

A Novel Spatial Fidelity with Learnable Nonlinear Mapping for Panchromatic Sharpening

Rui Wen, Liang-Jian Deng, *Member, IEEE*, Zhong-Cheng Wu, Xiao Wu, Gemine Vivone, *Senior Member, IEEE*

Abstract—The purpose of panchromatic sharpening, i.e., pansharpening, is to fuse a low spatial resolution multispectral (LRMS) image with a high spatial resolution panchromatic (PAN) image, aiming to obtain a high spatial resolution multispectral (HRMS) image. Pan sharpening models based on variational optimization consist of a spectral fidelity term, a spatial fidelity term, and a regularization term. Most of the methods assume that the existing PAN image and the homologous HRMS image satisfy the global or local linear relationship, which could be far from the real case, thus causing sub-optimal performance. Inspired by the nonlinear mapping ability of machine learning (ML) techniques, we propose a novel spatial fidelity term with learnable nonlinear mapping (LNM-SF), which trains an implicit functional operator via a specifically designed convolutional neural network (CNN) and efficiently constructs the nonlinear relationship between the known PAN and the latent HRMS images. Relying upon the above description of the spatial fidelity term, a new variational model with a learnable nonlinear mapping in the spatial fidelity term for pansharpening, named LNM-PS, is simply integrated by the conventional spectral fidelity term into the proposed LNM-SF. To effectively solve the resulting optimization problem, we develop an alternating direction method of multipliers (ADMM)-based algorithm with the fast iterative shrinkage-thresholding algorithm (FISTA) as inner solver. Extensive numerical experiments on different datasets, assessing the performance both at reduced-resolution and full-resolution, show the superiority of the proposed LNM-PS method. The code is available at <https://github.com/liangjiandeng/-LNM-PS>.

Index Terms—Learnable Nonlinear Mapping, Variational Model, Convolutional Neural Networks, Pansharpening, Remote Sensing Image.

I. INTRODUCTION

HIGH resolution remote sensing images have many applicable fields [1], [2], [3], [4] such as environmental monitoring, agriculture, and forecasting. However, due to hardware limitations [5], [6], [7], existing satellites, e.g., IKONOS, QuickBird, GaoFen-2, WorldView-2, and WorldView-3, have to make a trade-off between the spectral information and the spatial details of remote sensing images. Thus, they can only capture a low spatial resolution multispectral (LRMS) image

This research is supported by NSFC (12271083, 12171072), Natural Science Foundation of Sichuan Province (2022NSFSC0501), Key Projects of Applied Basic Research in Sichuan Province (Grant No. 2020YJ0216), and National Key Research and Development Program of China (Grant No. 2020YFA0714001). (Corresponding author: Liang-Jian Deng)

Rui Wen, Liang-Jian Deng, Zhong-Cheng Wu, and Xiao Wu are with the School of Mathematical Sciences, University of Electronic Science and Technology of China (UESTC), Chengdu, Sichuan, 611731, China (e-mails: wenrui202102@163.com; liangjian.deng@uestc.edu.cn; wuzhch97@163.com; wxwsx1997@gmail.com).

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

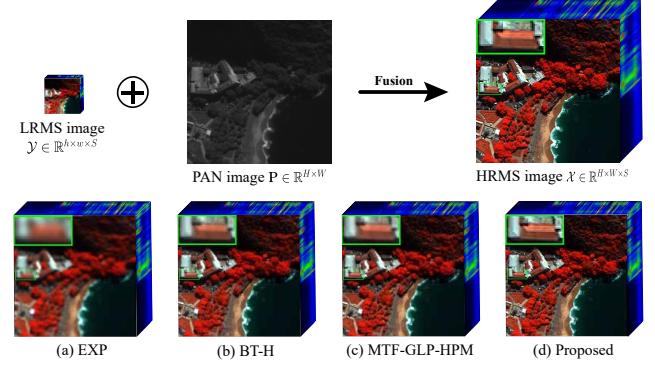


Fig. 1. Top row: a schematic representation of pansharpening exploiting the reduced-resolution Rio dataset (source: WorldView-3). Bottom row: (a) the EXP [8] and the fused images using (b) the BT-H [9], (c) the MTF-GLP-HPM [10], and (d) the proposed LNM-PS method. The 7th, 3rd, and 2nd channels are selected to generate the pseudo-color images.

and a high spatial resolution panchromatic (PAN) image. Hence, pansharpening emerged as a tool to combine the LRMS and PAN images to get high spatial resolution multispectral (HRMS) images. A schematic representation of pansharpening is depicted in Fig. 1.

A. Related Works

In recent years, pansharpening aroused interest in many fields and several fused methods have been proposed. They can be roughly divided into four categories [11], [12], [13], i.e., (i) component substitution (CS) approaches, (ii) multi-resolution analysis (MRA) techniques, (iii) variational optimization (VO) methods, and (iv) machine learning (ML) methods.

CS methods, also called spectral methods, are the earliest and also the most widely used because of their reduced computation cost. They rely upon the replacement of the spatial component of the LRMS image (obtained by a transformation of the spectral domain) with the PAN image. Representative instances are the principal component analysis (PCA) method [14], the Gram-Schmidt spectral sharpening (GS) approach [15], the band-dependent spatial-detail (BDS) method [16], and the partial replacement adaptive component substitution (PRACS) technique [17]. Compared with other categories, CS methods have a lower time consumption, while causing severe spectral distortion.

MRA approaches, also called spatial methods, extract spatial information from the known PAN image by spatial filtering, then injecting them into the UPMS image, i.e., the upsampled LRMS image, to obtain the final HRMS image.

Some instances of approaches include the Laplacian pyramids (LP) method [18], the morphological filters (MF) [19], the modulation transfer function generalized Laplacian pyramid with full-resolution regression-based injection (GLP-Reg-FS) method [20], and the modulation transfer function generalized Laplacian pyramid with high-pass modulation injection (MTF-GLP-HPM) technique [21]. Conversely to the CS methods, the fused images provided by the MRA methods easily suffer from the severe spatial distortion, while the spectral information of them is usually well-preserved.

VO techniques are rapidly gaining popularity in the field of pansharpening thanks to their exceptional flexibility and generalization, see, e.g., [22], [23], [24]. These methods usually describe an exact relationship between the known LRMS, PAN, and latent HRMS images based on some assumptions. The main processes involved in this category include: (i) constructing the energy function; and (ii) designing an algorithm to minimize this function, thus getting the desired image. The earliest variational approach is the P+XS method [25]. However, it obtains fusion results with inaccurate spatial details because of the global linear assumption between the PAN and HRMS images. To improve this approach, Fu *et al.* [26] proposed a local gradient constraint to extract spatial information, resulting in high-performance pansharpening results. However, it also fails to avoid spatial distortion due to the use of a local linear assumption that cannot be valid. In addition, many tensor decomposition-based methods have recently been applied to pansharpening, e.g., Li *et al.* [27] exploited the Tucker decomposition to estimate the core tensor and basis matrices from the desired HRMS and existing LRMS images, preserving the spatial details and spectral information of the fused HRMS image. Unfortunately, tensors also make the same linear assumption, which inevitably results in lower performance. Generally speaking, VO methods with a solid mathematical foundation can produce excellent results. Nevertheless, they cannot solve a nonlinear problem under the limitations of the existing regularization methodologies.

Recently, ML methods have been proposed and widely applied in several computer vision fields, see, e.g., hyperspectral image super-resolution [28], [29], [30], [31], image classification [32], [33], [34], and hyperspectral image denoising [35], [36], [37]. These methods obtain pansharpened images by inputting the existing LRMS and PAN images into the designed deep convolutional neural network, which has been pre-trained on a vast amount of data. In particular, many advanced ML approaches can faultlessly compensate for the shortcomings reflected by the VO techniques, e.g., the unsatisfactory accuracy and the use of the linear constraint between the known PAN and LRMS images. Masi *et al.* [38] proposed one of the first convolution neural networks (CNNs) for pansharpening. The network was named pansharpening neural network (PNN), and it uses a simple three-layers architecture to solve the problem. After that, many variants have been proposed. For instance, Liu *et al.* [39] proposed a novel fusion framework based on the generative adversarial networks (GANs), which takes into consideration both the application of neural networks to learn a nonlinear mapping and the solution to the upsampling problem. Besides, Zhang

et al. [40] presented a new pansharpening technique based on the deep neural networks (DNNs), called SSE-Net, which defines the spectral ratio loss to capture the nonlinear relationships among the channels of desired HRMS image, reducing the spectral distortion in the final outcome. On the one hand, these techniques can use more feature information getting high accuracy with an excellent nonlinear mapping ability. However, on the other hand, their generalization ability is not guaranteed.

Especially, hybrid approaches combining some of the above four classes have also been proposed, see, e.g., [41], [42], [43], [44]. These methods always plug the deep networks or deep priors via ML-based approaches into a part of the other three techniques, making the resulting hybrid model have the advantages of each of them, simultaneously. For instance, Deng *et al.* [45] propose the Fusion-net, which combines the CS or MRA with ML, achieving higher performance on various datasets. On the other hand, Ye *et al.* [46] creatively choose to plug the deep prior, which is composed of the proposed problem-specific recursive block and is trained in the gradient domain, into the spatial preservation term of the VO-based model and propose a hybrid pansharpening approach via a gradient-based deep network prior (GDN), achieving promising generalization in several datasets. Furthermore, Guo *et al.* [47] start from the Bayesian theory to attempt a nonlinear manner based on CNN, train parameters and introduce an anisotropic TV prior in multi-order gradient domains, and finally establish the posterior probability model, effectively preserving the more spatial structure and spectral details. Unlike the existing papers, we propose a novel and more comprehensive hybrid pansharpening framework, which employs the prominent ability of ML, e.g., nonlinear learning and matching. Ulteriorly, we lead it into pansharpening for the learning of spectral features, obtaining outstanding nonlinear presentation performance beyond traditional methods.

B. Contributions

Considering that the (local/global) linear hypothesis of the relationship between PAN and HRMS images cannot be valid, fusion results of traditional models are often unsatisfied. To overcome this limitation, we propose to learn the nonlinear relationship between the known PAN and the latent HRMS images by CNNs, which are pre-trained on a vast amount of data, precisely outlining the nonlinear relationship, thus breaking the bottleneck represented by the linear representation. More specifically, we exploit the nonlinear relationship in pansharpening to learn the spatial features from the PAN image in a better way, resulting in a new learnable nonlinear mapping spatial fidelity (called LNM-SF). Furthermore, we propose a novel variational model with learnable nonlinear mapping for pansharpening, named LNM-PS, by integrating the conventional spectral fidelity term and the proposed LNM-SF. To solve the LNM-PS method effectively, we develop an alternating direction method of multipliers (ADMM)-based algorithm with the fast iterative shrinkage-thresholding algorithm (FISTA) framework as the inner solver. Extensive experiments on reduced-resolution (i.e., simulated) and full-resolution (i.e., real) datasets assess the excellent performance

