Machine Learning in Pansharpening

A benchmark, from shallow to deep networks

LIANG-JIAN DENG, GEMINE VIVONE, MERCEDES E. PAOLETTI, GIUSEPPE SCARPA, JIANG HE, YONGJUN ZHANG, JOCELYN CHANUSSOT, AND ANTONIO PLAZA

xxxxx

Machine learning (ML) is influencing the literature in several research fields, often through state-of-the-art approaches. In the past several years, ML has been explored for pansharpening, i.e., an image fusion technique based on the combination of a multispectral (MS) image, which is characterized by its medium/low spatial resolution, and higher-spatial-resolution panchromatic (PAN) data. Thus, ML for pansharpening represents an emerging research line that deserves further investigation. In this article, we go through some powerful and widely used ML-based approaches for pansharpening that have been recently proposed in the related literature. Eight approaches are extensively compared. Implementations of these eight methods, exploiting a common software platform and ML library, are developed for comparison purposes. The ML framework for pansharpening will be freely distributed to the scientific community. Experimental results using data acquired by five commonly used sensors for pansharpening and well-established protocols for performance assessment (both at reduced resolution and at full resolution) are shown. The ML-based approaches are compared with a benchmark consisting of classical and variational optimization (VO)-based methods. The pros and cons of each pansharpening technique, based on the training-by-examples philosophy, are reported together with a broad computational analysis. The toolbox is provided in https://github.com/liangjiandeng/DLPan-Toolbox.

OVERVIEW

Pansharpening is the process of combining an MS image with a PAN image to produce an output that holds the same spatial resolution as the PAN image and the same spectral resolution as the MS image. To date, several techniques for this have been proposed. With the development of new hardware and software solutions, ML approaches, especially deep learning-based (DL) frameworks, have been significantly developed. However, a fair comparison of these techniques (including, for instance, their development on the same software platform using the same libraries, testing on data sets simulated in a conventional way, and so forth) is still an open issue. To this end, in this article, we go through shallow to deep networks based on widely used and powerful ML-based pansharpening approaches. In addition, traditional approaches, belonging to component substitution (CS), multiresolution analysis (MRA), and VO, are compared and discussed. A quantitative and qualitative assessment is presented in the "Experimental Results" section, exploiting protocols at reduced

Digital Object Identifier 10.1109/MGRS.2022.3187652
Date of current version: 21 July 2022

resolution and at full resolution. All the compared ML-based techniques are implemented using PyTorch. The source code will be freely distributed to the community (https://github.com/liangjiandeng/DLPan-Toolbox). The ML framework for pansharpening uses a uniform programming style to facilitate interpretability for users.

BACKGROUND AND RELATED WORKS

Recently, some books [1] and review articles [2]–[4] about pansharpening have been published, attesting to pansharpening's key role in the field of remote sensing image fusion. In addition, in recent years, several other surveys have been published, such as [5]–[7], confirming the increased interest in this area. Many techniques have been applied to the task of remote sensing pansharpening. They are usually divided into four classes [4], i.e., CS, MRA, VO, and ML. In this article, we consider the first three classes to be traditional methods since their first approaches were proposed a long time ago. Meanwhile, several works in the related literature, such as [2] and [4], have deeply analyzed these categories. The remaining methods, belonging to the ML class, are further investigated in this article. In the rest of this section, we go through the four main categories of pansharpening algorithms, introducing the related literature.

COMPONENT SUBSTITUTION

CS approaches (also called *spectral methods*) rely on the projection into a transformed domain of the original MS image to separate its spatial information and substitute it with the PAN image. Many pioneering pansharpening techniques belong to the CS class, thanks to their easy implementation. Two examples of CS approaches, proposed in the early 1990s, are intensity–hue–saturation [8], [9] and principal component analysis (PCA) [10], [11].

By considering various image transformations, a variety of techniques for incorporating PAN spatial information into original MS data have been developed. These methods are usually viewed as the second generation of CS techniques, mainly improving the injection rules by investigating the relationship between the pixel values of the PAN image and those of the MS channels. Representative approaches are the Gram–Schmidt (GS) method [12] and its adaptive version [13], nonlinear PCA [14], and partial replacement adaptive CS [15]. Beyond these CS strategies, some other recent approaches are based on 1) the local application of CS algorithms and 2) the joint estimation of detail injection and estimation coefficients. The former subclass mainly focuses on sliding widow-based methods [13] and approaches relying on clustering and segmentation [16], whereas the latter one includes band-dependent spatial detail (BDSD) methods (see, e.g., BDSD [17] and its robust version [18]).

MULTIRESOLUTION ANALYSIS

MRA methods apply a multiscale decomposition to the PAN image to extract its spatial components. This class is also referred to as *spatial methods*, as they work in the spatial domain. General-purpose decompositions have been considered in the pansharpening literature, including, for instance, Laplacian pyramids [19], wavelets [20], curvelets [21], and contourlets [22]. MRA-based fusion techniques present interesting features, such as temporal coherence [23], spectral consistency [2], and robustness to aliasing [24], thus deserving further investigation.

Recently, researchers have considered various decomposition schemes and several ways to optimize the injection model to improve MRA-based methods. Due to their superior performance in other image processing fields, nonlinear methods have been introduced into pansharpening; typical examples are least-squares support vector machines [25] and morphological filters [26]. Moreover, thanks to an in-depth analysis of the relationship among the obtained images [27], [28] and the influence of the atmosphere on the collected signals, a series of advanced injection models have been designed [27], [29], [30]. A further crucial step forward has been the introduction of information about the acquisition sensors, thus driving the decomposition phase [24], [31]. This symbolized the beginning of the second generation of MRA-based pansharpening. The application of adaptive techniques has been proposed to deal with unknown and difficult-to-predict features about acquisition sensors [32], [33] and to address the peculiarities of some target images [34].

Hybrid technologies combining MRA and CS methods (see, e.g., [4]) have also been proposed. They can be regarded as MRA approaches [24]. Within this category, two varieties have been considered, i.e., "MRA + CS" (MRA followed by CS) [35] and "CS + MRA" (CS followed by MRA) [27], [36]. Other notable examples in this subclass include the use of independent component analysis in combination with curvelets [37] and the use of PCA with contourlets [38] and guided filters [39].

VARIATIONAL OPTIMIZATION

The class of VO methods focuses on the solution of optimization models. In recent years, VO methods have become more and more popular thanks to the advances in convex optimization and inverse problems, such as MS pansharpening [40]–[44] and hyperspectral image fusion [45]–[47]. Most VO methods focus on the relationship between the input PAN image, the low-spatial-resolution MS (LRMS) image, and the desired high-spatial-resolution MS (HRMS) image to generate the corresponding model. However, the problem to be solved is clearly ill posed, thus requiring some regularizers introducing prior information about the solution (i.e., the HRMS). The target image is usually estimated under the assumption of proper coregistered PAN and LRMS images. Anyway, some papers (see, e.g., [48]) have been proposed to deal with registration issues.

The timeline of VO techniques starts in 2006, with the so-called panchromatic (P) + multispectral (XS) method [49]. Inspired by P + XS, researchers have proposed various regularization terms [50], [51] and new fidelity terms [52]–[54]. In [55], the authors indirectly model the connection between PAN and HRMS images by considering the spectral low-rank relationship between them. Apart from P + XS-like methods, other approaches belonging to the VO class mainly include Bayesian methods [56]–[59] and sparse representations [60]–[67].

MACHINE LEARNING

ML-based methods have shown great ability in fusing MS and PAN data, thanks to the recent advances in computer hardware and algorithms. Classical ML approaches mainly include dictionary learning methods [62]–[65] and compressed sensing techniques [60], [61]. Compressed sensing concerns acquiring and reconstructing a signal by efficiently solving underdetermined linear systems. The sparsity of a signal can be utilized to recover the signal through proper optimization, even with considerably fewer samples than the ones required by the Nyquist–Shannon sampling theorem. The mainstream perspective based on compressive sensing pansharpening views the linear observation models (both the one focused on LRMS and the one related to the PAN) as a measurement process in compressive sensing theory, then building effective and efficient algorithms to solve the related models under the sparsity assumption. Dictionary learning, a special representation strategy, is mainly based on sparse coding to find a sparse linear representation from the input data, forming a so-called dictionary matrix and the corresponding coefficients. The main idea of dictionary learning for pansharpening is to calculate (trained and untrained) dictionaries of LRMS and PAN images, then reconstruct the final HRMS pansharpened image by investigating the relationship between dictionaries and the corresponding coefficients.

Recently, DL techniques have swept across almost all the applications in remote sensing imaging, including MS pansharpening [68]–[83] and some closely related tasks such as remote sensing image superresolution (SR) [84]–[86] and hyperspectral image fusion [87]–[89]. The first work using a DL technique for pansharpening, dating to 2015, by Huang et al. [68], employed and modified an autoencoder scheme inspired by the sparse denoising task. In 2016, Masi et al. [69] built and trained the first fully convolutional neural network (CNN) for pansharpening, also called the *pansharpening NN* (*PNN*). The architecture mainly consists of three convolutional layers and is inspired by the SR CNN [90], whose task concerns the single image SR problem. Meanwhile, Zhong et al. [70], in 2016, proposed a new CS pansharpening method based on the GS transform, in which a commercially available SR CNN was exploited to up-sample the MS component.

Following these pioneering approaches, this topic received the interest of many researchers, as testified to by numerous publications, such as [72], [76], [77], and [81]–[83]. Thus, the use of CNNs has become a common choice for DL-based pansharpening. Unlike the PNN, which has a simple network architecture, the later pansharpening architectures have been deepened and widened, receiving more and more complex structures with many parameters to fit during the training phase to obtain superior performance. These methods can be found in [71], [75], and [79]. Another research line using residual learning has been developed to effectively alleviate the phenomenon of gradient vanishing and explosion, thus accelerating network convergence. Hence, residual learning has been widely applied to pansharpening; see, e.g., [73], [91], and [92]. A weak generalization ability of ML-based approaches can easily be observed, representing a key issue. Therefore, another research line is working toward the development and design of new network architectures and preprocessing operators to improve ML approach generalization; see, e.g., [73] and [74].

In addition to the preceding DL methods, hybrid methods to combine traditional techniques (e.g., CS, MRA, and VO methods) and ML approaches have recently become a promising direction in the field of remote sensing pansharpening; see, e.g., [47] and [92]–[100]. For example, in [92], motivated by avoiding linear injection models and replacing the detail injection phases in both CS and MRA methods, Deng et al. design a deep CNN, inspired by the CS and MRA schemes, to effectively manage the nonlinear mapping and image extraction features, thus yielding favorable performance. Moreover, with the development of DL and VO techniques, the literature is presenting combinations of these two classes. Three strategies have been developed: the unfolding VO model [97], the plug-and-play operator [93], and the VO+Net mixed model [47], [96], which can also be viewed as belonging to the VO class.

The outcomes of these latter approaches can benefit from the advantages of DL and VO classes, e.g., the good generalization ability of VO methods and the high performance of DL approaches. Specifically, in [94], Shen et al. incorporate the pansharpened outcomes learned from a DL model into a VO framework. This strategy is simple but quite effective in practical applications. Xie et al., in [95], use a strategy similar to [94] for the task of hyperspectral pansharpening, also producing promising outcomes. Differing from the strategy in [94] and [95], new DL network architectures propose to unfold traditional VO models. In [97], Feng et al. present a two-step optimization model based on spatial detail decomposition, then unfold the model under the gradient descent framework to further construct the corresponding end-to-end CNN architecture. Similar to [97], Xu et al., in [98], propose a model-driven deep pansharpening network by gradient projection. Specifically, two optimization problems regularized by the deep prior are formulated. The two problems are solved by a gradient projection algorithm in which the iterative steps are constructed by two network blocks that will be effectively trained in an end-to-end manner. Moreover, Cao et al., in [99], and Yin et al., in [100], present sparse coding-based strategies to unfold optimization models into subproblems that are replaced by learnable networks.

Recently, unsupervised learning strategies have been introduced into the field of pansharpening; see, e.g., [101]–[103]. Unsupervised learning explores hidden patterns and features without any labeled data, which means that there is no need to simulate data sets with labels for training. It is a direct approach to network training but strongly dependent on the effectiveness of the loss function. In [101], Ma et al. propose a novel unsupervised pansharpening approach that can avoid the degrading effect of downsampling high-resolution MS images. The technique also considers a generative adversarial network (GAN) strategy, yielding excellent results, in particular, on full-resolution data. Furthermore, Qu et al., in [103], present a self-attention mechanism-based unsupervised learning technique for pansharpening. This can address some challenges, e.g., poor performance on full-resolution images and the wide presence of mixed pixels. In [104], leveraging the target-adaptive strategy introduced in [74], Ciotola et al. present an unsupervised full-resolution training framework, demonstrating its effectiveness on different CNN architectures [71], [73], [74].

