simulated) Pléiades2 dataset (source: Pléiades).

TABLE V

QUALITY METRICS ON THE REDUCED RESOLUTION (SIMULATED) TOULOUSE AND PLÉIADES2 DATASETS. (BOLD: BEST; UNDERLINE: SECOND BEST)

Dataset	PAN size	Method	*ERGAS*	*SAM*	*Q4*
Toulouse	512 × 512	PanNet	5.8152	5.5342	0.7180
		Proposed	**3.1948**	**3.5488**	**0.8943**
Pléiades2	256 × 256	PanNet	5.5887	5.9802	0.8417
		Proposed	**2.6649**	**4.0249**	**0.9522**
Ideal value			**0**	**0**	**1**

TABLE VI

QUALITY METRICS COMPARING THE PROPOSED APPROACH WITH TWO DIFFERENT PDI TERMS ON THE REDUCED RESOLUTION (SIMULATED) RIO DATASET. (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	*ERGAS*	*SAM*	*Q8*
PanNet	3.8588	4.4518	0.8643
DiCNN	3.8738	4.2262	0.8600
PanNet-based	**3.6266**	4.0229	0.8716
DiCNN-based	3.6267	**4.0177**	**0.8743**
Ideal value	**0**	**0**	**1**

TABLE VII

QUALITY METRICS AND RUN TIMES OF THE DIFFERENT APPROACHES ON THE REDUCED RESOLUTION (SIMULATED) RIO DATASET. (BOLD: BEST; UNDERLINE: SECOND BEST)

Method	*ERGAS*	*SAM*	*Q8*	Time(s)
FUSE	5.2765	6.8140	0.7918	**0.0003**
Proposed	**3.6266**	**4.0229**	**0.8716**	4.2095
Ideal value	**0**	**0**	**1**	-

REFERENCES

[1] 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 *2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*. IEEE, 2003, pp. 90–94.

[2] 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 *IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 5449–5457.

[3] H. F. Shen, X. C. Meng, and L. P. Zhang, "An integrated framework for the spatio-temporal-spectral fusion of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7135–7148, 2016.

[4] X. Y. Fu, Z. H. Lin, Y. Huang, and X. H. Ding, "A variational pan-sharpening with local gradient constraints," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 10265–10274.

[5] L. J. Deng, W. H. Guo, and T. Z. Huang, "Single-image super-resolution via an iterative reproducing kernel Hilbert space method," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 11, pp. 2001–2014, 2015.

[6] M. Imani, "Band dependent spatial details injection based on collaborative representation for pansharpening," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 12, pp. 4994–5004, 2018.

[7] R. W. Dian, S. T. Li, L. Y. Fang, T. Lu, and J. M. Bioucas-Dias, "Nonlocal sparse tensor factorization for semiblind hyperspectral and multispectral image fusion," *IEEE Transactions on Cybernetics*, 2019.

[8] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4381–4398, 2020.

[9] P. Chavez, S. C. Sides, J. A. Anderson, et al., "Comparison of three different methods to merge multiresolution and multispectral data-landsat TM and SPOT panchromatic," *Photogrammetric Engineering and Remote Sensing*, vol. 57, no. 3, pp. 295–303, 1991.

[10] 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," *Information Fusion*, vol. 52, pp. 76–89, 2019.

[11] Z. Y. Zhang, T. Z. Huang, L. J. Deng, J. Huang, and H. X. Dou, "Pansharpening via RoG-based filtering," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 2790–2793.

[12] J. Liu, L. J. Deng, F. M. Fang, and T. Y. Zeng, "A Rudin-Osher-Fatemi model-based pansharpening approach using RKHS and AHF representation," *East Asian Journal on Applied Mathematics*, vol. 9, no. 1, pp. 13–27, 2019.

[13] 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 Transactions on Image Processing*, vol. 27, no. 9, pp. 4330–4344, 2018.