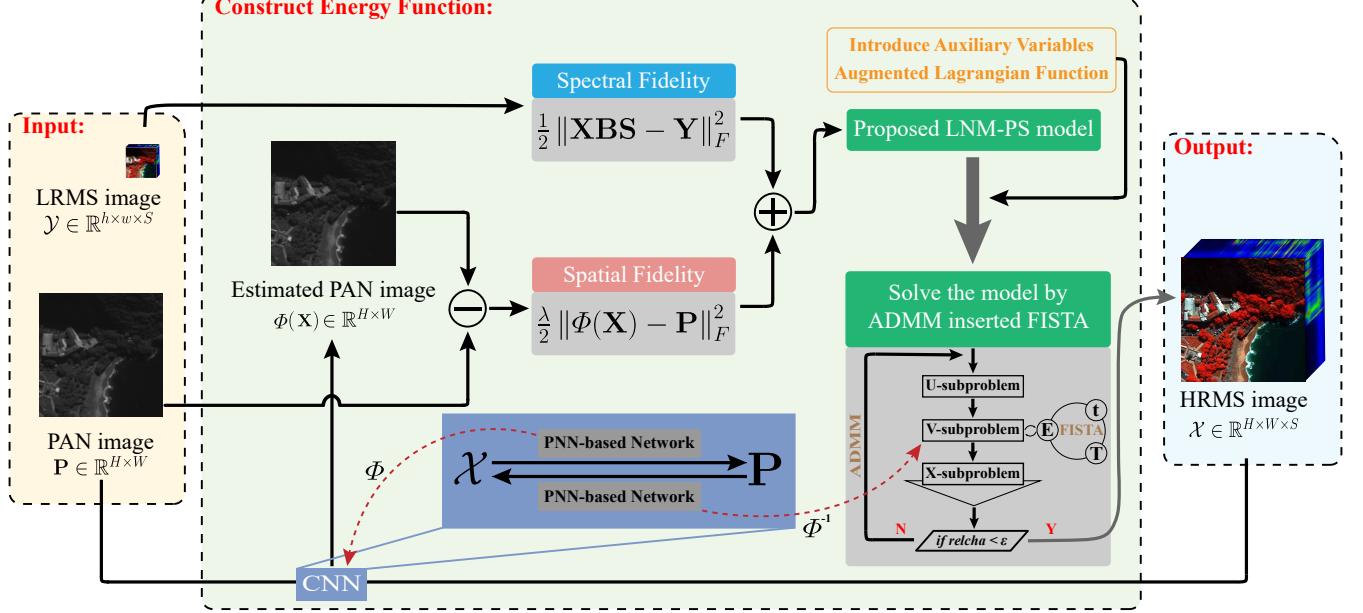


Fig. 2. Flowchart of the proposed LNM-PS method. The details of our framework can be found in Sect. III.

of the proposed LNM-PS framework compared to other state-of-the-art methods. Finally, some discussions are provided, i.e., the parameters analysis, the learnable nonlinear mapping, the generalization ability, the computational load, and the network structure.

The main contributions of our paper are summed up as follows:

- A new spatial fidelity with learnable nonlinear mapping (LNM-SF) is proposed to represent a more accurate relationship between the known PAN and latent HRMS images. Unlike the existing methods, a more reasonable explanation is given in the proposed method from modeling to designing algorithm about employing the nonlinear representation ability of ML methods, e.g., CNNs, as an implicit functional operator in the spatial fidelity term of a VO model.
- A novel variational model with learnable nonlinear mapping for pansharpening, named LNM-PS, is proposed via integrating the conventional spectral fidelity term and the LNM-SF.
- An ADMM-based algorithm with the FISTA framework as the inner solver is designed to solve the proposed hybrid model more effectively to have a simple yet effective pansharpening approach. More specifically, we derive the nonlinear function into linear functions by further exploring the properties of the trained nonlinear functions in some particular points, reducing the computational burden. Extensive experiments on several reduced-resolution and full-resolution datasets demonstrate the superiority of the proposed LNM-PS method compared with existing state-of-the-art approaches.

C. Paper Organization

The remaining of our paper is organized as follows. In Sect. II, the notation and the main motivations of the work are introduced. The proposed LNM-PS method is described in Sect. III. The solution of the created optimization problem is instead provided in Sect. IV. In Sect. V, the experimental analysis is shown comparing LNM-PS with some existing state-of-the-art methods. Finally, conclusions are drawn in Sect. VI.

II. NOTATION AND MOTIVATIONS

This section works on the presentation of the notation and the main motivations under the development of the proposed LNM-PS.

A. Notation

Lowercase letters denote scalars, i.e., a , matrices and vectors are denoted by uppercase and lowercase bold letters, respectively, i.e., X and \mathbf{x} , and calligraphic letters represent tensors, i.e., \mathcal{X} . We denote $\mathbf{1}$ as the all-ones matrix, whose dimension, when not given explicitly, is inferred from the context. The rest of the notation is summarized in Tab. I.

B. Motivations

A widely used hypothesis in some existing works is that the known PAN and latent HRMS images have the common spatial structure. According to this consideration, many methodologies extract the spatial details from the existing PAN image to reconstruct the desired HRMS image. Specifically, many VO-based methods assume that the PAN image is (globally or locally) linearly related to the HRMS image, see, e.g., [25], [43]. However, the relationship between the above two images

TABLE I
NOTATION.

Notation	Explanation
$\mathcal{X}, \mathbf{X}, \mathbf{x}, x$	Tensor, matrix, vector, scalar
$\mathcal{X} \in \mathbb{R}^{H \times W \times S}$	The HRMS image
$\mathbf{X} \in \mathbb{R}^{S \times HW}$	The mode-3 unfolding of \mathcal{X}
$\mathbf{X}_i \in \mathbb{R}^{H \times W}$	The i -th band of \mathcal{X}
$\mathcal{Y} \in \mathbb{R}^{h \times w \times S}$	The LRMS image
$\mathbf{Y} \in \mathbb{R}^{S \times hw}$	The mode-3 unfolding of \mathcal{Y}
$\mathbf{P} \in \mathbb{R}^{H \times W}$	The PAN image
x_{i_1, i_2, i_3}	the (i_1, i_2, i_3) -th element of \mathcal{X}
\circ	Hadamard product
$\ \mathcal{X}\ _F$	Frobenius norm
	$\ \mathcal{X}\ _F = \sqrt{\sum_{i_1, i_2, i_3} x_{i_1, i_2, i_3} ^2}$

(i.e., the known PAN and latent HRMS images) is often nonlinear, as demonstrated in Sect. III-B. Unfortunately, VO-based methods cannot effectively solve this nonlinear relationship. Thus, describing the nonlinear relationship between the above two images can represent the key to success for VO-based approaches.

Many researchers proposed CNN-based regularization terms to improve the performance of VO-based sharpening methods. For example, Wu *et al.* [41] employed the results of a deep convolution neural network (DCNN) as prior into a VO framework. Moreover, Dian *et al.* [42] introduced a special denoiser prior, which was learned by CNN, to fuse hyperspectral and multispectral images [48]. In both cases, the superiority (in terms of accuracy) of using CNNs in the fusion framework is demonstrated. However, the simple use of CNNs in the regularization term is insufficient since it fails to make the best of the strong ability of CNNs in the nonlinear mapping. Hence, how to exploit CNNs to map nonlinearities in data fidelity terms for VO-based pansharpening is a relevant problem. This paper is devoted to making the first attempts toward addressing this question.

Based on the above two considerations, a novel hybrid model is proposed, which improves traditional variational models by plugging a learnable nonlinear mapping into the spatial fidelity term (the LNM-SF), just including two data fidelity terms, i.e., the traditional spectral fidelity term and the proposed LNM-SF. This model breaks away from the classical linear constraints getting competitive results on different datasets against state-of-the-art techniques.

III. THE PROPOSED MODEL

A. The Spectral Fidelity Term

According to the theory of single image super-resolution (SISR), a widely used degradation model is as follows:

$$y = \Psi(x) + n, \quad (1)$$

where Ψ is the downsampling (decimation and filtering) operator, y is the observed low-performance image, x is the reference image, and n is a Gaussian noise.

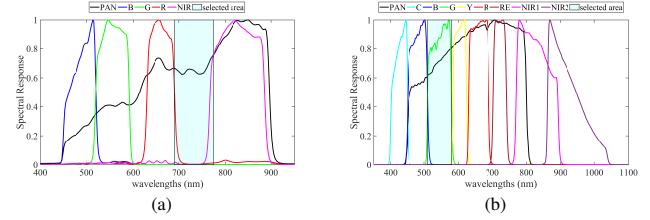


Fig. 3. The spectral responses for two sensors: (a) GaoFen-2 and (b) WorldView-3.

Similarly, in pansharpening, the LRMS image is viewed as the degraded version of the HRMS image. Mathematically, we have the following:

$$\mathbf{Y} = \mathbf{XBS} + \zeta, \quad (2)$$

where $\mathbf{B} \in \mathbb{R}^{HW \times HW}$ is the blurring matrix, $\mathbf{S} \in \mathbb{R}^{HW \times hw}$ denotes the decimation operator, and ζ is a zero-mean Gaussian noise. According to the above formula, the conventional spectral fidelity term is expressed as:

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

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

B. Learnable Nonlinear Mapping Spatial Fidelity Term

Many existing variational models extract the spatial information from the known PAN image by an inaccurate assumption, i.e., the linear relationship between the existing PAN and latent HRMS images, thus causing some spatial distortion [49]. The first idea linking PAN and HRMS images dated back to 2006 with the development of the P+XS framework [25], which assumes a global linear combination between the PAN and HRMS spectral bands. Mathematically, we have the following:

$$\mathbf{P} = \sum_{i=1}^S \alpha_i \mathbf{X}_i, \quad (4)$$

where $\{\alpha_i\}_{i=1}^S$ denote the weighting coefficients of the linear combination satisfying the constraint $\sum_{i=1}^S \alpha_i = 1$ and S is the number of the spectral channels of the HRMS image.

Based on the above-mentioned method, Xie *et al.* [43] proposed a novel way to calculate the weights denoted as $\mathbf{R} \in \mathbb{R}^{1 \times S}$:

$$\mathbf{R}^i = \frac{\sigma^i}{\sum_{i=1}^S (\sigma^i)}, \quad (5)$$

where σ^i denotes the largest eigenvalue of the i -th channel of the latent HRMS image, and \mathbf{R}^i represents the spectral response of the i -th channel. Thus, the spatial constraint between the known PAN and unknown HRMS images can be obtained as follows:

$$\mathbf{P} = \mathbf{RX}. \quad (6)$$

It is easy to see that both of them make the same assumption of a linear combination between the PAN image and the HRMS spectral bands. To further explore the relationship both in spectral and spatial between the PAN and HRMS

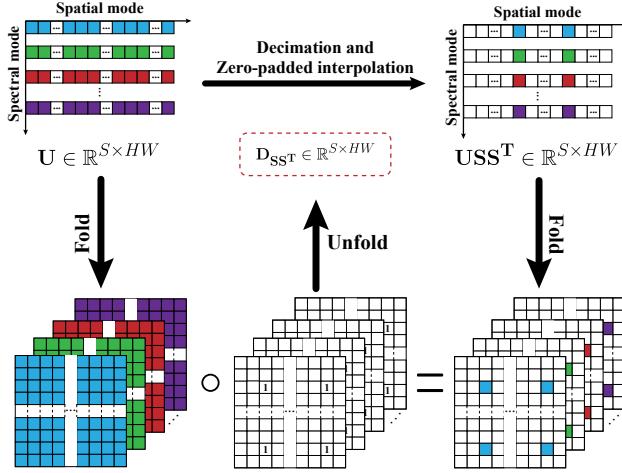


Fig. 4. The diagrammatic representation of (13) when $r = 4$. More details can be found in [41].

images, the spectral responses of two extensively used sensors for pansharpening are shown in Fig. 3, and the definition of spectral response is provided as follows.

Definition 3.1: *The spectral response is defined as the ratio of the number of photoelectrons, which are measured by the instrument system, to the radiance at a particular wavelength present at the entrance to the telescope aperture, i.e., the spectral response of a sensor only relies on its inherent parameters.* [12].

From the selected area of Fig. 3, we can see that the spectral response between the existing PAN and latent HRMS images is nonlinear, because the spectral response of PAN (the black curve) can not be a linear representation by other spectral response of channels of HRMS. It is worth noting that the linear assumption widely used in many existing VO-based methods still obtains good performance, but surely represents an approximation. On the contrary, the nonlinear relationship is more coherent with a real environment driving the development of new approaches towards considering this feature to improve fusion performance.