GAN techniques [105] have recently been applied to the field of image processing. GANs mainly concern learning generative models via an adversarial process; thus, two models are required to be trained simultaneously, i.e., generative models to capture a data distribution and adversarial models to compute the probability of a sample belonging to training data. GANs have been applied to the task of pansharpening; see, e.g., [78], [101], and [106]–[110]. In [78], Liu et al. utilize a GAN to address the task of remote sensing pansharpening. This method mainly contains a two-stream fusion architecture consisting of a generator to produce the desired HRMS image and a discriminator to judge whether the image is real or pansharpened. In [110], to further boost accuracy, the authors propose a GAN-based pansharpening framework containing two discriminators, the first dealing with image textures and the second accounting for image color.

Table 1 gives an overview of the four classes, focusing on aspects such as spatial fidelity, spectral fidelity, generalization, running time, and model interpretability. For example, it is easy to remark that ML methods generally get the best spatial and spectral performance but require training and testing data to have similar properties (e.g., a similar geographic area and acquisition time).

CONTRIBUTION

This article is focused on a deep analysis of the emerging class of pansharpening algorithms based on ML paradigms. A complete review of the related literature has been presented. Henceforth, the article provides a critical comparison of the state-of-the-art approaches belonging to the ML class. To this end, a toolbox exploiting a common software platform and open source ML library for all the ML approaches has been developed. We would like to stress that this is the only way to get a critical comparison of ML approaches. In fact, changing software platforms and ML libraries (e.g., TensorFlow and Caffe), results in different built-in functions, thus generating different behaviors (e.g., a different initialization of the weights of the network) of the same algorithm coded in a different environment.

To overcome this limitation, a Python toolbox based on the PyTorch ML library (which is widely used for applications such as computer vision and natural language processing) has been developed. The toolbox will be freely distributed to the scientific community related to ML and pansharpening. In this article, eight state-of-the-art approaches are selected and implemented in the common framework, following the original implementations proposed in related papers. A tuning phase to ensure the highest performance for each approach is performed. This represents a mandatory step to have a fair comparison because the eight approaches were originally developed on different software platforms and using different ML libraries. A broad experimental analysis, exploiting different test cases, is conducted with the aim of assessing the performance of each ML-based state-of-the-art approach.

Widely used sensors for pansharpening are involved [i.e., WorldView-2 (WV2), WorldView-3 (WV3), WorldView-4 (WV4), QuickBird (QB), and Ikonos]. Assessments at reduced resolution and at full resolution are exploited. Two test cases at reduced resolution are considered. The first concerns the use of a part of the training set not used for this aim. However, by taking into account a testing area very close to that used in the training phase, we have a sort of coupling among data (e.g., sharing features with the training samples, such as the atmospheric composition and conditions). Thus, to test the ability of the networks to work in a real scenario, we consider a second test case in which the images are acquired by the same sensor but over a different area and at a different time with respect to the data used for the training. The comparison of the ML-based approaches is also expanded to state-of-the-art methods belonging to different paradigms (i.e., CS, MRA, and VO), exploiting standard implementations [4]. Finally, a wide computational analysis is presented. Execution times for training and testing, convergence analysis, the number

TABLE 1. AN OVERVIEW OF THE PROS AND CONS OF THE FOUR PANSHARPENING CLASSES.

	CS	MRA	VO	ML
Spatial fidelity	⋆⋆	⋆	⋆⋆	⋆⋆⋆
Spectral fidelity	⋆	⋆⋆	⋆⋆	⋆⋆⋆
Generalization ability	⋆⋆⋆	⋆⋆⋆	⋆⋆	⋆
Running time	⋆⋆⋆	⋆⋆⋆	⋆	⋆⋆
Interpretability	⋆⋆⋆	⋆⋆⋆	⋆⋆⋆	⋆

Weak: ⋆; moderate: ⋆⋆; strong: ⋆⋆⋆.

of parameters, and so forth are highlighted. Moreover, the generalization ability of the networks with respect to the change of the acquisition sensor is discussed.

NOTATION

The notation is as follows. Vectors are indicated in bold lowercase (e.g., $\mathbf{x}$), with the ith element indicated as x_i. 2D and 3D arrays are expressed in bold uppercase (e.g., $\mathbf{X}$). An MS image $\mathbf{X} = \{\mathbf{X}_k\}_{k=1,\ldots,N}$ is a 3D array composed of N bands indexed by the subscript $k = 1,\ldots,N$; accordingly, $\mathbf{X}_k$ indicates the kth band of $\mathbf{X}$. A PAN image is a 2D matrix and is denoted as $\mathbf{P}$. $\mathbf{MS}$ is an MS image, $\widetilde{\mathbf{MS}}$ is an MS image up-sampled to the PAN scale, and $\widehat{\mathbf{MS}}$ is a fused image. Other symbols will be defined within the article as needed.

COMPONENT SUBSTITUTION, MULTIRESOLUTION ANALYSIS, AND VARIATIONAL OPTIMIZATION: A BRIEF OVERVIEW

In this section, we go through the CS, MRA, and VO categories, providing a brief overview of each class and instances of methods that are exploited in this article for comparison purposes. The methods belonging to the CS class rely on the projection of the MS image into a new space, where the spatial structure is separated from the spectral information [111]. Afterward, the transformed MS image can be sharpened by substituting the spatial component with the PAN image. Finally, the sharpening process is completed by the inverse transformation to return to the original space. CS methods obtain high fidelity in rendering details. Moreover, they are usually easy to implement and have a limited computational burden [2], [4].

Under the hypotheses of linear transformation and the substitution of a unique component, the CS fusion process can be simplified, obtaining faster implementation described by the following formulation [112]:

$$\widehat{\mathbf{MS}}_k = \widetilde{\mathbf{MS}}_k + \mathbf{G}_k \cdot (\mathbf{P} - \mathbf{I}_L), \quad (1)$$

in which $\widehat{\mathbf{MS}}_k$ is the kth fused band, $\widetilde{\mathbf{MS}}_k$ is the up-sampled image to the PAN scale, $\mathbf{P}$ is the PAN image, $\mathbf{G}_k$ is the injection gain matrix, the matrix multiplication is meant pointwise, and $\mathbf{I}_L$ is the so-called intensity component obtained by a weighted average of the MS spectral bands with weights w_k.

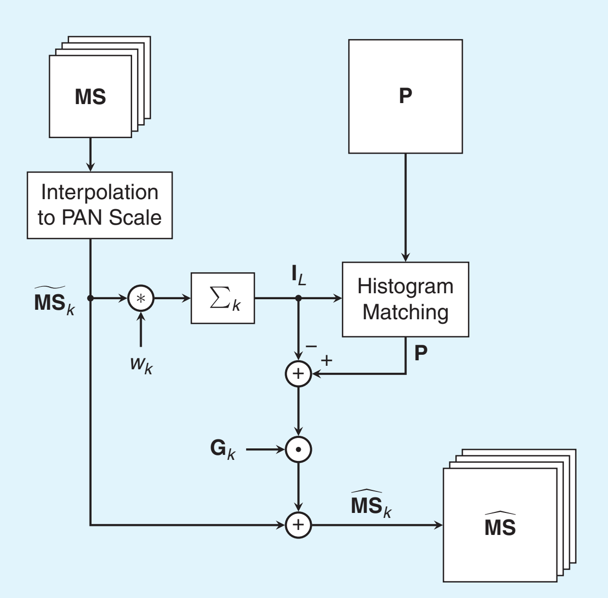


FIGURE 1. The CS-based methods.

Figure 1 describes the general fusion process for CS-based approaches. There are blocks related to the up-sampling, computation of $\mathbf{I}_L$, spectral matching between $\mathbf{P}$ and $\mathbf{I}_L$, and detail injection according to (1). Setting the injection gains in (1) as the pixel-wise division between $\widetilde{\mathbf{MS}}_k$ and $\mathbf{I}_L$, we have a multiplicative injection scheme, the widely known Brovey transform (BT) [113], [114]. An interpretation of the BT in terms of the radiative transfer model led to the development of a haze-corrected version, called *optimized BT with haze correction* (*BT-H*), recently proposed in [30]. The GS orthogonalization procedure has also been used for pansharpening [12]. This approach exploits the intensity component, $\mathbf{I}_L$ as the first vector of the new orthogonal basis. Pansharpening is obtained thanks to the substitution of $\mathbf{I}_L$ with the PAN image before inverting the transformation.

Several versions of GS are achieved by varying $\mathbf{I}_L$. The context-adaptive GSA (C-GSA) is obtained by separately applying an adaptive GS process (where $\mathbf{I}_L$ is obtained by a weighted average of the MS bands using proper weights [13]) to each cluster [16]. The BDSD framework, proposed for pansharpening in [17], exploits an extended version of (1) optimizing the minimum mean-square error for jointly estimating the weights and scalar gains [17]. A physically constrained (PC) optimization (i.e., the BDSD-PC) was recently proposed in [18]. MRA methods extract the PAN details, exploiting the difference between $\mathbf{P}$ and its low-pass spatial version, $\mathbf{P}_L$. The fused image is obtained as follows:

$$\widehat{\mathbf{MS}}_k = \widetilde{\mathbf{MS}}_k + \mathbf{G}_k \cdot (\mathbf{P} - \mathbf{P}_L). \quad (2)$$

These different approaches are characterized by the way in which they calculate $\mathbf{P}_L$ and estimate the injection gains $\mathbf{G}_k$. In a very general setting, $\mathbf{P}_L$ is achieved through an iterative decomposition scheme, MRA. The general fusion scheme is depicted in Figure 2. We observe blocks devoted to the up-sampling, calculation of the low-pass version $\mathbf{P}_L$ of the PAN image based on the resolution ratio R, and computation of the injection gains $\mathbf{G}_k$. MRA algorithms reduce spectral distortion but often result in greater spatial distortion [2], [4]. Among the MRA approaches, one much-debated subcategory is based on the generalized Laplacian pyramid (GLP). In this case, $\mathbf{P}_L$ can be performed with multiple fractional steps, utilizing Gaussian low-pass filters to carry out the analysis steps [19]. The corresponding differential representation is called the *Laplacian pyramid*.

However, high performance can be obtained with a single Gaussian low-pass filter tuned to closely match the MS sensor's modulation transfer function (MTF) [31], with a cut frequency equal to $1/R$ (where R is the resolution ratio between PAN and MS) and decimating by R [4]. In the literature, many instances of GLP approaches, relying on filters that exploit the MS sensor's MTF, have been proposed by changing the method to estimate the injection coefficients. In this article, we exploit high-pass modulation (HPM) injection [114], i.e., setting the injection gains as the pixel-wise division between $\widetilde{\mathbf{MS}}_k$ and $\mathbf{P}_L$, adopting a spectral matching procedure based on the multivariate linear regression between each MS band and a low-pass version of the PAN image, i.e., the MTF-GLP-HPM-R [115]. Moreover, we consider an MTF-GLP-full scale (FS) that is based on an FS fusion rule, thus removing the hypothesis of invariance among scales for the coefficient injection estimation phase [116].

The methods in the VO category rely on the definition of an optimization model. We exploit two instances of techniques belonging to the concepts of sparse representation and total variation (TV). In [66], an example of sparse representation for pansharpening is provided. In particular, the authors propose to generate the spatial details by using a dictionary of patches. Specifically, the dictionary $\mathbf{D}^h$ at FS is composed of patches representing high-spatial-resolution details. The coefficients $\boldsymbol{\alpha}$ of the linear combination are estimated by solving a sparse regression problem. Under the hypothesis of scale invariance, the coefficients can be estimated thanks to a reference image. The problem to solve is as follows:

$$\hat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{\alpha}\|_0 \quad \text{such that} \quad \mathbf{y} = \mathbf{D}^l \boldsymbol{\alpha}, \tag{3}$$

where $\mathbf{y}$ is a patch, $\|\cdot\|_0$ is the l_0 norm, and $\mathbf{D}^l$ is a dictionary of details at reduced resolution. The estimated coefficients are used for the representation of the full-resolution details (i.e., $\mathbf{y}^h = \mathbf{D}^h \boldsymbol{\alpha}$).

The cost function for the TV pansharpening method in [50] is given by the following TV-regularized least-squares problem:

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{M}\mathbf{x}\|^2 + \lambda \mathrm{TV}(\mathbf{x}), \tag{4}$$

where $\mathbf{y} = [\mathbf{y}_{\mathrm{MS}}^T, \mathbf{y}_{\mathrm{PAN}}^T]$, $\mathbf{y}_{\mathrm{MS}}$, and $\mathbf{y}_{\mathrm{PAN}}$ are the MS in a matrix format and the PAN in a vector; $\mathbf{M} = [\mathbf{M}_1^T, \mathbf{M}_2^T]$, $\mathbf{M}_1$ is a decimation matrix; $\mathbf{M}_2$ reflects that the PAN image is assumed to be a linear combination of the MS bands; λ is a weight; and $\mathrm{TV}(\cdot)$ is an isotropic TV regularizer. The pansharpened image $\mathbf{x}$ is obtained by minimizing the convex cost function in (4).