[14] Y. Yang, L. Wu, S. Y. Huang, Y. J. Tang, and W. G. Wan, "Pansharpening for multiband images with adaptive spectral–intensity modulation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 9, pp. 3196–3208, 2018.

[15] Y. Yang, L. Wu, S. Y. Huang, W. G. Wan, W. Tu, and H. Y. Lu, "Multiband remote sensing image pansharpening based on dual-injection model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1888–1904, 2020.

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

[17] H. F. Shen, M. H. Jiang, J. Li, Q. Q. Yuan, Y. C. Wei, and L. P. Zhang, "Spatial-spectral fusion by combining deep learning and variational model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, pp. 6169–6181, 2019.

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

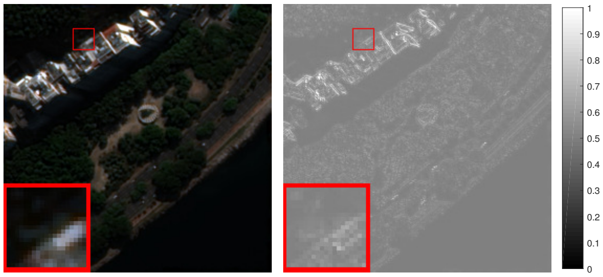
[19] P. Kwarteng and A. Chavez, "Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis," *Photogrammetric Engineering and Remote Sensing*, vol. 55, no. 1, pp. 339–348, 1989.

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

[21] 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.

[22] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 295–309, 2010.

[23] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli,



(a) FUSE

(b) Error

Fig. 11. The fusion results on the reduced resolution (simulated) Rio dataset (source: WorldView-3). (a) The FUSE pansharpened outcome; (b) the MAE map between the FUSE outcome and the GT image shown in Fig. 4.

G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565–2586, 2014.

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

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

[26] C. Chen, Y. Q. Li, W. Liu, and J. Z. Huang, "SIRF: Simultaneous satellite image registration and fusion in a unified framework," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4213–4224, 2015.

[27] K. Zhang, W. M. Zuo, S. H. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3929–3938.

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

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

[30] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 4, pp. 1188–1204, 2019.

[31] L. He, J. W. Zhu, J. Li, A. Plaza, J. Chanussot, and B. Li, "Hyperpnn: Hyperspectral pansharpening via spectrally predictive convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 8, pp. 3092–3100, 2019.

[32] Y. C. Wei, Q. Q. Yuan, H. F. Shen, and L. P. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1795–1799, 2017.

[33] Y. Y. Jiang, X. H. Ding, D. L. Zeng, Y. Huang, and J. Paisley, "Pansharpening with a hyper-laplacian penalty," in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 540–548.

[34] P. F. Liu, L. Xiao, and T. Li, "A variational pan-sharpening method based on spatial fractional-order geometry and spectral–spatial low-rank priors," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 3, pp. 1788–1802, 2017.

[35] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[36] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 1, pp. 228–236, 2007.

[37] R. W. Dian, S. T. Li, A. J. Guo, and L. Y. Fang, "Deep hyperspectral image sharpening," *IEEE Transactions on Neural Networks and Learning Systems*, , no. 99, pp. 1–11, 2018.

[38] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2014.

[39] M. K. Ng, R. H. Chan, and W. C. Tang, "A fast algorithm for deblurring models with Neumann boundary conditions," *SIAM Journal on Scientific Computing*, vol. 21, no. 3, pp. 851–866, 1999.

[40] Y. L. Wang, J. F. Yang, W. T. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[41] Q. Wei, N. Dobigeon, and J. Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4109–4121, 2015.

[42] G. Vivone, P. Addesso, R. Restaino, M. Dalla Mura, and J. Chanussot, "Pansharpening based on deconvolution for multi-band filter estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 540–553, 2019.

[43] G. Vivone, M. Simoes, M. Dalla Mura, R. Restaino, J. M. Bioucas Dias, G. Licciardi, and J. Chanussot, "Pansharpening based on semiblind deconvolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 1997–2010, 2015.