Thanks to the rapid development of CNNs for pansharpening, see, e.g., [50], [51], [52], [53], [54], [55], [56], the nonlinear mapping can be obtained by exploiting CNNs. Thus, we design a pre-trained CNN to learn the nonlinear relationship between the existing PAN and latent HRMS images. Then, we use this CNN structure as the embed operator to replace the traditional theory of linear algebra, making a substantial step toward constructing a nonlinear relationship in the data fidelity of VO-based methods. More details about how to train CNN can be found in Sect. V-B. The CNN-based spatial fidelity term can be gained as follows:

$$f_{\text{spatial}}(\mathbf{X}, \mathbf{P}) = \|\Phi(\mathbf{X}) - \mathbf{P}\|_F^2, \quad (7)$$

where Φ denotes nonlinear functional operator, which is pre-trained for the designed CNN and has no explicit expression.

For this reason, the resulting optimization model for pan-

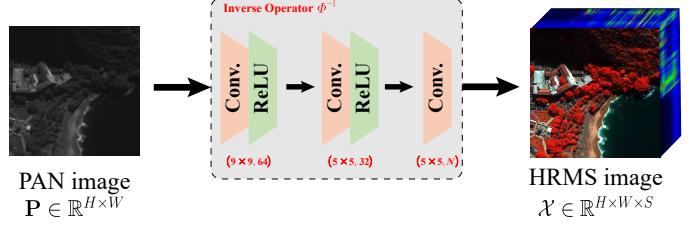


Fig. 5. The network representing the inverse operator of Φ (i.e., Φ^{-1}) designed by using the simple PNN [38].

sharpening can be obtained:

$$\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{XBS} - \mathbf{Y}\|_F^2 + \frac{\lambda}{2} \|\Phi(\mathbf{X}) - \mathbf{P}\|_F^2, \quad (8)$$

where $\lambda > 0$ is the regularization coefficient. It is worth noting that existing VO-based methods (see, e.g., [57], [41], and [24]) add a regularization term to further enhance the performance of the resulting pansharpening model. However, the improvement due to the regularization term is not sometimes significant, instead leading to an increment of the model's parameters requiring a greater computational burden to be solved. Thus, to simplify our model and promote the efficiency of the fusion framework, we consider no regularization term. The framework of the proposed LNM-PS optimization model is illustrated in Fig. 2.

Although convex, the objective function (8) is not advisable to directly calculate the derivative because it involves an extremely high burden on computation and storage. Afterward, an ADMM-based algorithm exploiting the FISTA framework as the inner solver will be designed to effectively pursue the iterative solution of the above optimization problem.

IV. THE PROPOSED ALGORITHM

This section is devoted to designing an ADMM [58]-based algorithm, which splits the proposed optimization problem into three simple and independent sub-problems, all of them having the corresponding closed-form solution. For simplicity, we can rewrite the resulting optimization model (8) as an isovalent constrained problem by introducing two auxiliary variables \mathbf{U} and \mathbf{V} . Hence, we have that:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \frac{1}{2} \|\mathbf{US} - \mathbf{Y}\|_F^2 + \frac{\lambda}{2} \|\Phi(\mathbf{V}) - \mathbf{P}\|_F^2, \\ \text{s.t.} \quad & \mathbf{U} = \mathbf{XB}, \quad \mathbf{V} = \mathbf{X}. \end{aligned} \quad (9)$$

The corresponding augmented Lagrangian function of the above optimization problem (9) can be expressed as:

$$\begin{aligned} \mathcal{L}_{\eta_1, \eta_2}(\mathbf{U}, \mathbf{V}, \mathbf{X}, \Theta_1, \Theta_2) = & \frac{1}{2} \|\mathbf{US} - \mathbf{Y}\|_F^2 \\ & + \frac{\lambda}{2} \|\Phi(\mathbf{V}) - \mathbf{P}\|_F^2 + \frac{\eta_1}{2} \left\| \mathbf{XB} - \mathbf{U} + \frac{\Theta_1}{\eta_1} \right\|_F^2 \\ & + \frac{\eta_2}{2} \left\| \mathbf{X} - \mathbf{V} + \frac{\Theta_2}{\eta_2} \right\|_F^2 + \text{const}, \end{aligned} \quad (10)$$

where Θ_1, Θ_2 denote the Lagrange multipliers, $\eta_1, \eta_2 > 0$ represent the penalty parameters, and *const* is a generic constant. At this time, (10) can be solved iteratively by updating the following three simpler subproblems.

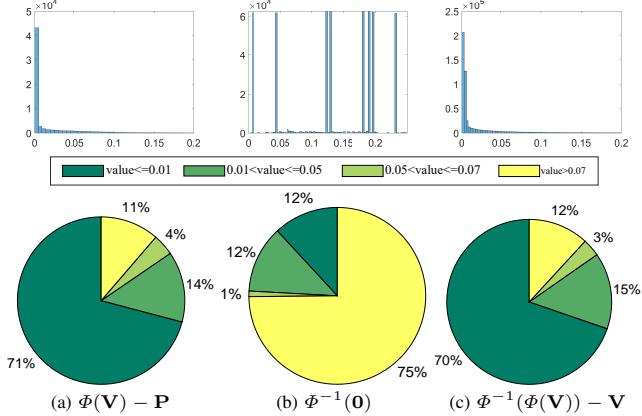


Fig. 6. The results generated by the trained nonlinear operator. First row: the histograms on the reduced-resolution Rio dataset (source: WorldView-3) using the trained functional operators Φ and Φ^{-1} . Second row: the corresponding pie charts. It can be seen that $\Phi(\mathbf{V}) - \mathbf{P}$, $\Phi^{-1}(\mathbf{0})$ and $\Phi^{-1}(\Phi(\mathbf{V})) - \mathbf{V}$ are not zero. In other words, there are errors that make the ideal properties not valid.

A. U-subproblem

According to (10), the **U**-subproblem can be expressed as the following optimization problem:

$$\arg \min_{\mathbf{U}} \frac{1}{2} \|\mathbf{US} - \mathbf{Y}\|_F^2 + \frac{\eta_1}{2} \left\| \mathbf{XB} - \mathbf{U} + \frac{\Theta_1}{\eta_1} \right\|_F^2, \quad (11)$$

which is a least-squares problem. Therefore, we can solve it by taking the derivative of (11), i.e.:

$$\mathbf{USS}^T - \mathbf{YS}^T - \eta_1 \mathbf{XB} + \eta_1 \mathbf{U} - \Theta_1 = 0, \quad (12)$$

where \mathbf{USS}^T can be regarded as the element-wise multiplication of \mathbf{D}_{SS^T} and \mathbf{U} , i.e.,

$$\mathbf{USS}^T = \mathbf{U} \circ \mathbf{D}_{SS^T} \quad (13)$$

with

$$\mathbf{D}_{SS^T} = [\text{vec}(\mathbf{K} \otimes \mathbf{1}), \text{vec}(\mathbf{K} \otimes \mathbf{1}), \dots, \text{vec}(\mathbf{K} \otimes \mathbf{1})]^T \quad (14)$$

where \circ and \otimes indicate the Hadamard and Kronecker product, respectively, $\text{vec}(\cdot)$ is the vectorization operator, and $\mathbf{K} \in \mathbb{R}^{r \times r}$ is a sparse matrix, e.g., when $r = 4$, $\mathbf{K}[(3, 3)] = 1$, whereas all the other cases is 0. The interesting readers can refer to the graphical representation as in Fig. 4 and to [41] for more details.

Therefore, the **U**-subproblem can be efficiently solved by:

$$\mathbf{U} := \frac{\mathbf{YS}^T + \eta_1 \mathbf{XB} + \Theta_1}{\mathbf{D}_{SS^T} + \eta_1 \mathbf{1}}, \quad (15)$$

note that the element-wise division is needed.

B. V-subproblem

By fixing other variables except for \mathbf{V} , the **V**-subproblem becomes:

$$\arg \min_{\mathbf{V}} \frac{\lambda}{2} \|\Phi(\mathbf{V}) - \mathbf{P}\|_F^2 + \frac{\eta_2}{2} \left\| \mathbf{X} - \mathbf{V} + \frac{\Theta_2}{\eta_2} \right\|_F^2, \quad (16)$$

which is a least-squares problem and easily to obtain the corresponding closed-form solution. However, since Φ cannot

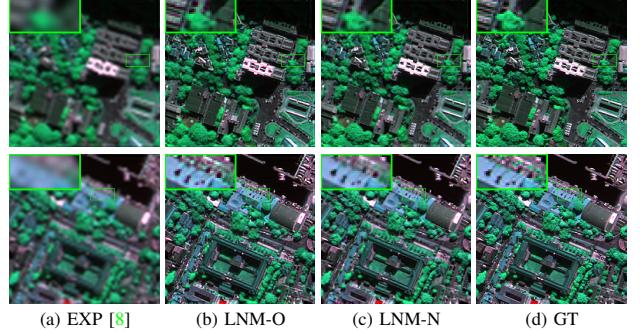


Fig. 7. The fusion results on the reduced-resolution Rio dataset (source: WorldView-3): (a) EXP [8], (b) LNM-O, (c) LNM-N, and (d) GT. The 3rd, 5th, and 6th bands are selected to generate the pseudo-color images.

be represented in the cooresponding matrix form, the FISTA framework [59] is applied to optimize this subproblem and separate this nonlinear operator, i.e., Φ [26].

Under the FISTA framework, (16) is divided into the following iterative procedure:

$$\mathbf{E} = \mathbf{T}^{p+1} - \Phi^{-1}(\Phi(\mathbf{V}) - \mathbf{P}) / L, \quad (17)$$

where L is the Lipschitz constant for $\Phi^{-1}(\Phi(\mathbf{V}) - \mathbf{P})$, whose value is generally set to 1, $\mathbf{E}, \mathbf{T} \in \mathbb{R}^{H \times W \times C}$ are the auxiliary variables, and p implies the p -th iteration of FISTA. Φ^{-1} denotes the inverse operator of Φ , which also is nonlinear and has no explicit expression, and both Φ and Φ^{-1} are continuous functions. The detailed network structure of Φ^{-1} is shown in Fig. 5.

However, when (17) is used to update \mathbf{E} , we obtain an unstable intermediate result, i.e., \mathbf{E} , which leads to an undesired fused image. The reason is that the operator Φ and its inverse operator Φ^{-1} are trained to satisfy the more precise and real nonlinear relationship between the existing PAN and latent HRMS images, which is meaningless for any temporary variable. Thus, when we use (17) employing the above operators to update the temporary variable \mathbf{E} , inaccurate results are produced. To avoid these results, we consider the following two situations to simplify (17).