A BENCHMARK RELYING ON RECENT ADVANCES IN MACHINE LEARNING FOR PANSHARPENING

ML for pansharpening mainly relates to the DL philosophy, as pointed out in the "Background and Related Works" section. The approaches in this class strongly depend on the reduced-resolution training set (or the full-resolution one if belonging to the unsupervised paradigm). The testing data sets are exploited to get the network outcomes by using the trained models. In what follows, we choose eight representative supervised ML pansharpening approaches, i.e., the deep CNN architecture for pansharpening (PanNet) [73], deep residual NN for pansharpening (DRPNN) [71], multiscale and multidepth CNN architecture for pansharpening (MSDCNN) [75], bidirectional pansharpening network (BDPN) [79], detail injection-based CNN (DiCNN) [91], PNN [69], advanced PNN using fine-tuning (A-PNN-FT) [74], and pansharpening by combining ML and traditional fusion schemes (FusionNet) [92], for a fair and critical comparison using the same training and testing data. It is worth remarking that we did not select unsupervised learning and GAN-based methods for comparison purposes since they can require different training data sets (with respect to the used ones), invalidating the fair comparison.

DEEP CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE FOR PANSHARPENING

In [73], Yang et al. design a deep CNN architecture, the PanNet, for the task of pansharpening, relying on high-frequency information inputs from LRMS and PAN images. The PanNet architecture considers domain-specific knowledge and mainly focuses on preserving spectral and spatial information in remote sensing images. It first up-samples the LRMS image to the PAN scale, aiming to keep the spectral information. A deep residual network is employed to learn the spatial mapping to obtain

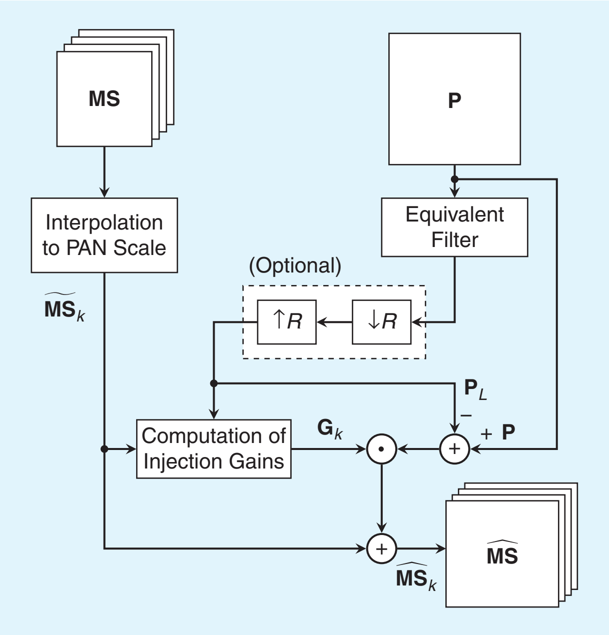

MS
P
Interpolation to PAN Scale
Equivalent Filter
(Optional)
$\widetilde{\mathbf{MS}}_k$
↑R
↓R
$\mathbf{P}_L$
−
$\mathbf{G}_k$
Computation of Injection Gains
⊙
+
+ P
+
$\widehat{\mathbf{MS}}$
$\widehat{\mathbf{MS}}_k$

FIGURE 2. The MRA-based methods. Some MRA approaches skip the dashed box.

the spatial details for the fused image. Specifically, the deep residual network mainly contains a preprocessing convolutional layer that increases the feature channels and a postprocessing convolutional layer that reduces the channels to the spectral bands. Furthermore, four residual network blocks [117] with a skip connection are employed to increase the network depth for better feature extraction.

In particular, the high-frequency spatial information of the LRMS and PAN images, which is obtained by using simple high-pass filters, is concatenated and exploited into the deep residual network for training. With this step, we can learn accurate spatial details that will be added to the LRMS to yield the final HRMS product. The output of the network is then compared with the ground truth (GT) image using an ℓ_2 loss function. Via an Adam optimizer with momentum, the weights on all the layers can be suitably updated. This strategy of focusing on high-frequency content is valid, even obtaining good generalization ability. Details of the PanNet (including the architecture, hyperparameter setting, and so forth) are available in Figure 3.

The idea of the PanNet is to design the network architecture in the high-pass domain rather than the image domain that is commonly used for most DL-based techniques. The domain-specific high-pass strategy can foster the network generalization capability since images obtained from different sensors have similar distributions for high-frequency information. Also, since most high-pass details are close to zero, there is a reduction of the mapping space, leading to easier training of the network. In summary, the PanNet demonstrates that the training and generalization abilities of a network can be improved by focusing on a specific domain, i.e., the high-pass domain, instead of the original one.

DEEP RESIDUAL NEURAL NETWORK FOR PANSHARPENING

Wei et al. [71] proposed a deep residual NN, the DRPNN, to address the task of pansharpening, as shown in Figure 4. They believed that a deeper CNN with more filtering layers tends to extract more abstract and representative features, and thus a higher prediction accuracy is expected. However, due to the gradient vanishing problem, weights of shallow layers cannot be optimized via backpropagation, which prevents the deep network from being fully learned. Deep residual learning [117] is an advanced method for solving this problem, in which the transformation $\mathcal{F}(\mathbf{X}) \approx \mathrm{CNN}(\mathbf{X})$ is replaced with $\mathcal{F}(\mathbf{X}) - \mathbf{X} \approx \mathrm{RES}(\mathbf{X})$ by setting a skip connection between separate layers, which facilitates adding more layers to the network to boost its performance.

In the DRPNN, Wei et al. built a deep residual skip before and after the convolutional filtering framework, containing 10 layers with all the kernel sizes set to 7 × 7. MS bands to be fused are interpolated to the PAN scale and then concatenated with the PAN image to form an input cube. After the deep residual feature extraction, a restoration layer with N groups of convolutional filters is employed to obtain the fused images. The outcome is used to calculate the ℓ_2 loss with the GT,

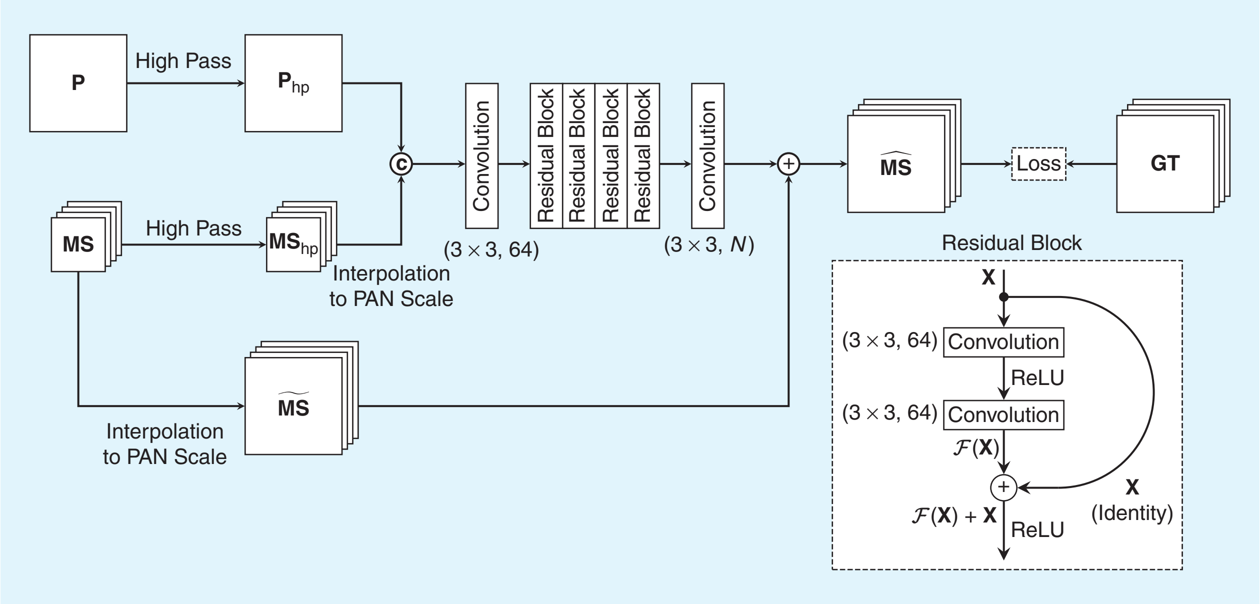


FIGURE 3. The PanNet exploiting the ℓ_2 loss function. Note that $(3 \times 3, 64)$ means that the size of the convolutional kernel is 3 × 3 with 64 channels, and the rectified linear unit (ReLU) is the activation function. The notations © and ⊕ stand for concatenation and summation, respectively. The $\mathbf{P}_{hp}$ and $\mathbf{MS}_{hp}$ are the high-pass-filtered versions of the **P** and **MS** images, respectively. The "Convolution" block represents a convolutional layer. The up-sampling is done using a 23-tap polynomial interpolator [118]. The definitions and notations in the following network architectures are the same as these ones, so we will not introduce them again.

and then the stochastic gradient descent (SGD) algorithm is utilized to train the DRPNN, which costs 300 epochs. In addition, Wei et al. set different learning rates for the first 10 layers and the last layer, which were 0.05 and 0.005, respectively, while the momentum was fixed at 0.95. Note that after every 60 epochs, the learning rate would fall by half.

The deep residual skip ensures that the model learns the difference between input and output, leading to quick and accurate training. The strategy of the skip connection is also used in the PanNet, which was published in the same period as the DRPNN. The DRPNN can achieve competitive outcomes thanks to its use of convolution kernels with a larger size, i.e., 7 × 7, which can cover a greater area. However, due to these larger kernels, the DRPNN has a relatively high number of parameters.

MULTISCALE AND MULTIDEPTH CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE FOR PANSHARPENING

In [75], Yuan et al. proposed a multiscale and multidepth CNN, the MSDCNN, for pansharpening. As demonstrated in Figure 5, the MSDCNN extracts deep and shallow features by using different convolutional filters with receptive fields of multiple scales and finally integrates them to yield better estimation. In pansharpening, coarse structures and texture details are of great importance for ideal restoration. At the same time, the sizes of the ground objects vary from very small neighborhoods to large regions containing thousands of pixels, and a ground scene can cover many objects with various sizes. Recalling that multiscale features respond differently to convolutional filters with different

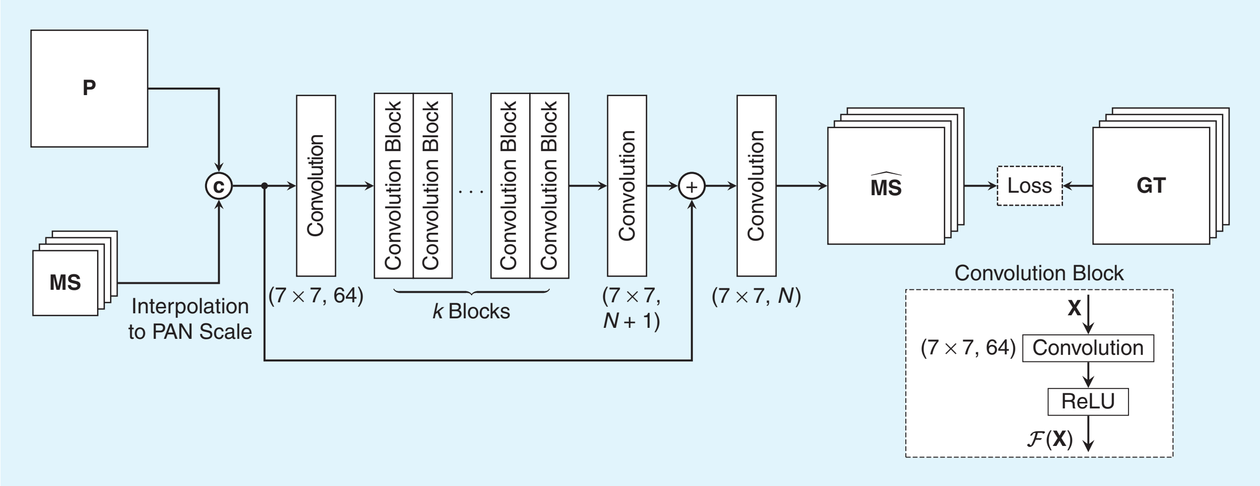


FIGURE 4. The DRPNN using the ℓ_2 loss function.