[44] W. Y. Xie, J. Lei, Y. H. Cui, Y. S. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1529–1543, 2019.

[45] W. S. Dong, F. Z. Fu, G. M. Shi, X. Cao, J. J. Wu, G. Y. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2337–2352, 2016.

[46] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[47] J. H. Yang, X. L. Zhao, T. H. Ma, Y. Chen, T. Z. Huang, and M. Ding, "Remote sensing images destriping using unidirectional hybrid total variation and nonconvex low-rank regularization," *Journal of Computational and Applied Mathematics*, vol. 363, pp. 124–144, 2020.

[48] J. Huang, T. Z. Huang, X. L. Zhao, and L. J. Deng, "Joint-sparse-blocks regression for total variation regularized hyperspectral unmixing," *IEEE Access*, vol. 7, pp. 138779–138791, 2019.

[49] X. Li, J. Huang, L. J. Deng, and T. Z. Huang, "Bilateral filter based total variation regularization for sparse hyperspectral image unmixing," *Information Sciences*, vol. 504, pp. 334–353, 2019.

[50] C. Chen, Y. Q. Li, W. Liu, and J. Z. Huang, "Image fusion with local spectral consistency and dynamic gradient sparsity," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2760–2765.

[51] X. Bresson and T. F. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problems and Imaging*, vol. 2, no. 4, pp. 455–484, 2008.

[52] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[53] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.

[54] 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 Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3012–3021, 2007.

[55] 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 3rd Summaries of the Annual JPL Airborne Geoscience Workshop*, vol. 1, pp. 147–149, 1992.

[56] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 4, pp. 313–317, 2004.

[57] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[58] 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 Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2300–2312, 2002.

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

[60] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 72, no. 5, pp. 591–596, 2006.

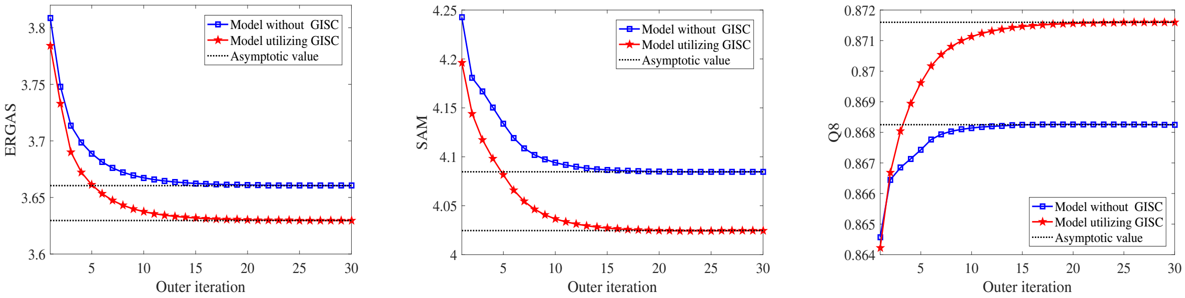


Fig. 12. The performance over the outer iterations for the models that use the $\mathcal{P}$ and $\widetilde{\mathcal{P}}$, respectively, on the reduced resolution (simulated) Rio dataset. It is worth to be remarked that optimal parameter configurations for both the models are used for fairness.

[61] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, “Multispectral and panchromatic data fusion assessment without reference,” *Photogrammetric Engineering and Remote Sensing*, vol. 74, no. 2, pp. 193–200, 2008.

[62] R. W. Dian, S. T. Li, and X. D. Kang, “Regularizing hyperspectral and multispectral image fusion by CNN denoiser,” *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[63] R. W. Dian, S. T. Li, and L. Y. Fang, “Learning a low tensor-train rank representation for hyperspectral image super-resolution,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2672–2683, 2019.

[64] R. W. Dian and S. T. Li, “Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization,” *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 5135–5146, 2019.