In the ideal situation, the trained functional operator Φ and its inverse operator Φ^{-1} satisfy the following properties:

$$\begin{cases} \Phi(\mathbf{V}) = \mathbf{P}, \\ \Phi^{-1}(\mathbf{0}) = \mathbf{0}, \\ \Phi^{-1}(\Phi(\mathbf{V})) = \mathbf{V}, \\ \mathbf{X} = \mathbf{V}. \end{cases} \quad (18)$$

According to (18), we derive (17) as:

$$\begin{aligned} \mathbf{E} &= \mathbf{T}^{p+1} - \Phi^{-1}(\Phi(\mathbf{V}) - \mathbf{P}) / L \\ &= \mathbf{T}^{p+1} - \Phi^{-1}(\mathbf{0}) \\ &= \mathbf{T}^{p+1} - \mathbf{0} \\ &= \mathbf{T}^{p+1} - (\Phi^{-1}(\Phi(\mathbf{V})) - \Phi^{-1}(\mathbf{P})) \\ &= \mathbf{T}^{p+1} - (\mathbf{V} - \Phi^{-1}(\mathbf{P})). \end{aligned} \quad (19)$$

Thus, (17) can be simplified as follows:

$$\mathbf{E} = \mathbf{T}^{p+1} - (\mathbf{V} - \Phi^{-1}(\mathbf{P})), \quad (20)$$

TABLE II
QUANTITATIVE RESULTS FOR EXP [8], LNM-O, AND LNM-N, ON TWO TEST CASES OF THE REDUCED-RESOLUTION RIO DATASET (SOURCE: WORLDVIEW-3). THE APPROACHES ARE RUN ON THE CPU. (BOLD: BEST RESULTS)

Data	Index	EXP [8]	LNM-O	LNM-N	Ideal value
Metropolis	PSNR	27.814	31.378	33.145	$+\infty$
	SSIM	0.697	0.876	0.910	1
	SAM	7.003	6.810	5.659	0
	SCC	0.847	0.932	0.956	1
	ERGAS	8.230	5.371	4.377	0
	Q8	0.705	0.873	0.921	1
	Runtime(s)	0.012	218.9	2.640	0
Rio	PSNR	28.914	32.244	33.808	$+\infty$
	SSIM	0.690	0.873	0.902	1
	SAM	7.541	7.118	6.032	0
	SCC	0.819	0.914	0.942	1
	ERGAS	7.602	5.039	4.255	0
	Q8	0.685	0.868	0.913	1
	Runtime(s)	0.012	220.1	2.710	0

which is only related to the operator Φ^{-1} , thus saving time compared with (17) that involves the two operators, i.e., Φ and Φ^{-1} . Moreover, the above derivation process illustrates that although the trained functional operator Φ is nonlinear, it can still obtain the distribution property similar to the linear function in some special points, thanks to the iteration process of the algorithm and the special properties of Φ . This reduces both the complexity of the proposed LNM-PS model and the computational burden.

In the real situation, we find that (18) is not strictly satisfied, as shown in Fig. 6. The detailed operator is that we first unfold the three results, i.e., $\Phi(\mathbf{V}) - \mathbf{P}$, $\Phi^{-1}(\mathbf{0})$ and $\Phi^{-1}(\Phi(\mathbf{V})) - \mathbf{V}$, via the designed PNN-based network, as the corresponding arrays, then we plot the histograms and furthermore show the pie charts to help readers better understand. It is easy to see from Fig. 6 that $\Phi(\mathbf{V}) - \mathbf{P}$, $\Phi^{-1}(\mathbf{0})$, and $\Phi^{-1}(\Phi(\mathbf{V})) - \mathbf{V}$ are not zero. The reason that causes this case is the existence of weights and biases in each layer when we train these two nonlinear mapping operators. However, this is the only way that must be passed in the real situation, i.e., it is an indisputable fact that the above three results are not the non-zero tensors. However, we can denote them, mathematically, i.e., there are errors as ϵ_1 , ϵ_2 , and ϵ_3 , making Φ and Φ^{-1} having similar properties as (18), i.e.,

$$\begin{cases} \Phi(\mathbf{V}) = \mathbf{P} + \epsilon_1, \\ \Phi^{-1}(\mathbf{0}) = \mathbf{0} + \epsilon_2, \\ \Phi^{-1}(\Phi(\mathbf{V})) = \mathbf{V} + \epsilon_3, \\ \mathbf{X} = \mathbf{V}. \end{cases} \quad (21)$$

However, these errors are acceptable, i.e., (18) is still hold-

ing in the real situation since the improvement of the resulting performance is remarkable. To validate the effectiveness of this derivation (i.e., simplifying (17) with (20)), we choose two test cases related to the reduced-resolution Rio dataset acquired by the extensively used WorldView-3 sensor for pansharpening. We denote the two algorithms using (17) and (20) as LNM-O and LNM-N, where O and N stand for the original version and the new version, respectively. The visual comparison of the two algorithms on the two test cases is displayed in Fig. 7. The 3rd, 7th, and 6th bands are selected to generate the pseudo-color images. The quantitative results are shown in Tab. II. One can see from the Table that LNM-O and LNM-N achieve better fusion results compared with the baseline EXP [8]. Moreover, LNM-O needs more computation to get the fusion results, even getting lower performance than LNM-N.

To sum up, this transformation from (17) to (20) is applicable to effectively update the variable \mathbf{E} .

Under the FISTA framework, the \mathbf{V} -subproblem can be rewritten as the following minimization problem with the updated \mathbf{E} in (20):

$$\arg \min_{\mathbf{V}} \frac{\lambda}{2} \|\mathbf{V} - \mathbf{E}\|_F^2 + \frac{\eta_2}{2} \left\| \mathbf{X} - \mathbf{V} + \frac{\Theta_2}{\eta_2} \right\|_F^2, \quad (22)$$

which is a least-squares problem and can be simple to achieve the following closed-form solution:

$$\mathbf{V} := \frac{\lambda \mathbf{E} + \eta_2 \mathbf{X} + \Theta_2}{(\lambda + \eta_2) \mathbf{1}}, \quad (23)$$

notably, the element-wise division is needed, and the solution of \mathbf{V} -subproblem can be summed up in Algorithm 1.

C. X-subproblem

Similarly, \mathbf{X} can be updated by solving the following optimization problem:

$$\arg \min_{\mathbf{X}} \frac{\eta_1}{2} \left\| \mathbf{X} \mathbf{B} - \mathbf{U} + \frac{\Theta_1}{\eta_1} \right\|_F^2 + \frac{\eta_2}{2} \left\| \mathbf{X} - \mathbf{V} + \frac{\Theta_2}{\eta_2} \right\|_F^2, \quad (24)$$

which is a simple least-squares problem. Thus, we have that:

$$\mathbf{X} := \mathcal{F}^{-1} \left(\frac{\mathbf{M} + \mathbf{N}}{\eta_1 \mathcal{F}(\mathbf{B}) \circ \mathcal{F}(\mathbf{B})^\dagger + \eta_2 \mathbf{1}} \right). \quad (25)$$

In (25), the expressions of \mathbf{M} and \mathbf{N} are as follows:

$$\begin{aligned} \mathbf{M} &= \eta_1 \mathcal{F}(\mathbf{U}) \circ \mathcal{F}(\mathbf{B})^\dagger - \mathcal{F}(\Theta_1) \circ \mathcal{F}(\mathbf{B})^\dagger, \\ \mathbf{N} &= \eta_2 \mathcal{F}(\mathbf{V}) - \mathcal{F}(\Theta_2), \end{aligned} \quad (26)$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ denote the fast Fourier transform (FFT) and the corresponding inverse operator, respectively, the symbol \dagger is the complex conjugate, and the division is component-wise.

D. Updating Multipliers Θ_1, Θ_2

Finally, on the basis of ADMM framework, the corresponding Lagrangian multipliers Θ_1 and Θ_2 can be updated via:

$$\begin{aligned} \Theta_1 &:= \Theta_1 + \eta_1 (\mathbf{X} \mathbf{B} - \mathbf{U}), \\ \Theta_2 &:= \Theta_2 + \eta_2 (\mathbf{X} - \mathbf{V}). \end{aligned} \quad (27)$$

Algorithm 1: The FISTA-based framework for solving V-subproblem (16)

```

Input:  $\mathbf{X}$ ,  $\mathbf{P}$ ,  $\Theta_2$ ,  $\lambda$ ,  $\eta_2$ ,  $L$ ,  $k_{imax}$ 
Initialization:  $\mathbf{T}^0$ ,  $\mathbf{V}^0$ ,  $t^0$ 
for  $p = 1$  to  $k_{imax}$  do
    Update  $\mathbf{E}$  via (20);
    Update  $t$  via  $t = \left(1 + \sqrt{1 + 4(t^p)^2}\right) / L$ ;
    Update  $\mathbf{T}$  via  $\mathbf{T} = \mathbf{V}^{p+1} + \frac{t^p - 1}{t^{p+1}} (\mathbf{V}^{p+1} - \mathbf{V}^p)$ ;
end
Output:  $\mathbf{V}$  via (23)

```

Algorithm 2: The designed ADMM-based algorithm for more effectively solve the proposed optimization model (8)

```

Input:  $\mathbf{Y}$ ,  $\mathbf{P}$ ,  $\Theta_1$ ,  $\Theta_2$ ,  $\lambda$ ,  $\eta_1$ ,  $\eta_2$ ,  $k_{omax}$ 
Initialization:  $\mathbf{X}^0 = \mathbf{0}$ 
while  $relcha > \varepsilon$  and  $k < k_{omax}$  do
    Update  $\mathbf{U}$  via (15);
    Update  $\mathbf{V}$  via Algorithm 1;
    Update  $\mathbf{X}$  via (25);
    Update multipliers  $\Theta_1, \Theta_2$  via (27);
end
Output: Fused HRMS image,  $\mathbf{X}$ 

```

The termination criterion is considered in the ADMM algorithm based on the relative change ($relcha$) between two successive pansharpened outputs that should be less than a set tolerance value ε , i.e.:

$$relcha = \|\mathbf{X}^{k+1} - \mathbf{X}^k\|_F / \|\mathbf{X}^k\|_F < \varepsilon. \quad (28)$$

where k denotes the k -th iteration of ADMM. The procedure to solve the resulting optimization model (8) is summarized in Algorithm 2, where k_{omax} is the maximum number of iterations. The convergence of the designed iterative scheme is theoretically guaranteed, see [60] for more details.

V. EXPERIMENTAL RESULTS

This section is committed to demonstrating the effectiveness of LNM-PS approach. More specifically, we compare the LNM-PS with some state-of-the-art approaches on remote sensing datasets (i.e., Guangzhou, Indianapolis, and Rio datasets) acquired by several widely used sensors for pansharpening, i.e., GaoFen-2, QuickBird, and WorldView-3. We choose several techniques from different pansharpening categories. Furthermore, two assessment procedures, including reduced-resolution and full-resolution, are considered. On the other hand, to obtain more convincing results, all the approaches in the benchmark are also run on the same software and hardware platforms, i.e., MATLAB (R2021b) and a computer of 16Gb RAM with AMD Ryzen 5 5600X 6-Core at 3.7 GHz. It is worth to be noted that the tolerance value, ε , and the number of inner iterations, k_{imax} , are empirically set to 2×10^{-4} and 2 in all the experiments, respectively. Finally, we set k_{omax} to 30, as shown in Sect. V-F2.

A. Datasets

The datasets are obtained from the website¹. We randomly sample the original LRMS and PAN images to have patches from the selected datasets. Wald's protocol [64] is adopted and Gaussian filters are designed to match the modulation transfer functions (MTFs) of the acquiring sensors followed by a decimation obtaining a scaling ratio of 4 [65]. Thus, the following three training datasets are obtained:

- GaoFen-2 training set: 21607 pairs of {PAN: $64 \times 64 \times 1$, MS: $16 \times 16 \times 4$, GT: $64 \times 64 \times 4$ }.
- QuickBird training set: 20685 pairs of {PAN: $64 \times 64 \times 1$, MS: $16 \times 16 \times 4$, GT: $64 \times 64 \times 4$ }.
- WorldView-3 training set: 8806 pairs of {PAN: $64 \times 64 \times 1$, MS: $16 \times 16 \times 8$, GT: $64 \times 64 \times 8$ }.

According to the same simulation procedure, we generate the testing datasets (not included in the training set):

- The Guangzhou dataset, obtained by the GaoFen-2 (GF-2) sensor, has an LRMS image with 4 spectral channels and a size of 64×64 and 256×256 for the LRMS and PAN images, respectively.
- The Indianapolis dataset, obtained by the QuickBird (QB) sensor, has an LRMS image with 4 spectral channels and a size of 64×64 and 256×256 for the LRMS and PAN images, respectively.
- The Rio dataset, acquired by the WorldView-3 (WV-3) sensor, has an LRMS image with 8 spectral channels and a size of 64×64 and 256×256 for the LRMS and PAN images, respectively.