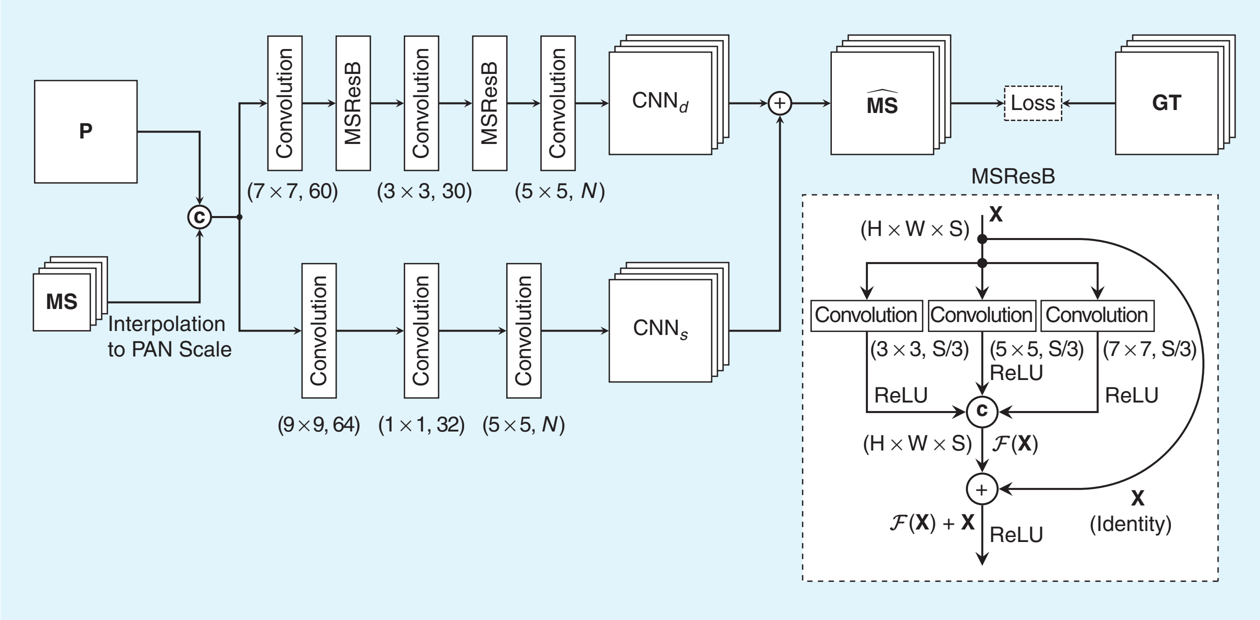


FIGURE 5. The MSDCNN using the ℓ_2 loss function. Here, $\mathbf{CNN}_d$ and $\mathbf{CNN}_s$ stand for deep and shallow features, respectively.

sizes, the authors proposed a multiscale block containing three parallel convolutional layers with kernel sizes of three, five, and seven.

Furthermore, they employed a short skip connection for each multiscale block, which formed the multiscale residual block (MSResB), as in Figure 5. Passing the input image cube through the deep extraction branch, the deep features $\mathbf{CNN}_d$, which have been reduced to the same spectral dimensionality as the ideal MS images, can be extracted. On the other hand, the shallow features $\mathbf{CNN}_s$ are yielded by a shallow network branch with three convolutional layers, where the kernel sizes are nine, one, and five, respectively. Furthermore, the output feature numbers of the convolutional layers in both branches are reduced as the depth increases. The MSDCNN is trained for 300 epochs by using the ℓ_2 loss function with the SGD optimization algorithm, where the momentum μ is equal to 0.9 and the learning rate ϵ is 0.1.

Overall, the MSDCNN benefits from several features obtained by convolving one feature with kernels of different sizes (called *multiscaled operation*). By this strategy, different features with various receptive fields are concatenated to improve the feature extraction. Beyond the multiscaled operation in the so-called deep branch, the other branch conducts three plain convolutions to obtain the “shallow features.” We think the plain convolution layers in the shallow branch might not be necessary since they make the network outputs from the two branches too flexible, resulting in uncertainty in learning deep and shallow features.

BIDIRECTIONAL PANSHARPENING NETWORK

In traditional MRA-based pansharpening methods, multiscale details of the PAN image are used to improve the resolution of the MS image. The accuracy of multiscale details is directly related to the quality of the pansharpened image. Insufficient details lead to blurring effects; excessive details result in artifacts and spectral distortions. To more accurately extract the multiscale details of an HRMS image, Zhang et al. [79] propose a two-stream network for pansharpening, the BDPN. The network adopts a bidirectional pyramid structure to separately process the MS image and the PAN image, following the general idea of MRA. Multilevel details are extracted from the PAN image and injected into the MS image to reconstruct the pansharpened image. The detail extraction branch uses stacked residual blocks to extract details, while the image reconstruction branch uses subpixel convolutional layers to up-sample the MS image. The multiscale structure helps the network to extract multiscale details from the PAN image. It allows part of the computation to be located at reduced-resolution features, thus easing the computation burden.

In the network’s training, a multiscale loss function is used to accelerate the rate of convergence. At the beginning, reconstructed images at all the scales are supervised. As the training continues, the weight of the low-resolution scales gradually declines. A detailed flowchart of the BDPN is provided in Figure 6. Although the idea of a bidirectional structure has been proposed in other multiresolution fusion applications, such as deep image SR [119], the BDPN used it first for pansharpening. However, because of the use of too many multiscaled convolution layers, the BDPN has a large number of parameters, similar to the DRPNN. This disadvantage can be alleviated by exploiting more effective convolutions.

DETAIL INJECTION-BASED CONVOLUTIONAL NEURAL NETWORK

He et al. [91] proposed a detail injection procedure based on DL end-to-end architectures to learn the MS details while enhancing the physical interpretability of the

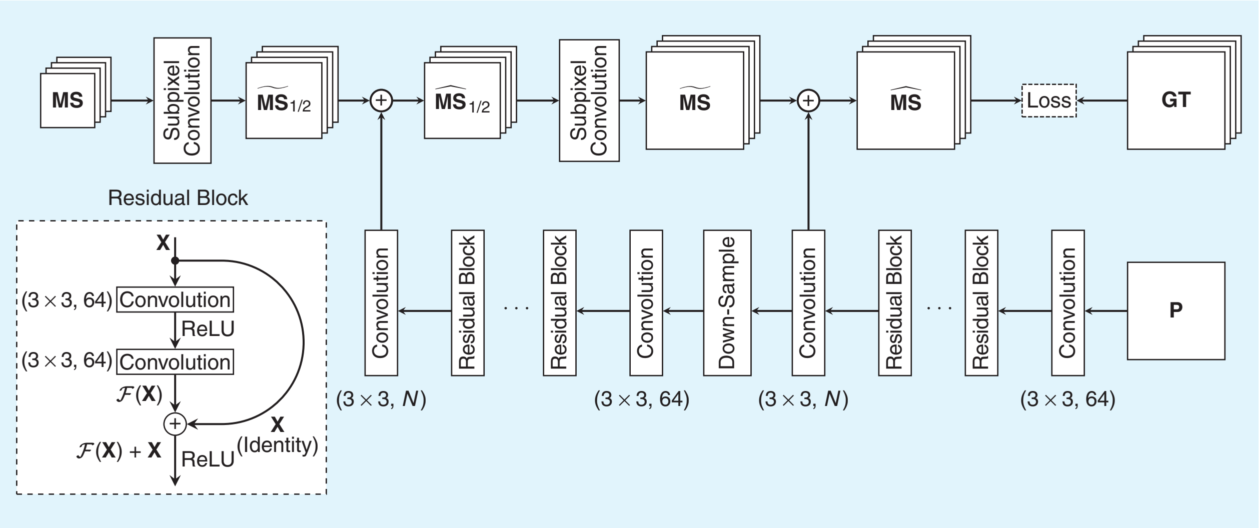


FIGURE 6. The BDPN exploiting the Charbonnier loss function, where “Down-Sample” means the reduction of the spatial resolution by a factor of two and “1/2” stands for up-sampling by a factor of two. The “Subpixel Convolution” block represents a subpixel convolutional layer used to up-sample the MS image.

pansharpening process. Two DiCNN models are implemented following the three-layer architecture for SR proposed by Dong et al. [90]. Figure 7 gives a graphical overview of the network used in this work based on the first proposed model in [91].

The adopted DiCNN receives as input the concatenation along the spectral dimension of the PAN image, **P**, and the MS image up-sampled to the PAN scale, $\widetilde{\mathbf{MS}}$. As a result, the volume $\mathbf{G} \in \mathbb{R}^{H \times W \times N+1} = (\widetilde{\mathbf{MS}}, \mathbf{P})$ is obtained as input, where $H \times W$ indicates the spatial dimensions and N denotes the number of spectral bands of the MS plus the PAN image. This input volume **G** is processed by a stack of three 3 × 3 convolution layers, where the first and second layers are followed by the nonlinear activation function rectified linear unit (ReLU) to explore the nonlinearities of the data. In this regard, the stack of convolution layers exploits the relations between the up-sampled MS and PAN images to obtain MS details that can enhance the original MS data, involving the mapping function $\hat{\mathbf{D}}(\mathbf{G};\theta)$ that obtains the details of the MS fused image from the inputs **G**, with θ representing the set of the learnable parameters of the convolutions.

Moreover, the DiCNN employs residual learning to enhance the feature extraction process by propagating only the up-sampled MS image through a shortcut connection. This not only maintains the same number of spectral bands between the shortcut data and the obtained details (avoiding the implementation of an auxiliary convolution within the shortcut) but also provides an explicit physical interpretation. Indeed, in contrast to other deep models that work as black boxes, the DiCNN introduces a domain-specific structure with a meaningful interpretation. As a result, the output $\hat{\mathbf{D}}(\mathbf{G};\theta)$ can be directly exploited to enhance the up-sampled MS image to produce the desired HRMS image.

In this sense, the main goal of the DiCNN is to minimize the loss function $l(\theta)$ defined by (5), with the aim of appropriately adjusting the network parameters that best fit the data:

$$\begin{aligned} l(\theta) &= \left\| \hat{\mathbf{D}}(\mathbf{G};\theta) + \widetilde{\mathbf{MS}} - \mathbf{Y} \right\|_F^2 \\ &= \frac{1}{N_p} \sum_{i=1}^{N_p} \left\| \hat{\mathbf{D}}^{(i)}(\mathbf{G}^{(i)};\theta) + \widetilde{\mathbf{MS}}^{(i)} - \mathbf{Y}^{(i)} \right\|_{F'}^2 \end{aligned} \tag{5}$$

where **Y** denotes the GT image, N_p is the number of input patches, i is the index of the current patch, and $\|\cdot\|_F$ is the Frobenius norm. This guarantees that the DiCNN approach can directly learn the details of the MS data. Overall, the strategy of the skip connection, as for the PanNet, DRPNN, and MSDCNN, is employed again in the DiCNN to have fast convergence with accurate computation. Thanks to the use of only three convolution layers, the DiCNN involves significantly fewer network parameters while providing competitive pansharpened performance.

PANSHARPENING NEURAL NETWORK

The pansharpening CNN model by Masi et al. [69], the PNN, was among the first pansharpening solutions based on CNNs. Inspired by the SR network for natural images proposed in [90], the PNN is a very simple three-layer fully convolutional model. Table 2 reports the main hyperparameters related to PNN implementation for the proposed toolbox, where, differing from the original version, the hyperparameters have been set equal for all sensors, with the obvious exception of the number of input and output channels of the whole network that are related to the number of spectral bands of the MS image. The three convolutional layers are interleaved by ReLU activations. Prior to feeding the network, the input MS component is upscaled to the PAN size via 23-tap polynomial interpolation and concatenated with the PAN to form a single-input data cube.

Although the PNN exploits the CNN architecture for single-image SR in [90], just extending it to the pansharpening task, this approach holds quite an important role in the DL-based pansharpening community. In fact, it was the first attempt to address the pansharpening issue by using a fully CNN, resulting in an important benchmark for subsequently developed DL-based pansharpening techniques. Since the main structure of the PNN involves only three simple convolution layers without skip connections, its parameters are not significant in terms of getting relatively slow convergence.

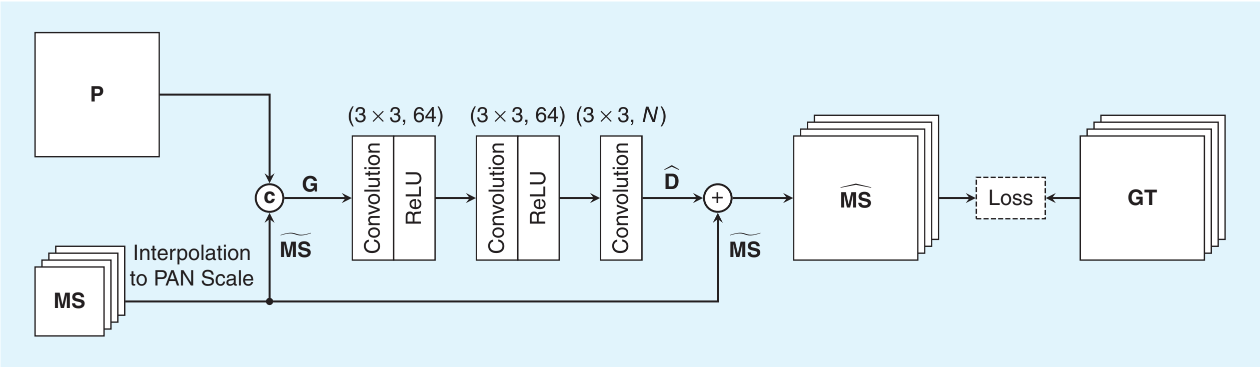


FIGURE 7. The DiCNN exploiting the Frobenius loss function.