B. Trained Functional Operator

A CNN is needed to define spatial fidelity, i.e., Φ^{-1} . Thus, we adopt a 3 Conv-ReLU layer CNN as Φ^{-1} . To better illustrate our algorithm, we also train Φ , which leverages on the same network with new weights fed by Φ^{-1} to learn how to convert HRMS to PAN. However, the final algorithm only needs to employ Φ^{-1} , see Sect. IV-B. Notably, Φ and Φ^{-1} use standard Pytorch convolution operations with padding and different optimizer settings compared to the PNN [38]. We adopt Adam optimizers with the learning rate being 2×10^{-4} and $(0.9, 0.999)$. The CNN implementation is performed in Pytorch 1.11.1 using an Nvidia GeForce RTX 3080 GPU.

C. Benchmarking

We compare LNM-PS with several state-of-the-art pansharpening approaches. The EXP [8] is considered as the baseline. Into the CS class, we chose the haze corrected version of the Brovey transformation (BT-H) [9] and the robust band-dependent spatial-detail (BD-PC) method [61]. Inside the MRA family, the GLP with MTF-matched filters and the HPM injection (HPM) method [10] and the context-based GLP (C-GLP) [62] are considered. In addition, the VO-based method with local gradient constraints (LGC) [26] and the context-aware details injection fidelity (CDIF) [24] are also added. To complete the benchmarking, the pansharpening neural network

¹<http://openremotesensing.net/kb/codes/pansharpening/>

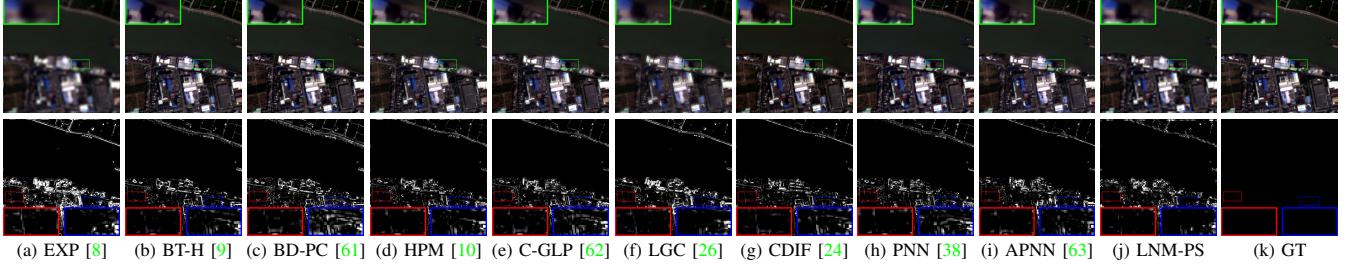


Fig. 8. Visual results on the reduced-resolution Guangzhou dataset (source: GaoFen-2). The PAN size is 256×256 . First row: the fused images. Second row: the homologous residual maps when we used the GT image as the reference image. The Blue (3rd), Green (2nd), and Red (1st) channels are selected to generate these images.

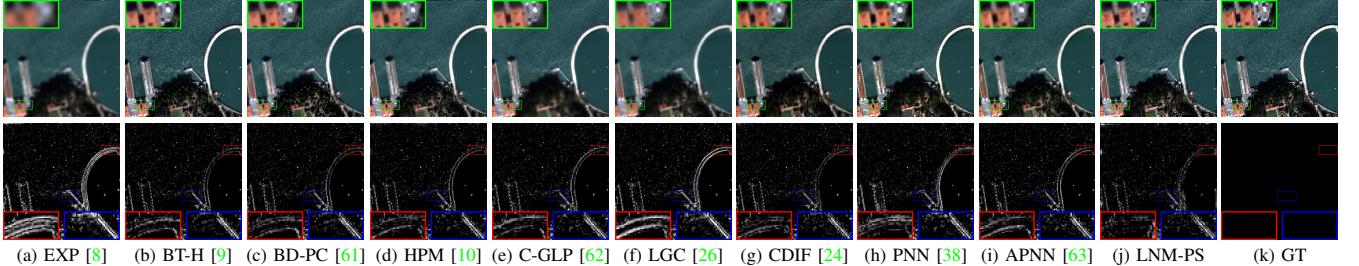


Fig. 9. Visual results on the reduced-resolution Indianapolis dataset (source: QuickBird). The PAN size is 256×256 . First row: the fused images. Second row: the homologous residual maps when we used the GT image as the reference image. The Blue (3rd), Green (2nd), and Red (1st) channels are selected to generate these images.

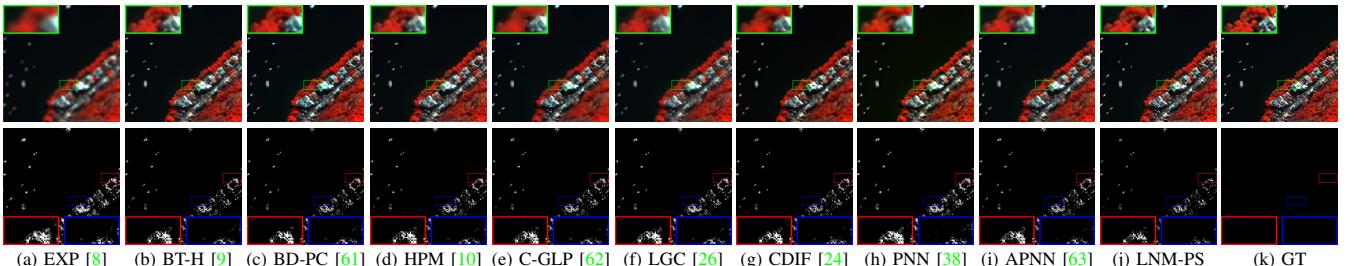


Fig. 10. Visual results on the reduced-resolution Rio dataset (source: WorldView-3). The PAN size of this image is 256×256 . First row: the fused images from different frameworks. Second row: the homologous residual maps when we used the GT image as the reference image. The NIR (7th), Green (3rd), and Blue (2nd) channels are selected to generate these Pseudo color images.

(PNN) [38] and the corresponding target adaptive version (APNN) [63] are also chosen belonging to the ML class. It is worth remarking that no very recent ML-based approaches have been adopted, and the reasons why we did not consider them in our paper can be found in Sect. V-F5.

D. Quality Metrics

In this paper, the peak signal-to-noise ratio (PSNR) [66], the structural similarity (SSIM) [66], the spectral angle mapper (SAM) [67], the spatial correlation coefficient (SCC) [68], the erreur relative globale adimensionnelle de synthèse (ER-GAS) [69], and the $Q2^n$ index [70], [71], i.e., the Q4 for 4-channels MS images or the Q8 for 8-channels MS images, are chosen for the assessment at the reduced-resolution. Instead, at the full-resolution, the quality with no reference (QNR) [72] index, including a spectral quality index, D_λ , and a spatial quality index, D_s , is used.

E. Qualitative and Quantitative Comparison

1) Reduced-Resolution Assessment: This sub-section is committed to assessing the performance of LNM-PS on three reduced-resolution datasets (i.e., GuangZhou, Indianapolis, and Rio), gained by GaoFen-2, QuickBird, and WorldView-3 sensors, respectively.

The visual comparisons and the corresponding residual images of several pansharpening methods on different datasets are displayed in Figs. 8-10. Focusing on Fig. 8, it can be observed from the visual results that the images in Figs. 8 (a), (f), and (i) are more blurry than the other compared methods on the eaves and the river bank in the close-ups, while the rest of the results can clearly retain spatial details in those zones. The related residual images are depicted in the second row of Fig. 8. Having a look at the close-ups, it is easy to conclude that the LNM-PS can produce more visually appealing fusion outcomes than the competing methods. Similar conclusions can be done inspecting Figs. 9 and 10.

Furthermore, Tabs. III-V report the quantitative results. It is clear that the LNM-PS is the best approach in all the test cases and for all the image quality metrics. Compared to approaches belonging to the same class (i.e., VO), the LNM-PS gets an

TABLE III

QUANTITATIVE RESULTS (MEAN AND STANDARD DEVIATION) FOR ALL THE COMPARED APPROACHES INCLUDED IN THE BENCHMARK ON 84 IMAGES FROM THE REDUCED-RESOLUTION GUANGZHOU DATASET (SOURCE: GAOFEN-2). THE CNN-BASED APPROACHES ARE RUN ON GPU (G), WHEREAS THE OTHERS ON CPU (C). (BOLD: BEST RESULTS; UNDERLINED: SECOND BEST RESULTS)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q4	Runtime(s)
EXP [8]	37.961 ± 2.221	0.912 ± 0.031	2.027 ± 0.346	0.903 ± 0.034	2.548 ± 0.429	0.775 ± 0.043	0.001(C)
BT-H [9]	41.374 ± 1.613	0.960 ± 0.011	1.849 ± 0.318	0.953 ± 0.023	1.718 ± 0.287	0.890 ± 0.035	<u>0.019(C)</u>
BD-PC [61]	40.822 ± 1.752	0.955 ± 0.013	1.873 ± 0.329	0.951 ± 0.023	1.835 ± 0.295	0.886 ± 0.032	0.029(C)
HPM [10]	40.598 ± 1.830	0.948 ± 0.017	1.796 ± 0.333	0.947 ± 0.023	1.868 ± 0.298	0.874 ± 0.035	0.085(C)
C-GLP [62]	41.128 ± 1.614	0.957 ± 0.011	1.816 ± 0.314	0.951 ± 0.022	1.799 ± 0.247	0.888 ± 0.033	0.268(C)
LGC [26]	40.404 ± 1.988	0.952 ± 0.014	1.728 ± 0.304	0.949 ± 0.019	1.927 ± 0.288	0.869 ± 0.026	4.541(C)
CDIF [24]	41.644 ± 1.642	0.961 ± 0.011	1.710 ± 0.308	0.957 ± 0.020	1.666 ± 0.258	0.898 ± 0.029	13.018(C)
PNN [38]	40.385 ± 1.637	0.948 ± 0.014	2.286 ± 0.419	0.948 ± 0.022	2.171 ± 0.360	0.886 ± 0.031	0.204(G)
APNN [63]	41.294 ± 1.880	0.961 ± 0.013	1.877 ± 0.592	0.954 ± 0.022	1.788 ± 0.822	<u>0.901 \pm 0.038</u>	2.359(G)
LNM-PS	42.113 ± 1.899	0.962 ± 0.012	1.500 ± 0.263	0.961 ± 0.018	1.522 ± 0.257	0.913 ± 0.028	2.302(C)
Ideal value	$+\infty$	1	0	1	0	1	0

TABLE IV

QUANTITATIVE RESULTS (MEAN AND STANDARD DEVIATION) FOR ALL THE COMPARED APPROACHES INCLUDED IN THE BENCHMARK ON 40 IMAGES FROM THE REDUCED-RESOLUTION INDIANAPOLIS DATASET (SOURCE: QUICKBIRD). THE CNN-BASED APPROACHES ARE RUN ON GPU (G), WHEREAS THE OTHERS ON CPU (C). (BOLD: BEST RESULTS; UNDERLINED: SECOND BEST RESULTS)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q4	Runtime(s)
EXP [8]	27.140 ± 1.130	0.648 ± 0.040	8.993 ± 0.768	0.742 ± 0.027	12.613 ± 0.995	0.582 ± 0.032	0.001(C)
BT-H [9]	31.374 ± 1.003	0.849 ± 0.016	<u>7.647 ± 0.695</u>	0.912 ± 0.014	7.771 ± 0.573	0.850 ± 0.019	<u>0.019(C)</u>
BD-PC [61]	31.413 ± 0.922	0.846 ± 0.016	<u>8.570 ± 1.000</u>	0.907 ± 0.016	7.775 ± 0.573	0.857 ± 0.030	0.028(C)
HPM [10]	31.241 ± 1.168	0.852 ± 0.014	8.168 ± 0.925	0.902 ± 0.034	8.454 ± 2.320	0.858 ± 0.023	0.082(C)
C-GLP [62]	30.842 ± 0.974	0.827 ± 0.021	8.186 ± 0.770	0.893 ± 0.017	8.395 ± 0.627	0.832 ± 0.036	0.259(C)
LGC [26]	29.715 ± 1.033	0.790 ± 0.022	8.350 ± 0.730	0.881 ± 0.014	9.530 ± 0.705	0.770 ± 0.020	4.291(C)
CDIF [24]	<u>31.994 ± 0.960</u>	<u>0.857 ± 0.016</u>	7.842 ± 0.765	<u>0.921 ± 0.013</u>	<u>7.386 ± 0.561</u>	<u>0.871 ± 0.019</u>	11.997(C)
PNN [38]	30.483 ± 1.071	0.850 ± 0.017	8.257 ± 0.494	0.887 ± 0.012	8.519 ± 0.631	0.854 ± 0.017	0.175(G)
APNN [63]	31.017 ± 1.007	0.842 ± 0.017	7.853 ± 0.706	0.900 ± 0.014	8.195 ± 0.585	0.849 ± 0.015	2.219(G)
LNM-PS	34.156 ± 0.904	0.906 ± 0.012	6.652 ± 0.519	0.951 ± 0.011	5.660 ± 0.457	0.920 ± 0.019	2.280(C)
Ideal value	$+\infty$	1	0	1	0	1	0

TABLE V

QUANTITATIVE RESULTS (MEAN AND STANDARD DEVIATION) FOR ALL THE COMPARED APPROACHES INCLUDED IN THE BENCHMARK ON 15 IMAGES FROM THE REDUCED-RESOLUTION RIO DATASET (SOURCE: WORLDVIEW-3). THE CNN-BASED APPROACHES ARE RUN ON GPU (G), WHEREAS THE OTHERS ON CPU (C). (BOLD: BEST RESULTS; UNDERLINED: SECOND BEST RESULTS)

Method	PSNR	SSIM	SAM	SCC	ERGAS	Q8	Runtime(s)
EXP [8]	30.048 ± 2.024	0.761 ± 0.080	6.148 ± 2.068	0.839 ± 0.046	7.991 ± 2.071	0.577 ± 0.094	0.002(C)
BT-H [9]	33.491 ± 2.029	<u>0.895 ± 0.036</u>	<u>5.058 ± 1.558</u>	0.928 ± 0.021	5.367 ± 1.423	0.774 ± 0.095	<u>0.019(C)</u>
BD-PC [61]	33.329 ± 1.920	0.880 ± 0.042	6.031 ± 2.065	0.923 ± 0.024	5.613 ± 1.572	0.768 ± 0.100	0.053(C)
HPM [10]	32.566 ± 2.219	0.888 ± 0.040	5.645 ± 1.933	0.905 ± 0.043	7.952 ± 5.260	0.774 ± 0.097	0.111(C)
C-GLP [62]	32.346 ± 1.708	0.832 ± 0.058	6.019 ± 2.025	0.902 ± 0.033	6.420 ± 1.851	0.678 ± 0.125	0.347(C)
LGC [26]	32.458 ± 1.990	0.865 ± 0.043	5.350 ± 1.741	0.915 ± 0.023	6.038 ± 1.604	0.727 ± 0.093	8.552(C)
CDIF [24]	<u>33.835 ± 1.971</u>	0.891 ± 0.035	5.163 ± 1.668	<u>0.933 ± 0.020</u>	<u>5.262 ± 1.546</u>	<u>0.784 ± 0.104</u>	31.883(C)
PNN [38]	31.158 ± 1.708	0.855 ± 0.049	7.402 ± 1.462	0.898 ± 0.032	6.716 ± 1.510	0.721 ± 0.108	0.498(G)
APNN [63]	32.980 ± 1.747	0.882 ± 0.037	5.344 ± 1.647	0.921 ± 0.024	5.711 ± 1.642	0.753 ± 0.103	2.694(G)
LNM-PS	35.276 ± 1.943	0.928 ± 0.024	4.681 ± 1.357	0.954 ± 0.013	4.266 ± 1.082	0.825 ± 0.094	3.181(C)
Ideal value	$+\infty$	1	0	1	0	1	0



Fig. 11. Fusion results (including close-ups on the left-upper corner) for the compared approaches included in the benchmark on the four-bands Guangzhou dataset at full-resolution (source: GaoFen-2). The PAN size of this image is 512×512 . (a)-(j) Visual results of several selected fusion approaches in true colors and (k) the PAN image.

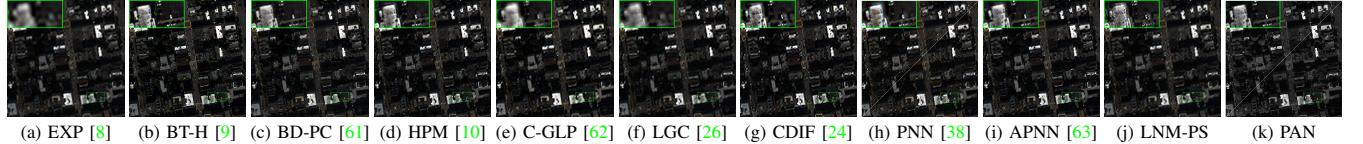


Fig. 12. Fusion results (including close-ups on the left-upper corner) for the compared approaches included in the benchmark on the four-bands Indianapolis dataset at full-resolution (source: QuickBird). The PAN size of this image is 512×512 . (a)-(j) Visual results of several selected fusion approaches in true colors and (k) the PAN image.

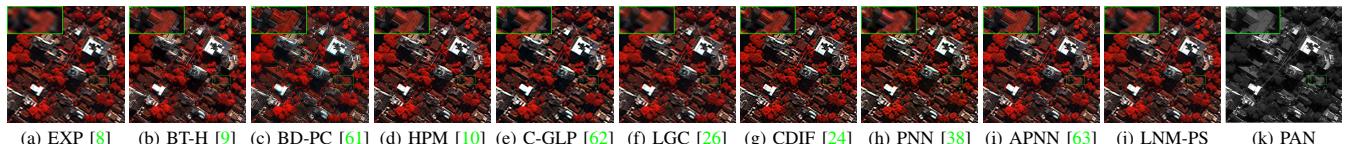


Fig. 13. Fusion results (including close-ups on the left-upper corner) for the compared approaches included in the benchmark on the eight-bands Rio dataset at full-resolution (source: WorldView-3). The PAN size of this image is 512×512 . (a)-(j) Visual results of several selected fusion approaches using a pseudo-colors composition (NIR (7th), green (3rd), and blue (2nd) channels) and (k) the PAN image.

TABLE VI

QUANTITATIVE RESULTS (MEAN AND STANDARD DEVIATION) FOR ALL THE COMPARED APPROACHES INCLUDED IN THE BENCHMARK ON (A) 50 IMAGES FROM THE FULL-RESOLUTION GUANGZHOU DATASET (SOURCE: GAOFEN-2); (B) 38 IMAGES FROM THE FULL-RESOLUTION INDIANAPOLIS DATASET (SOURCE: QUICKBIRD). THE CNN-BASED APPROACHES ARE RUN ON GPU (G), WHEREAS THE OTHERS ON CPU (C). (BOLD: BEST RESULTS; UNDERLINE: SECOND BEST RESULTS)

Method	(a) Full-Resolution GaoFen-2 Dataset			Runtime(s)	(b) Full-Resolution QuickBird Dataset			Runtime(s)
	D _λ	D _S	QNR		D _λ	D _S	QNR	
EXP [8]	0.005 ± 0.004	0.059 ± 0.025	0.937 ± 0.022	0.028(C)	0.002 ± 0.002	0.128 ± 0.026	0.870 ± 0.025	0.029(C)
BT-H [9]	0.100 ± 0.027	0.209 ± 0.037	0.712 ± 0.049	<u>0.054(C)</u>	0.059 ± 0.027	0.224 ± 0.062	0.731 ± 0.073	<u>0.066(C)</u>
BD-PC [61]	0.072 ± 0.036	0.198 ± 0.040	0.745 ± 0.062	0.062(C)	0.039 ± 0.038	0.215 ± 0.068	0.757 ± 0.087	0.072(C)
HPM [10]	0.112 ± 0.042	0.187 ± 0.047	0.723 ± 0.072	0.184(C)	0.067 ± 0.034	0.179 ± 0.062	0.768 ± 0.077	0.211(C)
C-GLP [62]	0.041 ± 0.018	0.080 ± 0.038	0.882 ± 0.049	1.101(C)	0.055 ± 0.040	0.122 ± 0.074	0.832 ± 0.096	1.359(C)
LGC [26]	<u>0.006 ± 0.004</u>	<u>0.029 ± 0.011</u>	<u>0.966 ± 0.012</u>	28.017(C)	<u>0.007 ± 0.003</u>	<u>0.036 ± 0.010</u>	<u>0.958 ± 0.009</u>	38.367(C)
CDIF [24]	0.044 ± 0.035	0.068 ± 0.021	0.892 ± 0.057	40.642(C)	0.033 ± 0.025	0.089 ± 0.063	0.883 ± 0.078	48.619(C)
PNN [38]	0.088 ± 0.028	0.213 ± 0.035	0.719 ± 0.050	0.239(G)	0.042 ± 0.029	0.165 ± 0.066	0.801 ± 0.080	0.234(G)
APNN [63]	0.039 ± 0.035	0.085 ± 0.032	0.880 ± 0.051	2.478(G)	0.020 ± 0.018	0.085 ± 0.035	0.898 ± 0.054	2.419(G)
LNM-PS	0.007 ± 0.003	0.026 ± 0.007	0.967 ± 0.006	5.682(C)	<u>0.007 ± 0.003</u>	0.019 ± 0.006	0.975 ± 0.008	6.139(C)
Ideal value	0	0	1	0	0	0	1	0

even better computational burden representing the fastest VO-based approach among the compared ones. This is obtained thanks to the use of just two data fidelity terms.

2) *Full-Resolution Assessment*: This sub-section corroborates the performance of the proposed pansharpening method on three full-resolution datasets obtained by above three sensors, respectively. The visual results are displayed in Figs. 11–13. One can observe from them that although EXP [8] retains better spectral information, it loses many details. Besides, the pseudo-color representation of the fused products generated by the PNN [38] are plenty of artifacts, see, e.g., Fig. 13 (h).

Instead, the LNM-PS consistently provides high-performance results. The quantitative assessment is reported in Tabs. VI–VII. The best overall results (i.e., the best values of the QNR index) are obtained by the proposed LNM-PS approach driven by the best values obtained by the spatial counterpart.

F. Discussions

In this sub-section, several discussions about the proposed LNM-PS framework as the parameters tuning, the proposed learnable nonlinear mapping, the generalization ability, the computational burden, and the network structure, are provided.