TABLE 2. THE OPTIMAL PARAMETERS FOR THE EIGHT COMPARED ML-BASED METHODS.

PARAMETER	PNN	A-PNN-FT	DRPNN	MSDCNN	PanNet	DiCNN	BDPN	FusionNet
Epoch number	12,000	10,000	500	500	450	1,000	1,000	400
Minibatch size	64	64	64	64	32	64	8	32
Optimization algorithm	SGD	SGD	Adam	Adam	Adam	Adam	Adam	Adam
Initial learning rate	0.0289*bands	0.0289*bands	2×10^{-4}	1×10^{-4}	0.001	2×10^{-4}	0.0001	0.0003
Learning rate tuning strategy	Fixed initial learning rate (FIL)	FIL	×0.5 per 60 epochs	×0.5 per 60 epochs	FIL	×0.5 per 200 epoch	×0.8 per 100 epochs	FIL
Filter size for each layer	$9\times9, 5\times5$	$9\times9, 5\times5$	3×3	3×3	3×3	3×3	3×3	3×3
Filter number for each layer	64, 32	64, 32	32	32	64	32	64	32
Type of loss function	ℓ_2	ℓ_1	ℓ_2	ℓ_2	ℓ_2	Frobenius	Charbonnier	ℓ_2
Number of layers	3	3	11	12	10	3	43	10

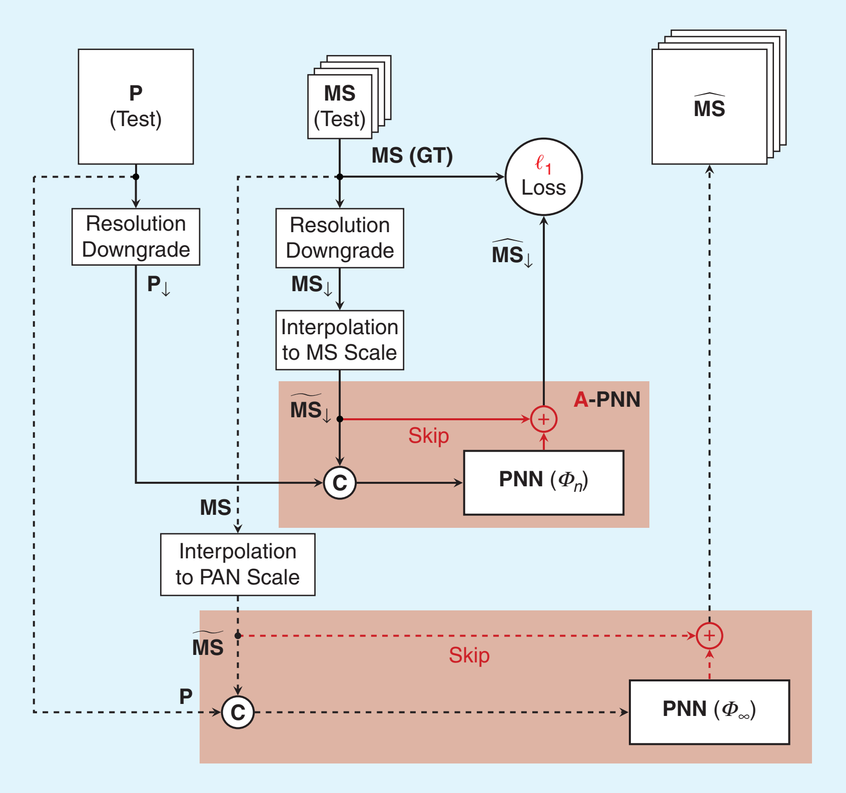


FIGURE 8. The A-PNN-FT, with reduced-resolution adaptation (solid lines) and full-resolution test (dashed) phases. The core A-PNN model is highlighted by opaque blocks and differs from the PNN by introducing a skip connection (red lines) and the ℓ_1 loss, which replaces an ℓ_2. The symbol ↓ indicates a resolution-downgraded version of the image.

ADVANCED PANSHARPENING NEURAL NETWORK USING FINE-TUNING

Two years later, Scarpa et al. [74] proposed an advanced version of the PNN, the A-PNN-FT, that presented three main innovations: residual learning, ℓ_1-loss, and fine-tuning for target adaptation. Residual learning [117] is an important innovation in DL, introduced with the primary purpose of speeding up the training process for very deep networks, as it helps prevent vanishing gradient problems. However, it has proved to be a natural choice for resolution enhancement [120]. In fact, the desired superresolved image can be viewed as being composed of its low- and high-frequency components, the former being essentially the input low-resolution image and the latter being the missing (or residual) part to be actually restored. Residual schemes naturally address SR and pansharpening problems in light of this partition, avoiding the unnecessary reconstruction of the whole desired output and reducing the risk of altering the low-frequency content of the image (i.e., spectral distortion). As a matter of fact, the majority of the recent DL pansharpening models embed residual modules [71], [73], [74], [76], [78], [79]. Specifically, for the A-PNN-FT, a single input–output skip connection added to the PNN model converts the model in a global residual module, as highlighted by the semitransparent blocks of Figure 8, which summarizes the overall A-PNN-FT algorithm. Solid-line connections refer to the fine-tuning phase.

Differing from the usual training, where data samples do not come from test images, in fine-tuning, the same test image is used for parameter updates, as shown in Figure 8. This makes perfect sense thanks to the self-supervised learning allowed by the resolution downgrade process that generates labeled samples from the input itself. Further details about the training (pretraining for the A-PNN-FT) of all the toolbox models are provided in a dedicated section of this article. When fine-tuning begins, the model parameters Φ_0 correspond to those computed in pretraining, and they are associated to what is referred to as the *A-PNN*. After a predetermined number of tuning iterations (50, by default) on the target (rescaled) test image, the parameters are frozen (say, Φ_∞), and eventually (follow the dashed lines) the full-resolution test image can be pansharpened using the "refined" A-PNN, that is, the A-PNN-FT.

PANSHARPENING BY COMBINING MACHINE LEARNING AND TRADITIONAL FUSION SCHEMES

The traditional approaches, such as CS and MRA, have achieved competitive outcomes in pansharpening. Nevertheless, they are under the assumption of linear injection models, which can be unsuitable in terms of the real spectral responses for sensors typically used in pansharpening. This motivates utilizing nonlinear approaches, such as ML, to avoid the limitation of the linear injection models. In [92], Deng et al. exploit the combination of ML techniques and traditional fusion schemes, i.e., CS and MRA, to address the task of pansharpening. The overall network architecture, the FusionNet, estimates the nonlinear injection models that rule the combination of the up-sampled LRMS image and the extracted details exploiting the two philosophies. In particular, the extracted details can be calculated by directly inputting the difference between the duplicated PAN image and the up-sampled LRMS image into a deep residual network.

This strategy of directly differing the duplicated PAN and the up-sampled LRMS images is simple. However, it can preserve the latent spatial and spectral information from PAN and LRMS images, respectively. In addition, the extracted details are taken into account in a preprocessing convolutional layer to increase the feature channels; the extracted details then pass four residual network blocks to increase the network depth for better feature extraction. The generated features are convoluted by a postprocessing layer to get HRMS details consisting of the same LRMS spectral band number. Moreover, the learned HRMS details are directly added to the up-sampled LRMS to yield the HRMS outcome. The FusionNet exploits an Adam optimizer with momentum and a fixed learning rate to train the network. The conventional ℓ_2 function is selected as a loss function to measure the distance between the HRMS outcome and the GT image. Figure 9 contains further details about the FusionNet approach.

Thanks to the combination of ML techniques and traditional fusion schemes to design the network architecture, the FusionNet can have better and faster regression between inputs and labels, generating competitive results when training and testing data sets have similar structures. However, since the FusionNet is also built by plain convolution layers, like the PNN and DiCNN (even with skip connections), its network generalization is weaker than that of the PanNet and A-PNN-FT, which are based on specific operations, such as learning in the high-pass domain and fine-tuning.

EXPERIMENTAL RESULTS

This section is devoted to the description of experimental results. The quality assessment protocols are detailed together with the data sets and the benchmark used for comparison purposes. Afterward, the generation of the training data and the parameter tuning is provided. Finally, the results at reduced and full resolutions are summarized, including a discussion about the computational burden, convergence, and other details of ML-based approaches.

QUALITY ASSESSMENT OF FUSION PRODUCTS

The quality assessment of pansharpening methods and data products is a highly debated problem. Wald's protocol [121] provides an answer to this issue by introducing two main properties (i.e., consistency and synthesis) that a fused product should satisfy. To verify the synthesis

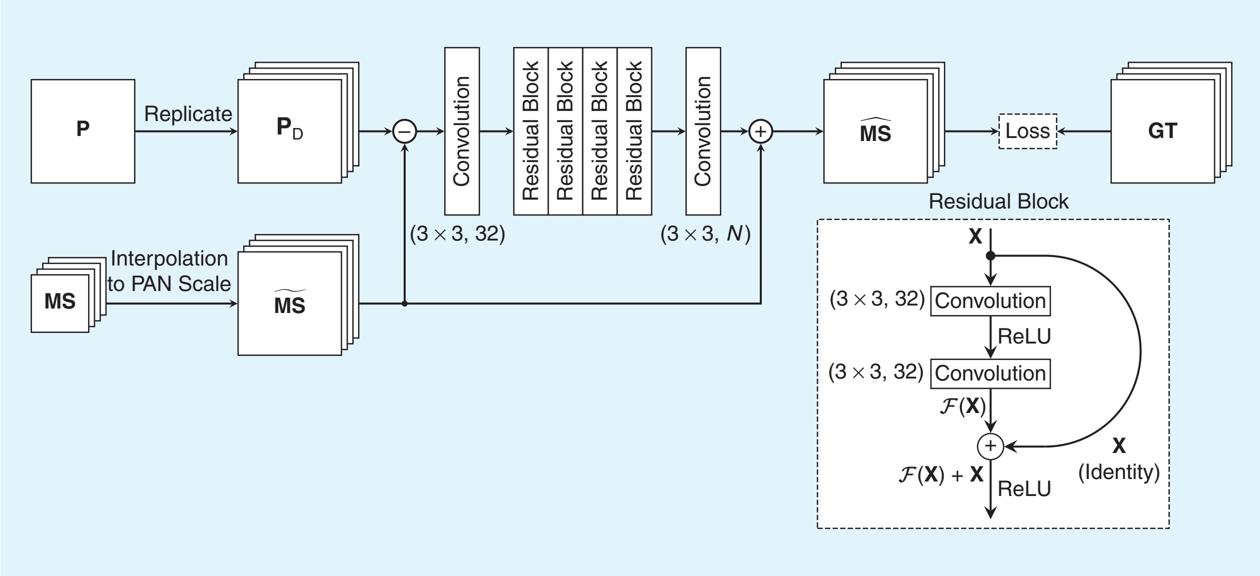


FIGURE 9. The FusionNet exploiting the ℓ_2 loss function. Here, $\mathbf{P}_D$ is the replicated version of $\mathbf{P}$ along the spectral dimension.

property, a reduced-resolution assessment is considered. Thus, the original MS and PAN images are degrading by spatially filtering them to a reduced resolution. Then, the pansharpening algorithm is applied to these data, and the outcome is compared with the original MS data playing the role of the reference image. The more the fused and the reference images are similar, the higher the performance of the pansharpening approach. Clearly, the choice of the filters to spatially degrade the MS and PAN products could bias the assessment. Generally, spatial filters matching the MS sensor's MTFs are exploited to degrade the MS image. Instead, ideal filters are adopted to reduce the resolution of the PAN image [4]. The similarity between the fused and reference images is measured by exploiting the following multidimensional indexes: the spectral angle mapper (SAM) [122], relative dimensionless global error in synthesis (ERGAS) [123], and multiband extension of the universal image quality index, $Q2^n$ [124]. The ideal results are zero for the SAM and ERGAS and one for $Q2^n$.

Unfortunately, a sole reduced-resolution assessment is not enough to state the superiority of a pansharpening algorithm. Indeed, an implicit hypothesis of "invariance among scales" is maintained when working at reduced resolution. Thus, even though this assessment is very accurate, it is based on the validity of the assumption. To this end, a full-resolution assessment is also considered. In this case, no hypothesis is present, but the lack of a reference image reduces the accuracy of the performance assessment. In this article, the hybrid quality with no reference (HQNR) index is used. This borrows the spatial distortion index, D_S, from the QNR index [125], and the spectral distortion index, D_λ, from Khan's protocol [126]. The two distortions are combined as follows:

$$\mathrm{HQNR} = (1 - D_\lambda)^\alpha \ (1 - D_S)^\beta, \tag{6}$$

where $\alpha = \beta = 1$. Ideal values for the D_S and the D_λ indexes are zero; thus, the optimal value for the HQNR is one.