TABLE VII

QUANTITATIVE RESULTS (MEAN AND STANDARD DEVIATION) FOR ALL THE COMPARED APPROACHES INCLUDED IN THE BENCHMARK ON 31 IMAGES FROM THE FULL-RESOLUTION RIO DATASET (SOURCE: WORLDVIEW-3). THE CNN-BASED APPROACHES ARE RUN ON GPU (G), WHEREAS THE OTHERS ON CPU (C). (BOLD: BEST RESULTS; UNDERLINE: SECOND BEST RESULTS)

Method	D_λ	D_S	QNR	Runtime(s)
EXP [8]	0.007 ± 0.007	0.068 ± 0.026	0.925 ± 0.026	0.028(C)
BT-H [9]	0.129 ± 0.042	0.214 ± 0.055	0.687 ± 0.074	<u>0.056(C)</u>
BD-PC [61]	0.104 ± 0.043	0.218 ± 0.062	0.702 ± 0.084	0.068(C)
HPM [10]	0.158 ± 0.072	0.220 ± 0.092	0.663 ± 0.123	0.185(C)
C-GLP [62]	0.045 ± 0.019	0.076 ± 0.042	0.884 ± 0.054	0.991(C)
LGC [26]	0.008 ± 0.004	<u>0.028 ± 0.014</u>	<u>0.964 ± 0.014</u>	28.151(C)
CDIF [24]	0.061 ± 0.059	0.065 ± 0.023	0.878 ± 0.063	40.174(C)
PNN [38]	0.122 ± 0.047	0.233 ± 0.067	0.676 ± 0.088	0.314(G)
APNN [63]	0.038 ± 0.036	0.073 ± 0.028	0.892 ± 0.046	2.543(G)
LNM-PS	0.005 ± 0.004	0.027 ± 0.007	0.969 ± 0.007	5.212(C)
Ideal value	0	0	1	0

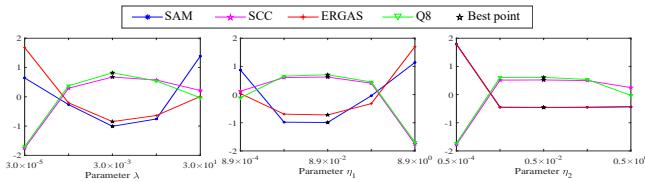


Fig. 14. The robustness analysis using four image quality indexes as metrics (including the SAM, the SCC, the ERGAS, and the Q8) for the parameters λ , η_1 , and η_2 on the reduced-resolution Rio dataset (source: WorldView-3). For convenience, the black stars are used to point out the best points. It is noteworthy that for better comparison, we standardize the used quality indexes as $(\text{index} - \text{mean}(\text{index})) / \text{std}(\text{index})$.

1) *Parameters Tuning*: From the created LNM-PS model (10), we can see that three key parameters are concerned, i.e., λ , η_1 , and η_2 . Fig. 14 depicts the performance when we vary the above three parameters on the reduced-resolution Rio dataset. It is worth noting that we fix all the parameters except the one to be analyzed in this experiment, the goal is to simplify the selection step of these parameters. It can be observed that $\lambda = 3.0 \times 10^{-3}$, $\eta_1 = 8.9 \times 10^{-2}$, and $\eta_2 = 0.5 \times 10^{-2}$ are the best choices whatever the quality metric that we adopted. Thus, we used this configuration for our approach.

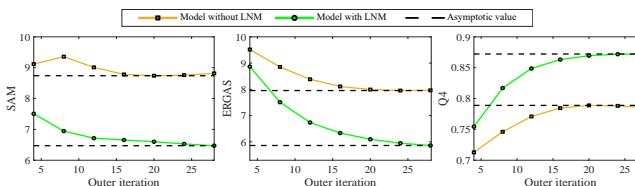


Fig. 15. The performance over the ADMM iterations for the models that use the linear operator and LNM, respectively, on the reduced-resolution Indianapolis dataset (source: QuickBird). It is worth noting that for fairness contrast, the optimal parameter configurations for both models are used.

2) *Analysis About Learnable Nonlinear Mapping*: To assess the improvement in using LNM, we illustrate in Fig. 15 the curves over the ADMM iterations of the several image quality indexes, i.e., SAM, ERGAS, and Q4, on the reduced-

resolution Indianapolis dataset (source: QuickBird) using nonlinear and the linear operator, respectively. The advantages when using LNM are clear. Moreover, it is clear that our fusion framework always gets high performance when the number of outer iterations is fixed at 30.

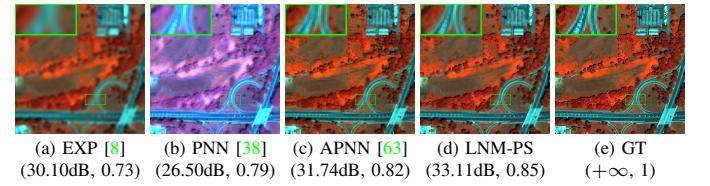


Fig. 16. The visual fusion results on the reduced-resolution Stockholm dataset (source: WorldView-2), where the involved ML-based methods are trained on the WorldView-3 dataset. (\cdot, \cdot) denotes two image quality indexes of the corresponding images, i.e., PSNR and SSIM. The NIR (7th), Green (3rd), and Blue (2nd) channels are selected to generate these Pseudo color images.

3) *Generalization Ability*: To test the generalization of the trained functional operator, i.e., Φ^{-1} , we process the reduced-resolution Stockholm dataset (source: WorldView-2) using the CNN trained on the reduced-resolution Rio dataset (source: WorldView-3). The visual and quantitative comparison of the results on this reduced-resolution Stockholm dataset is shown in Fig. 16. The 7th, 3rd, and 2nd channels are selected to generate a pseudo-color representation of the fused products. Moreover, the PSNR and SSIM metrics are calculated for a quantitative evaluation. It is easy to see that the functional operator obtained by the CNN and trained on a vast amount of data has a better generalization ability with respect to the two compared ML-based methods, i.e., PNN [38] and APNN [63], thus alleviating the generalization ability problem of methods belonging to the ML family.

4) *Computational Burden*: The total computational burden of our method is roughly divided into two parts, i.e., VO-part and ML-part. We performed a first computational analysis depending on several different PAN sizes, as shown in Fig. 17. Moreover, we further explored the computational burden of Algorithm 2 focusing on each subproblem, with four different patch sizes of the PAN image, as depicted in Fig. 18. It is

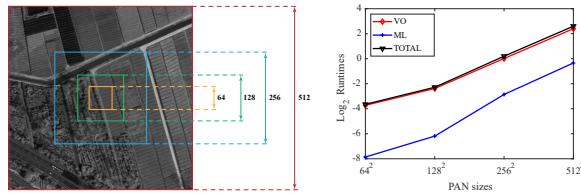


Fig. 17. Computational analysis of the proposed approach on the full-resolution Guangzhou dataset (source: GaoFen-2). Left: The PAN image cropped into patches with multiple sizes. Right: \log_2 Runtimes considering different PAN sizes.

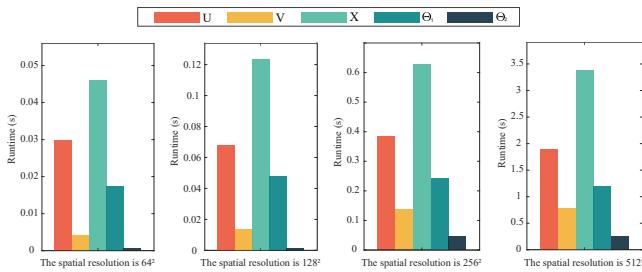


Fig. 18. The computation load of the corresponding five subproblems in Algorithm 2 evaluated on the full-resolution Guangzhou dataset (source: GaoFen-2).

clear that the subproblem **X**, which is related to the FFT, is the most demanding from a computational point of view.

5) *Analysis of the network structure:* In this paper, we design a simple network based on the PNN [38], rather than other more recent and elaborated networks, to obtain the inverse operator of Φ , i.e., Φ^{-1} . The purpose of this choice is to show that the advantages of our model are obtained even with the exploitation of a simple network to generate Φ^{-1} . Anyway, the obtained results of the proposed LNM-PS are competitive even with some popular, more recent and elaborated networks. For instance, considering Fusion-net [45] and LPPN [73], the corresponding fusion results are illustrated in Fig. 19 comparing the proposed approach with Fusion-net and LPPN using six image quality metrics and on the reduced-resolution Guangzhou and Indianapolis datasets obtained by the GaoFen-2 and QuickBird sensors, respectively. Notably, the quality indexes are normalized (i.e., $\text{index}/\max(\text{index})$) to get a maximum value equal to 1 for a better visual inspection. It is easy to see from Fig. 19 that the results of the proposed LNM-PS are just slightly less than Fusion-net on the Guangzhou dataset, while the advantages of LNM-PS are clear on the Indianapolis dataset. When it comes to LPPN, we can see that the corresponding standard deviations of the chosen six image quality indexes are relatively unstable, especially in SCC and Q4 about the Indianapolis datasets. Besides, another advantage between the proposed LNM-PS with the existing ML-based methods is the trained nonlinear operators are suitable for all datasets from the same sensor, while the existing ML-based approaches always need lots of time to re-train the parameters of the designed network. Hence, the use of more powerful and recent ML approaches in our framework could help in getting better results, but this analysis and the exploitation of different ML approaches in our methodology are left for future developments, thus promoting in this paper a lightweight

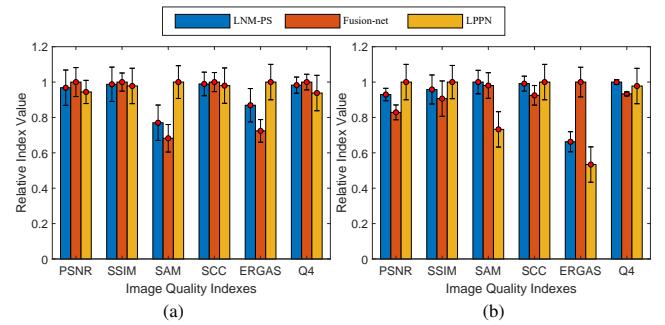


Fig. 19. The normalized quality metrics on (a) the Guangzhou and (b) Indianapolis datasets.

method with high performance.

VI. CONCLUSIONS

In this paper, a novel hybrid framework that combines VO-based with ML-based approaches has been proposed to achieve pansharpening, thus expecting to obtain a greater model interpretability, higher performance, and a better generalization ability. The proposed framework exploits a new learnable nonlinear mapping (LNM) as a spatial fidelity term. The introduction of a simple yet effective pansharpening model only including two data fidelity terms represents a further contribution of this paper. Furthermore, an ADMM-based algorithm with the FISTA framework as the inner solver has been designed to provide a solution for the proposed model. In particular, we derived the nonlinear function into linear functions by further exploring the properties of the trained nonlinear function in some special points, thus reducing the computational burden of the proposed approach. Both qualitative and quantitative comparisons on reduced-resolution and full-resolution datasets have been considered to assess the superiority of our LNM-PS with respect to several state-of-the-art techniques. Finally, a robustness analysis, the analysis of the learnable nonlinear mapping, the generalization ability, the computational burden, and the designed network structure have been proposed to the readers.

REFERENCES

- [1] T. Xu, T.-Z. Huang, L.-J. Deng, and N. Yokoya, "An iterative regularization method based on tensor subspace representation for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, 2022, doi: [10.1109/TGRS.2022.3176266](https://doi.org/10.1109/TGRS.2022.3176266).
- [2] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, J. Huang, J. Chanussot, and G. Vivone, "LRTCPan: Low-rank tensor completion based framework for pansharpening," *IEEE Transactions on Image Processing*, 2023, doi: [10.1109/TIP.2023.3247165](https://doi.org/10.1109/TIP.2023.3247165).
- [3] L. He, D. Xi, J. Li, H. Lai, A. Plaza, and J. Chanussot, "Dynamic hyperspectral pansharpening CNNs," *IEEE Transactions on Geoscience and Remote Sensing*, 2023, doi: [10.1109/TGRS.2023.3250621](https://doi.org/10.1109/TGRS.2023.3250621).
- [4] Y. Shi, A. Tan, N. Liu, W. Li, R. Tao, and J. Chanussot, "A pansharpening method based on hybrid-scale estimation of injection gains," *IEEE Transactions on Geoscience and Remote Sensing*, 2023, doi: [10.1109/TGRS.2023.3241111](https://doi.org/10.1109/TGRS.2023.3241111).
- [5] Y. Yang, L. Wu, S. Huang, W. Wan, W. Tu, and H. Lu, "Multiband remote sensing image pansharpening based on dual-injection model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1888–1904, 2020.
- [6] Y. Yang, H.-Y. Lu, S.-Y. Huang, and W. Tu, "Remote sensing image fusion based on fuzzy logic and salience measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 11, pp. 1943–1947, 2020.