DATA SETS

Several different test cases acquired by five widely used sensors for pansharpening are considered. For all the sensors, assessments at reduced resolution and at full resolution, following the indications in the "Quality Assessment of Fusion Products" section, are provided. The characteristics of the employed data sets are detailed as follows.

WORLDVIEW-2 DATA SETS

These data were acquired by the WV2 sensor, which works in the visible and near-infrared spectrum range. The MS sensor is characterized by eight spectral bands (coastal, blue, green, yellow, red, red edge, near-infrared region 1, and near-infrared region 2), and a PAN channel is available. The spatial sampling interval (SSI) is 1.84 m for MS and 0.46 m for PAN, respectively. The resolution ratio R is equal to four. The radiometric resolution is 11 b.

The following three data sets are exploited:

1) WV2 Washington, representing a mixed area in Washington, United States, characterized by an elevated presence of high buildings, vegetated areas, and a river (the size of an MS spectral band is 6,248 × 5,964); see Figure 10.
2) WV2 Stockholm, depicting a mixed zone with several water bodies in the urban area of Stockholm, Sweden (the size of an MS spectral band is 1,684 × 2,176); see Figure 10.
3) WV2 Rio, showing a mixed area of the city of Rio de Janeiro, Brazil, characterized by vegetated and urban features and a small portion of the ocean at the top right of the image (the size of an MS spectral band is 512 × 512); see Figure 11.

The first two data sets are used for training and testing the networks at reduced resolution, following Wald's protocol, as in the "Quality Assessment of Fusion Products" section. The third is exploited to test the ability of the networks in real conditions, namely, with a different data set acquired by the same sensor over a different area of the world and at a different time, thus having different features, such as atmospheric conditions, haze, landscapes, solar elevation angles, and so forth. In this case, reduced- and full-resolution assessments are performed.

WORLDVIEW-3 DATA SETS

These data were acquired by the WV3 sensor, which works in the visible and near-infrared spectrum range. The MS sensor is characterized by eight spectral bands (the same as the WV2 MS sensor), and a PAN channel is available. The SSI is 1.2 m for MS and 0.3 m for PAN, respectively, and R is equal to four. The radiometric resolution is 11 b.

Three data sets are exploited, as follows:

1) WV3 Tripoli, representing an urban area of Tripoli, Libya (the size of an MS spectral band is 1,800 × 1,956); see Figure 10.
2) WV3 Rio, a mixed data set showing both vegetated and man-made structures in the surroundings of Rio de Janeiro (the size of an MS spectral band is 2,380 × 3,376); see Figure 10.
3) WV3 New York, depicting the urban area of New York City, with more tall buildings with respect to European urban scenarios (the size of an MS spectral band is 512 × 512); see Figure 11.

Again, the first two data sets are used for training and testing the networks at reduced resolution, following Wald's protocol. The real test cases (at reduced resolution and at full resolution) are performed by exploiting the third set.

WORLDVIEW-4 DATA SETS

These data were acquired by the WV4 sensor, which works in the visible and near-infrared spectrum range. The MS sensor is characterized by four spectral bands (blue, green, red, and near-infrared region), and a PAN channel is available. The SSI is 1.24 m for MS and 0.31 m for PAN, respectively, while R is equal to four. The radiometric resolution is 11 b.

The following two data sets are exploited:

- WV4 Acapulco, representing a mixed area with sea, land, and urban areas in the surroundings of the city of Acapulco, Mexico (the size of an MS spectral band is 4,096 × 4,096); see Figure 10.
- WV4 Alice, a mixed data set mainly showing urban and bare soil features related to the city of Alice Springs, Australia (the size of an MS spectral band is 512 × 512); see Figure 11.

Once more, the first data set is used for training and testing the networks at reduced resolution, following Wald's protocol. The real test cases (at reduced resolution and at full resolution) are performed by exploiting the second.

QUICKBIRD DATA SETS

These data were acquired by the QB sensor, which works in the visible and near-infrared spectrum range. The MS sensor is characterized by four spectral bands (blue, green, red, and near-infrared region). A PAN channel is available. The SSI is 2.44 m for MS and 0.61 m for PAN, respectively; R is equal to four. The radiometric resolution is 11 b.

The following two data sets are exploited:

1) QB Indianapolis, representing a mixed area with a high presence of man-made structures as well as water bodies and green areas captured over the city of Indianapolis, Indiana (the size of an MS spectral band is 3,624 × 4,064); see Figure 10.

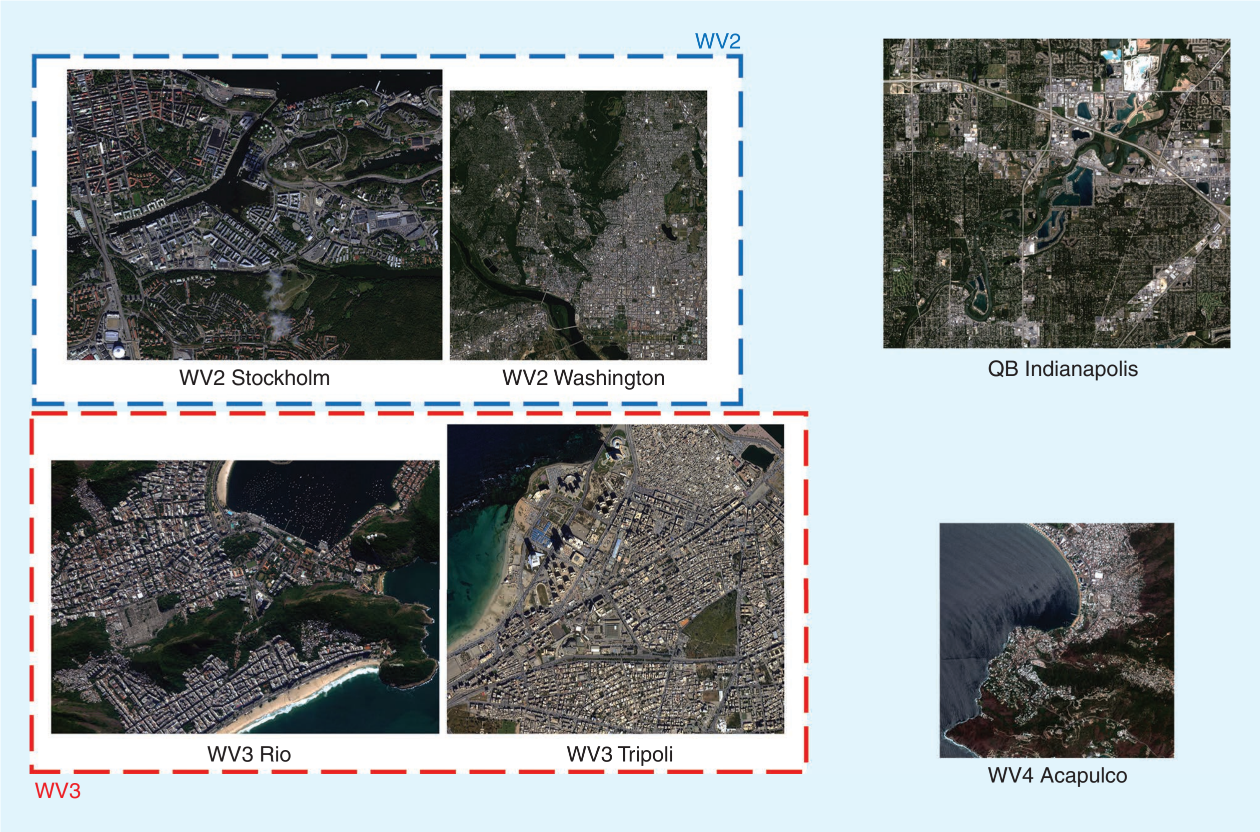


FIGURE 10. The data sets used for training the ML approaches (selected bands: red, green, and blue). Note that the images related to the data sets are intensity stretched to aid visual inspection.

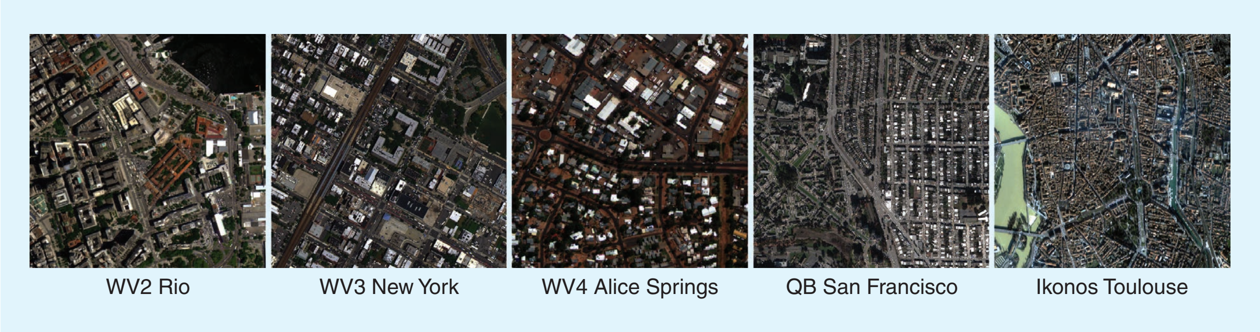


FIGURE 11. The data sets used for testing the ML approaches (selected bands: red, green, and blue). Note that the images related to the data sets are intensity stretched to aid visual inspection.

2) QB San Francisco, showing the urban area of San Francisco (the size of an MS spectral band is 512 × 512); see Figure 11.

The first is used for training and testing the networks at reduced resolution, following Wald's protocol. The real test cases (at reduced resolution and at full resolution) are performed by exploiting the second.

IKONOS DATA SET

This data set represents an area of the city of Toulouse, France. It was acquired by the Ikonos sensor, which works in the visible and near-infrared spectrum range. The MS sensor is characterized by four spectral bands, as with the QB sensor, and a PAN channel is available. The resolution cell is 4 × 4 m for the MS bands and 1 × 1 m for the PAN channel; therefore, R is equal to four. The radiometric resolution is 11 b. The size of an MS spectral band is 512 × 512 pixels (see Figure 11). This data set is used to assess the generalization ability of the networks with respect to the changing of both the acquisition sensor and captured scenario. In particular, we exploit networks trained on the QB data set and evaluated on this IKONOS data set.

BENCHMARK

Several state-of-the-art algorithms are employed for comparison purposes, presented as follows:

- *EXP*: MS image interpolation using a polynomial kernel with 23 coefficients
- *CS methods*:
  - BT-H [30]
  - BDSD with physical constraints (PCs) [18]
  - context-based GS adaptive (C-GSA) with local parameter estimation exploiting clustering [16]
- *MRA methods*:
  - GLP with MTF-matched filters with an FS regression-based injection model (MTF-GLP-FS) [116]
  - GLP with MTF-matched filters and an HPM injection model with a preliminary regression-based spectral matching phase (MTF-GLP-HPM-R) [115]
- *VO methods*:
  - pansharpening based on sparse representation of injected details (SR-D) [66]
  - pansharpening based on TV [50];
- *ML methods*:
  - *PanNet*: PanNet [73]
  - *DRPNN*: DRPNN [71]
  - *MSDCNN*: MSDCNN [75]
  - *BDPN*: BDPN [79]
  - *DiCNN*: DiCNN [91]
  - *PNN*: PNN [69]
  - *A-PNN-FT*: A-PNN-FT [74]
  - *FusionNet*: FusionNet [92].

A more detailed description of the methods can be found in the "Component Substitution, Multiresolution Analysis, and Variational Optimization: A Brief Overview" and "A Benchmark Relying on Recent Advances in Machine Learning for Pansharpening" sections and the related references.

GENERATION OF TRAINING DATA

The building of the training set is a crucial step for ML-based pansharpening approaches. Although, in the literature, there are plenty of state-of-the-art ML-based methods, the process of generating training sets is often different, leading to unfair comparisons. This section is devoted to the illustration of the whole procedure of generating training samples for ML-based pansharpening. Moreover, the MATLAB code for simulating training sets will be distributed to the community.

The overall procedure of generating training samples is depicted in Figure 12, which involves the follows three main steps:

1) *Data download*: Because of license limitations, we are not permitted to share the original data. Readers can directly download them from commercial websites (for WV data, readers can refer to https://resources.maxar.com/).
2) *Data simulation*: After downloading the source images, we can read the original PAN and MS images. Afterward, according to Wald's protocol, we filter the original MS image matching the corresponding sensor's MTF (the MATLAB code for the filtering using the MTF can be found at the following link: https://github.com/liangjiandeng/DLPan-Toolbox/tree/main/02-Test-toolbox-for-traditional-and-DL(Matlab)/Tools) and the original PAN image by using an almost ideal filter, then downgrading the filtered images by the nearest-neighbor interpolation with a scale factor of four. The down-sampled MS image will be up-sampled to the PAN scale by a 23-tap polynomial interpolator. Hence, we exploit the following in the training phase: 1) the down-sampled PAN image, 2) the down-sampled MS image, 3) the original MS image as the GT, and 4) the up-sampled version of the down-sampled MS image (UMS). Refer to Figure 12 and Table 3 for more details about the simulation process and the data used in this work, respectively.
3) *Data patching*: The simulated images in step 2 are too big (considering the limited storage capabilities of the GPUs) to feed the pansharpening networks. Thus, we need to crop these simulated images into small patches. We segment the GT, UMS, PAN, and MS images into several small patches with sizes of 64 × 64 × 8 (with an overlap of 16 spatial pixels), 64 × 64 × 8 (with an overlap of 16 spatial pixels), 64 × 64 × 1 (with an overlap of 16 spatial pixels), and 16 × 16 × 8 (with an overlap of four spatial pixels due to the scale factor of four), respectively (the MATLAB code for patching the training data sets can be found at the following link: https://github.com/liangjiandeng/DLPan-Toolbox). Finally, we have 9,000 training samples (i.e., 9,000 patch images) and 1,000 validation samples for the WV3, WV4, and QB data sets,

and 14,496 training samples and 1,611 validation samples for the WV2 data set, which can avoid overfitting during the training phase. Refer to Figure 12 for more details.

PARAMETER TUNING

This section shows the parameter settings for all the compared ML-based pansharpening methods, including information about the epoch number, learning rate, optimization algorithm, loss function, and so forth. These details can be found in Table 2. Note that some ML-based methods were originally implemented on other platforms (e.g., TensorFlow and MatConv). When we migrated the codes to PyTorch, the original parameters were tuned again, accounting for the different behavior of built-in functions (e.g., a different weight initialization) in the adopted software platform.

ASSESSMENT ON WORLDVIEW-2 DATA

In this section, we analyze the outcomes obtained on WV2 test cases. Multiple reduced-resolution testing data sets are first evaluated. Then, another data set is used to assess the performance at reduced resolution and at full resolution.

PERFORMANCE ON 12 REDUCED-RESOLUTION TESTING DATA SETS

We first evaluate the quantitative performance of all the compared pansharpening methods on 12 WV2 reduced-resolution testing data sets acquired over Stockholm, i.e., the testing data in Table 3 (A) (see the WV2 Stockholm data set in Figure 10). Note that multiple testing samples are captured over a similar area at the same time as those of the training data set but exploit different cuts. By looking at the quantitative performance displayed in Table 4, it is easy to see that the ML-based approaches obtain better average indicators than the traditional techniques and have smaller standard derivations, indicating better robustness. Specifically, the FusionNet outperforms the other approaches on these testing data. The PanNet, DRPNN, and DiCNN also have competitive performance. Overall, because the training data set has properties similar to those of the testing samples, the outcomes of ML-based methods show clear superiority with respect to the traditional techniques. This corroborates the ability of the networks during the training phase to properly fit their weights, thus easily solving problems similar to the ones proposed in this testing phase.

PERFORMANCE ON THE REDUCED-RESOLUTION WORLDVIEW-2 RIO DATA SET

This section evaluates the performance of all the compared methods on a single reduced-resolution WV2 test case. Differing from the reduced-resolution WV2 Stockholm testing samples in the preceding section, in this case, the single WV2 testing data set is acquired at different times over the city of Rio. Readers can examine the WV2 Rio testing image in Figure 11. Specifically, Figure 13 reveals that most of the traditional methods have more visual appeal than the ML-based approaches. Generally speaking, only small differences among the compared techniques can be identified. One exception is represented by the outcome provided by the PNN, which has high spectral distortion. The A-PNN-FT (which is an extension of the PNN) has competitive visual performance with high spectral preservation, thus demonstrating the effectiveness of using the fine-tuning strategy for PNN-based approaches.

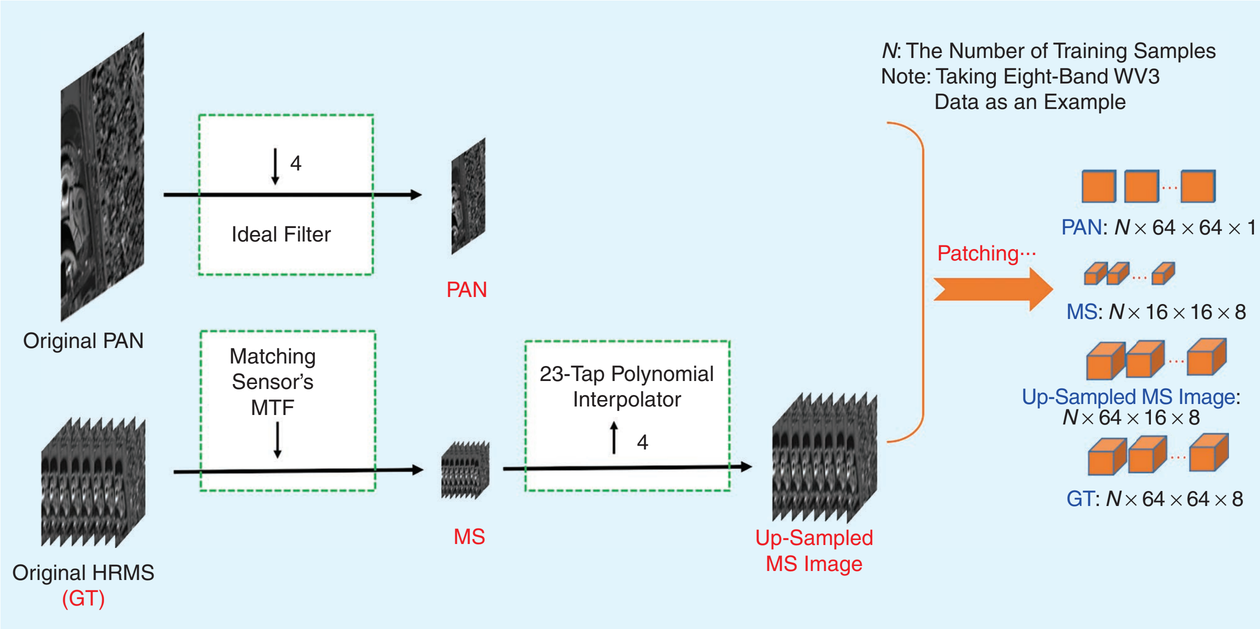
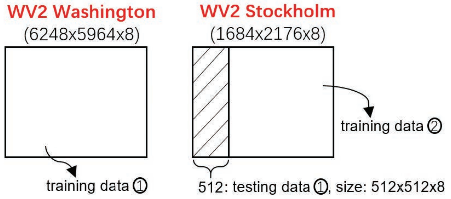
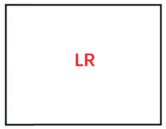
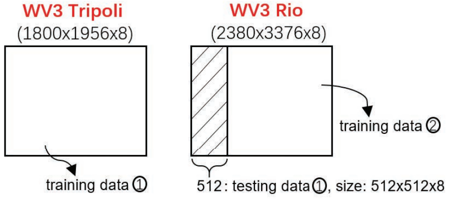
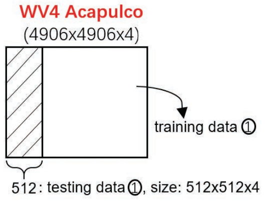
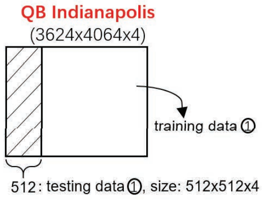


FIGURE 12. The generation process of training samples by Wald's protocol. Note that the names highlighted in red refer to the generated training data used to feed the networks, i.e., the GT, LRMS image, PAN, and up-sampled MS image.

TABLE 3. THE DETAILS OF THE TRAINING AND TESTING DATA SETS USED IN THIS WORK (SEE FIGURES 10 AND 11).

	Data Source	Training Data	Testing Data	Testing Data (For Generalization Ability)
(A) WorldView-2 (8 bands)	 WV2 Washington (6248x5964x8) training data ① WV2 Stockholm (1684x2176x8) training data ② 512: testing data ①, size: 512x512x8	training data: ① + ② total number: 10,000	(1) reduced res. (12 samples): testing data ① (2) reduced res.(1 sample): Rio_wv2_rr.mat (another WV2 sample over the city of Rio) (3) full res.(1 sample): Rio_wv2_fr.mat (another WV2 sample over the city of Rio)	Trained on QB Indianapolis Data (see left) (1) reduced res.(1 sample): Toulouse_ikonos_rr.mat (an IKONOS sample over the city of Toulouse) (2) full res.(1 sample): Toulouse_ikonos_fr.mat (an IKONOS sample over the city of Toulouse) (1) Toulouse_ikonos_rr.mat (128x128x4) (2) Toulouse_ikonos_fr.mat (512x512x4) 
(B) WorldView-3 (8 bands)	 WV3 Tripoli (1800x1956x8) training data ① WV3 Rio (2380x3376x8) training data ② 512: testing data ①, size: 512x512x8	training data: ① + ② total number: 10,000	(1) reduced res. (4 samples): testing data ① (2) reduced res.(1 sample): NY_wv3_rr.mat (another WV3 sample over the city of New York) (3) full res.(1 sample): NY_wv3_fr.mat (another WV3 sample over the city of New York)	
(C) WorldView-4 (4 bands)	 WV4 Acapulco (4906x4906x4) training data ① 512: testing data ①, size: 512x512x4	training data: ① total number: 10,000	(1) reduced res. (8 samples): testing data ① (2) reduced res.(1 sample): Alice_wv4_rr.mat (another WV4 sample over the city of Alice) (3) full res.(1 sample): Alice_wv4_fr.mat (another WV4 sample over the city of Alice)	
(D) QuickBird (4 bands)	 QB Indianapolis (3624x4064x4) training data ① 512: testing data ①, size: 512x512x4	training data: ① total number: 10,000	(1) reduced res. (7 samples): testing data ① (2) reduced res.(1 sample): SF_qb_rr.mat (another QB sample over the city of San Francisco) (3) full res.(1 sample): SF_qb_fr.mat (another QB sample over the city of San Francisco)	

Res.: resolution.

Quantitative results are reported in Table 5. From the table, the differences among all the compared methods are easily seen. Most of the traditional state-of-the-art methods achieve very high indicators, thus demonstrating their spatiospectral preservation ability. The BT-H method achieves the highest Q8 indicator, and the MTF-GLP-HPM-R obtains the lowest ERGAS among all the compared methods. In contrast, we can observe that the results of the ML-based methods differ from one another. Some ML-based approaches, e.g., the PanNet, A-PNN-FT, DRPNN, and MS-DCNN, have promising results, whereas other methods, such as the PNN, BDPN, and DiCNN, have a significant performance reduction. A possible reason is that the testing data set used here is quite different with respect to the training data, e.g., a different acquiring area and time.

Generally speaking, the more parameters there are to be trained, the greater the amount of data required to estimate them. Moreover, to improve the generalization ability, the training set should consist of samples acquired in several areas and in different conditions to convey to a network complete knowledge of the problem at hand. In the absence of a huge and variegated training set, this kind of analysis will reward only networks designed with a higher generalization ability and networks, including the A-PNN-FT, that exploit mechanisms such as the fine-tuning that allows adaptation to specific problems presented during the testing phase. Thus, the A-PNN-FT has the lowest SAM metric, which indicates better spectral preservation. Overall, among the ML-based methods, the PanNet yields promising outcomes, but its performance is still worse than most of the traditional approaches. More conclusions about the compared ML-based methods can be found in Table 5 for each quality metric.

TABLE 4. THE AVERAGE RESULTS OF THE APPROACHES BELONGING TO THE BENCHMARK FOR THE REDUCED-RESOLUTION WV2 STOCKHOLM TESTING DATA SET, I.E., ON THE 12 WV2 TESTING DATA SETS IN TABLE 3 (A).

	Q8 (±STANDARD DEVIATION)	SAM (±STANDARD DEVIATION)	ERGAS (±STANDARD DEVIATION)
		CS/MRA/VO	
GT	1 ± 0	0 ± 0	0 ± 0
EXP	0.5812 ± 0.0569	7.4936 ± 1.2394	7.0288 ± 0.8265
BT-H	0.8501 ± 0.041	6.5042 ± 1.3519	4.1552 ± 0.5579
BDSD-PC	0.843 ± 0.0477	7.1664 ± 1.2654	4.3242 ± 0.5203
C-GSA	0.8323 ± 0.0442	7.8657 ± 1.3074	4.6591 ± 0.4554
SR-D	0.8321 ± 0.0457	6.6042 ± 1.3383	4.3915 ± 0.6267
MTF-GLP-HPM-R	0.8356 ± 0.0446	7.3204 ± 2.0298	5.0992 ± 2.3204
MTF-GLP-FS	0.8347 ± 0.0391	7.4497 ± 1.6581	4.5257 ± 0.6078
TV	0.794 ± 0.0834	7.2902 ± 0.9685	4.84 ± 0.3226
		ML	
PanNet	0.913 ± 0.0551	4.4143 ± 0.6642	2.7713 ± 0.3156
DRPNN	0.9109 ± 0.0528	4.473 ± 0.6906	2.8552 ± 0.3393
MSDCNN	0.9079 ± 0.054	4.5698 ± 0.7250	2.9078 ± 0.3469
BDPN	0.8924 ± 0.0578	5.1381 ± 0.8587	3.2144 ± 0.3781
DiCNN	0.9111 ± 0.0528	4.4857 ± 0.7061	2.8411 ± 0.3365
PNN	0.9043 ± 0.0573	4.6774 ± 0.7064	2.9374 ± 0.3369
A-PNN-FT	0.8991 ± 0.0519	4.9263 ± 0.8348	3.1363 ± 0.3887
FusionNet	<u>**0.9169 ± 0.0532**</u>	<u>**4.2632 ± 0.6336**</u>	<u>**2.6911 ± 0.3115**</u>

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

TABLE 5. A QUANTITATIVE COMPARISON OF THE OUTCOMES OF THE BENCHMARK ON THE REDUCED-RESOLUTION WV2 RIO DATA SET (SEE FIGURE 11).