- [7] Y.-W. Zhuo, T.-J. Zhang, J.-F. Hu, H.-X. Dou, T.-Z. Huang, and L.-J. Deng, "A deep-shallow fusion network with multidetail extractor and spectral attention for hyperspectral pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7539–7555, 2022.
- [8] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, 2002.
- [9] S. Lolli, L. Alparone, A. Garzelli, and G. Vivone, "Haze correction for contrast-based multispectral pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2255–2259, 2017.
- [10] 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," *Proc. 22nd Digit. Avionics Syst. Conf.*, pp. 90–94, 2003.
- [11] 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," *Inform. Fusion*, vol. 46, pp. 102–113, 2019.
- [12] G. Vivone, M. Dalla Mura, A. Garzelli, R. Restaino, G. Scarpa, M. O. Ulfarsson, L. Alparone, and J. Chanussot, "A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 1, pp. 53–81, 2021.
- [13] G. Vivone, M. Dalla Mura, A. Garzelli, and F. Pacifici, "A benchmarking protocol for pansharpening: Dataset, preprocessing, and quality assessment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6102–6118, 2021.
- [14] P. Chavez, Jr and A. Kwarteng, "Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, pp. 339–348, 1989.
- [15] C.-A. Laben and B.-V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," no. 6011875, January 2000.
- [16] 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, 2008.
- [17] Q.-Z. Xu, B. Li, Y. Zhang, and L. Ding, "High-fidelity component substitution pansharpening by the fitting of substitution data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7380–7392, 2014.
- [18] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, 1983.
- [19] R. Restaino, G. Vivone, M. Dalla 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, 2016.
- [20] G. Vivone, R. Restaino, and J. Chanussot, "Full scale regression-based injection coefficients for panchromatic sharpening," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3418–3431, 2018.
- [21] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 930–934, 2014.
- [22] 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," *Inform. Fusion*, vol. 52, pp. 76–89, 2019.
- [23] L.-J. Deng, G. Vivone, W.-H. Guo, M. Dalla Mura, and J. Chanussot, "A variational pansharpening approach based on reproducible kernel hilbert space and heaviside function," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4330–4344, 2018.
- [24] J.-L. Xiao, T.-Z. Huang, L.-J. Deng, Z.-C. Wu, and G. Vivone, "A new context-aware details injection fidelity with adaptive coefficients estimation for variational pansharpening," *IEEE Trans. Geosci. Remote Sens.*, 2022, doi: [10.1109/TGRS.2022.3154480](https://doi.org/10.1109/TGRS.2022.3154480).
- [25] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for P+XS image fusion," *Int. J. Comput. Vis.*, vol. 69, no. 1, pp. 43–58, 2006.
- [26] X.-Y. Fu, Z.-H. Lin, Y. Huang, and X.-H. Ding, "A variational pan-sharpening with local gradient constraints," *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 10257–10266, 2019.
- [27] H. Li, W. Li, G. Han, and F. Liu, "Coupled tensor decomposition for hyperspectral pansharpening," *IEEE Access*, vol. 6, pp. 34206–34213, 2018.
- [28] J.-F. Hu, T.-Z. Huang, L.-J. Deng, H.-X. Dou, D.-F. Hong, and G. Vivone, "Fusformer: A transformer-based fusion network for hyperspectral image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, 2022, doi: [10.1109/LGRS.2022.3194257](https://doi.org/10.1109/LGRS.2022.3194257).
- [29] H. Lu, Y. Yang, S. Huang, X. Chen, B. Chi, A. Liu, and W. Tu, "AWFLN: An adaptive weighted feature learning network for pansharpening," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023, doi: [10.1109/TGRS.2023.3241643](https://doi.org/10.1109/TGRS.2023.3241643).
- [30] S.-Q. Deng, L.-J. Deng, X. Wu, R. Ran, D. Hong, and G. Vivone, "PSRT: Pyramid shuffle-and-reshuffle transformer for multispectral and hyperspectral image fusion," *IEEE Transactions on Geoscience and Remote Sensing*, 2023, doi: [10.1109/TGRS.2023.3244750](https://doi.org/10.1109/TGRS.2023.3244750).
- [31] R. Ran, L.-J. Deng, T.-X. Jiang, J.-F. Hu, J. Chanussot, and G. Vivone, "GuidedNet: A general CNN fusion framework via high-resolution guidance for hyperspectral image super-resolution," *IEEE Transactions on Cybernetics*, 2023, doi: [10.1109/TCYB.2023.3238200](https://doi.org/10.1109/TCYB.2023.3238200).
- [32] X.-Y. Cao, J. Yao, Z.-B. Xu, and D.-Y. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4604–4616, 2020.
- [33] D.-F. Hong, L.-R. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, 2021.
- [34] D.-F. Hong, L.-R. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, 2021.
- [35] X.-Y. Cao, X.-Y. Fu, C. Xu, and D.-Y. Meng, "Deep spatial-spectral global reasoning network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, doi: [10.1109/TGRS.2021.3069241](https://doi.org/10.1109/TGRS.2021.3069241).
- [36] Q. Zhang, Q.-Q. Yuan, M.-P. Song, H.-Y. Yu, and L.-P. Zhang, "Co-operated spectral low-rankness prior and deep spatial prior for hsi unsupervised denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 6356–6368, 2022.
- [37] Y. Chen, W. He, X.-L. Zhao, T.-Z. Huang, J. Zeng, and H. Lin, "Exploring nonlocal group sparsity under transform learning for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, 2022, doi: [10.1109/TGRS.2022.3202359](https://doi.org/10.1109/TGRS.2022.3202359).
- [38] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, p. 594, 2016.
- [39] Q.-J. Liu, H.-Y. Zhou, Q.-Z. Xu, X.-Y. Liu, and Y.-H. Wang, "PS-GAN: A generative adversarial network for remote sensing image pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10227–10242, 2021.
- [40] K. Zhang, A. Wang, F. Zhang, W. Diao, J. Sun, and L. Bruzzone, "Spatial and spectral extraction network with adaptive feature fusion for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, 2022, doi: [10.1109/TGRS.2022.3187025](https://doi.org/10.1109/TGRS.2022.3187025).
- [41] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, J.-F. Hu, and G. Vivone, "VO+Net: An adaptive approach using variational optimization and deep learning for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, 2022, doi: [10.1109/TGRS.2021.3066425](https://doi.org/10.1109/TGRS.2021.3066425).
- [42] R.-W. Dian, S.-T. Li, and X.-D. Kang, "Regularizing hyperspectral and multispectral image fusion by CNN denoiser," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1124–1135, 2021.
- [43] W.-Y. Xie, J. Lei, Y.-H. Cui, Y.-S. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1529–1543, 2020.
- [44] J. Qu, W. Dong, Y. Li, S. Hou, and Q. Du, "An interpretable unsupervised unrolling network for hyperspectral pansharpening," *IEEE Transactions on Cybernetics*, pp. 1–14, 2023, doi: [10.1109/TCYB.2023.3241165](https://doi.org/10.1109/TCYB.2023.3241165).
- [45] L.-J. Deng, G. Vivone, C. Jin, and J. Chanussot, "Detail injection-based deep convolutional neural networks for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6995–7010, 2021.
- [46] F. Ye, Y.-C. Guo, and P.-X. Zhuang, "Pan-sharpening via a gradient-based deep network prior," *Signal Processing: Image Communication*, vol. 74, pp. 322–331, 2019.
- [47] P.-H. Guo, P.-X. Zhuang, and Y.-C. Guo, "Bayesian pan-sharpening with multiorder gradient-based deep network constraints," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 950–962, 2020.
- [48] G. Vivone, "Multispectral and hyperspectral image fusion in remote sensing: A survey," *Inf. Fusion*, vol. 89, pp. 405–417, 2023.
- [49] P. Gemma, "Image fusion for enhanced visualization: A variational approach," *Int. J. Comput. Vis.*, vol. 83, no. 1, pp. 1–11, 2009.
- [50] T.-J. Zhang, L.-J. Deng, T.-Z. Huang, J. Chanussot, and G. Vivone, "A triple-double convolutional neural network for panchromatic sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, 2022, doi: [10.1109/TNNLS.2022.3155655](https://doi.org/10.1109/TNNLS.2022.3155655).
- [51] X.-Y. Fu, W. Wang, Y. Huang, X.-H. Ding, and J. Paisley, "Deep multiscale detail networks for multiband spectral image sharpening,"

- IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 2090–2104, 2021.
- [52] P.-X. Zhuang, Q.-S. Liu, and D. Xing-Hao, “Pan-GGF: A probabilistic method for pan-sharpening with gradient domain guided image filtering,” *Signal Process.*, vol. 156, pp. 177–190, 2019.
- [53] P.-H. Guo, P.-X. Zhuang, and Y.-C. Guo, “Bayesian pan-sharpening with multiorder gradient-based deep network constraints,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 950–962, 2020.
- [54] W. Diao, F. Zhang, H. Wang, W. Wan, J. Sun, and K. Zhang, “HLF-Net: Pansharpening based on high- and low-frequency fusion networks,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [55] N. Wang, X. Meng, X. Meng, and F. Shao, “Convolution-embedded vision transformer with elastic positional encoding for pansharpening,” *IEEE Transactions on Geoscience and Remote Sensing*, 2022, doi: [10.1109/TGRS.2022.3227405](https://doi.org/10.1109/TGRS.2022.3227405).
- [56] W. Zhu, J. Li, Z. An, and Z. Hua, “Mutiscale hybrid attention transformer for remote sensing image pansharpening,” *IEEE Transactions on Geoscience and Remote Sensing*, 2023, doi: [10.1109/TGRS.2023.3239013](https://doi.org/10.1109/TGRS.2023.3239013).
- [57] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, G. Vivone, J.-Q. Miao, J.-F. Hu, and X.-L. Zhao, “A new variational approach based on proximal deep injection and gradient intensity similarity for spatio-spectral image fusion,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6277–6290, 2020.
- [58] 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, 2010.
- [59] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
- [60] R. Glowinski and P. Le Tallec, “Augmented lagrangian and operator-splitting methods in nonlinear mechanics,” *Philadelphia, PA, USA: SIAM*, pp. 287–295, 1989.
- [61] G. Vivone, “Robust band-dependent spatial-detail approaches for panchromatic sharpening,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6421–6433, 2019.
- [62] G. Vivone, S. Marano, and J. Chanussot, “Pansharpening: Context-based generalized laplacian pyramids by robust regression,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6152–6167, 2020.
- [63] G. Scarpa, S. Vitalé, and D. Cozzolino, “Target-adaptive CNN-based pansharpening,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, 2018.
- [64] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, “Contrast and error-based fusion schemes for multispectral image pansharpening,” *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 930–934, 2014.
- [65] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, “A critical comparison among pansharpening algorithms,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, 2014.
- [66] W. Zhou, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [67] A. F. G. Roberta H. Yuhas and J. W. Boardman, “Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm,” *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, vol. 1, pp. 147–149, 1992.
- [68] X. Oztuz, 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, 2005.
- [69] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, “Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, 2007.
- [70] A. Garzelli and F. Nencini, “Hypercomplex quality assessment of multi/hyperspectral images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 662–665, 2009.
- [71] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, “A global quality measurement of pan-sharpened multispectral imagery,” *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 313–317, 2004.
- [72] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, “Multispectral and panchromatic data fusion assessment without reference,” *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, 2008.
- [73] C. Jin, L.-J. Deng, T.-Z. Huang, and G. Vivone, “Laplacian pyramid networks: A new approach for multispectral pansharpening,” *Inform. Fusion*, vol. 78, pp. 158–170, 2022.