	Q8	SAM	ERGAS
	CS/MRA/VO		
GT	1	0	0
EXP	0.7283	4.8597	6.7878
BT-H	**0.9441**	3.5368	3.3027
BDSD-PC	0.9316	4.032	3.7105
C-GSA	0.9407	3.8848	3.3972
SR-D	0.9375	3.7881	3.3127
MTF-GLP-HPM-R	0.9436	3.8778	**3.2173**
MTF-GLP-FS	0.9426	3.8129	3.2578
TV	0.9341	4.1811	3.7521
	ML		
PanNet	0.9329	4.2582	3.8532
DRPNN	0.9301	5.092	4.191
MSDCNN	0.92	5.4779	3.8565
BDPN	0.8888	5.9709	5.5306
DiCNN	0.8925	5.6765	5.4202
PNN	0.8866	9.4634	6.5718
A-PNN-FT	<u>0.9374</u>	<u>**3.53**</u>	<u>3.3032</u>
FusionNet	0.9069	5.122	4.1184

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

PERFORMANCE ON THE FULL-RESOLUTION WORLDVIEW-2 RIO DATA SET

Apart from the evaluation of the reduced-resolution data sets, an assessment at full resolution is also needed. To this end, an original full-resolution WV2 Rio data set is used (see Figure 11). Note that the full-resolution WV2 Rio data are also acquired over a different area and at a different time when compared with the WV2 training data. Since there is no GT image, we exploit proper indexes with no reference. We select the HQNR (consisting of the combination of D_λ and D_S) to have a quantitative evaluation of the performance, as introduced in the "Quality Assessment of Fusion Products" section. Table 6 reports the quantitative results. It is easy to observe that most of the traditional state-of-the-art approaches have high performance (even comparing them with the results of the ML-based methods). In particular, two traditional methods, i.e., the SR-D and TV, rank first and third among the 16 compared techniques.

The PanNet is the best ML-based approach and holds the second position in the overall ranking, showing its good network generalization. The reason could relate to the network training conducted only on high-frequency information. Some ML-based approaches yield lower indexes than the traditional methods because the ML-based methods are practically trained on different training samples (as discussed in the previous section) and, in this case, on reduced-resolution samples with data with a lower spatial resolution (this is the main drawback of training ML-based approaches in a supervised manner). This conclusion also refers to the PNN, which obtains the worst HQNR among all the techniques, not only because of its relatively small size but most likely for the lack of residual modules (skip connections), which makes the network prone to spectral distortion on new data sets. After the introduction of a skip connection (the A-PNN) and implementing a fine-tuning strategy, the network (i.e., the A-PNN-FT) can achieve better results (reaching the fourth position in the ranking), thus corroborating the generalization ability of adaptive fine-tuning combined with the robustness provided by properly set residual skip connections.

ASSESSMENT ON WORLDVIEW-3 DATA

In this section, we repeat the same three tests but use WV3 data. Multiple reduced-resolution testing data sets are evaluated first. Then, another data set is used to assess the performance at reduced resolution and at full resolution.

PERFORMANCE ON FOUR REDUCED-RESOLUTION TESTING DATA SETS

This section first evaluates all the compared pansharpening methods on four reduced-resolution WV3 Rio testing data sets that share a similar geographic area and have the same acquiring time as one of the data sets used for training [see the testing data in Table 3 (B) and the WV3 Rio image in Figure 10]. Table 7 reports the average numerical results of all 16 compared state-of-the-art pansharpening methods on the four reduced-resolution WV3 Rio testing data sets. From the table, it is clear that all the ML-based approaches outperform the traditional methods on the three related reduced-resolution indicators, i.e., Q8, SAM, and ERGAS. Note that the FusionNet has the best Q8, SAM, and ERGAS metrics (and most of the best standard deviations), showing its promising ability on testing data acquired over geographic areas similar to those in the training data. Among the ML-based methods, the PanNet, DRPNN, MSDCNN, and DiCNN can be grouped in a second-best class. Indeed, their performance is slightly worse than that of the FusionNet and A-PNN-FT but better than the BDPN and PNN. Among the traditional methods, the BT-H outperforms the others. TV has the worst Q8 and ERGAS indicators. The main reason for the outstanding performance of the

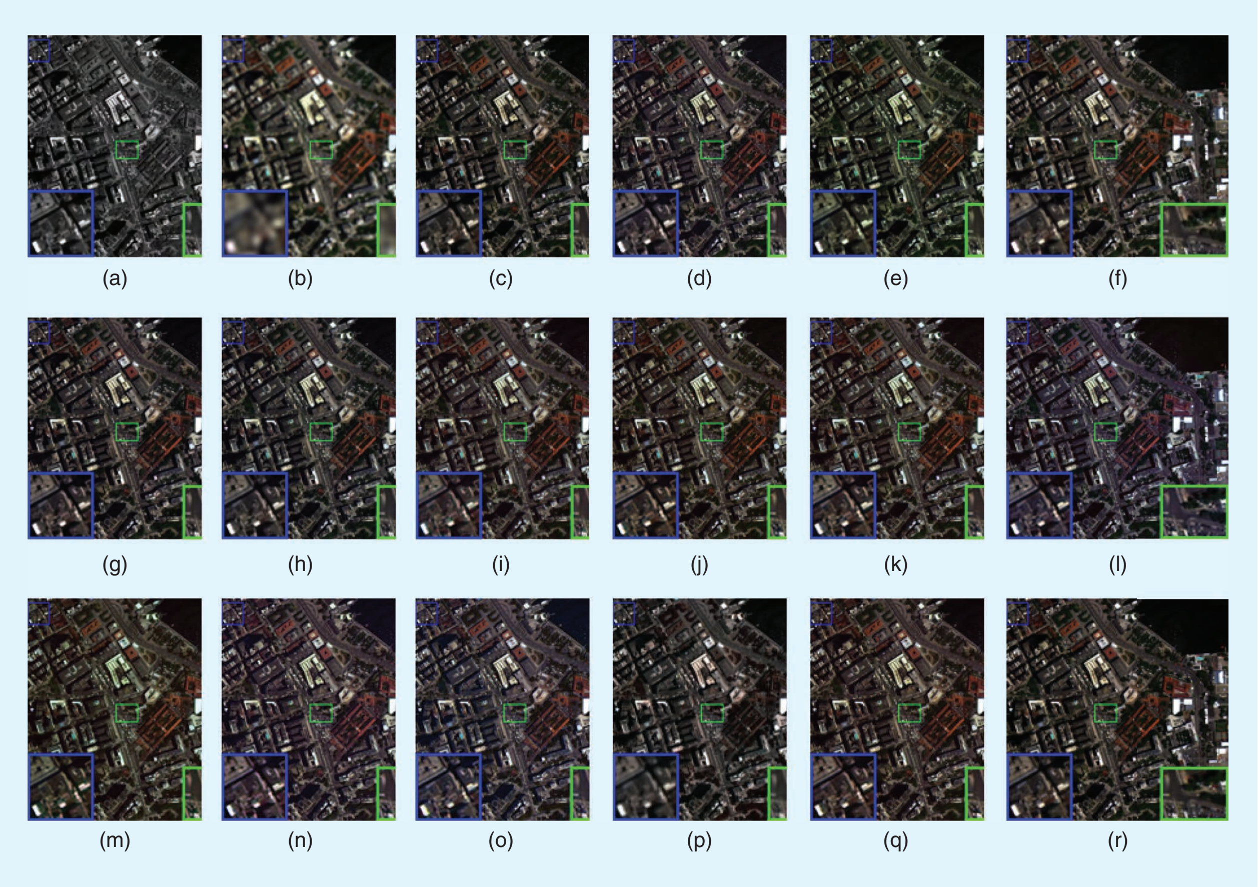


FIGURE 13. Visual comparisons in natural colors of the evaluated approaches on the reduced-resolution WV2 Rio data set (see Figure 11). The (a) PAN, (b) EXP, (c) BT-H, (d) BDSD-PC, (e) C-GSA, (f) SR-D, (g) MTF-GLP-HPM-R, (h) MTF-GLP-FS, (i) TV, (j) PanNet, (k) DRPNN, (l) MSDCNN, (m) BDPN, (n) DiCNN, (o) PNN, (p) A-PNN-FT, (q) FusionNet, and (r) GT.

ML-based methods is the same as in the "Performance on 12 Reduced-Resolution Testing Data Sets" section, i.e., the similarity between training and testing data sets.

PERFORMANCE ON THE REDUCED-RESOLUTION WORLDVIEW-3 NEW YORK DATA SET

We evaluate the quantitative performance of all the compared pansharpening methods on a new reduced-resolution WV3 data set used only for testing purposes. It was acquired over New York City and appears in Figure 11. The testing data set has a different geographical area captured at a different time when compared with the training data set. According to Table 8, the traditional approaches outperform most of the ML-based methods on all the quality metrics. The BDSD-PC, belonging to the class of traditional methods, has the two best indicators, i.e., Q8 and ERGAS, while another traditional technique, i.e., the BT-H, has the lowest SAM. Nevertheless, some ML-based approaches, such as the PanNet, DRPNN, BDPN, and A-PNN-FT, obtain competitive outcomes and represent the second-best class among all the compared methods. In contrast, the other three ML-based methods—the DiCNN, PNN, and FusionNet—have the worst performance. In particular, the PNN achieves the largest SAM (almost 5° more than the second-worst method), indicating higher spectral distortion than the other approaches. The reason why these three ML-based methods are worse than the other ones relates to their simpler network architecture with fewer parameters, which could not fit well the problem's nonlinearities.

PERFORMANCE ON THE FULL-RESOLUTION WORLDVIEW-3 NEW YORK DATA SET

Similar to the "Performance on the Full-Resolution World-View-2 Rio Data Set" section, this section compares the qualitative and quantitative performance of all the methods on the full-resolution WV3 New York data set (see Figure 11). This full-resolution testing data set was acquired New York City, and it has a different geographical area and acquisition time than the training data. Again, the HQNR is used for the performance assessment. Table 9 shows that the traditional techniques outperform most of the ML-based methods. The best results are reported for TV. The SR-D holds the third position in the ranking. Considering all the methods, the traditional MTF-GLP-HPM-R and MTF-GLP-FS techniques show superior performance over the ML-based approaches except for the PanNet and A-PNN-FT, which attained the second and fourth positions, respectively. Among the ML-based techniques, the PanNet and A-PNN-FT achieved the first two positions, thanks to their competitive generalization abilities. Moreover, some ML-based methods, including the MSDCNN, DiCNN, and PNN, have relatively large gaps compared with the PanNet and A-PNN-FT methods. The rest of the ML-based methods, such as the DRPNN, BDPN, and FusionNet, achieve a performance similar to the traditional BT-H, BDSD-PC, and C-GSA methods.

Figure 14 presents a visual comparison of all the compared methods on the full-resolution WV3 New York

TABLE 6. A QUANTITATIVE COMPARISON OF THE OUTCOMES OF THE BENCHMARK ON THE FULL-RESOLUTION WV2 RIO DATA SET (SEE FIGURE 11).

	D_λ	D_S	HQNR
	CS/MRA/VO		
EXP	0.0374	0.0717	0.8936
BT-H	0.0601	0.071	0.8732
BDSD-PC	0.0653	0.0435	0.894
C-GSA	0.0664	0.0653	0.8727
SR-D	**0.0153**	0.0286	**0.9566**
MTF-GLP-HPM-R	0.026	0.0594	0.9161
MTF-GLP-FS	0.0269	0.0652	0.9097
TV	0.0332	0.0269	0.9407
	ML		
PanNet	0.0292	**0.0171**	0.9542
DRPNN	0.0629	0.0311	0.908
MSDCNN	0.0872	0.0498	0.8674
BDPN	0.0909	0.0486	0.8649
DiCNN	0.1043	0.0478	0.8529
PNN	0.1678	0.051	0.7897
A-PNN-FT	0.0379	0.0396	0.924
FusionNet	0.0647	0.0179	0.9185

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

TABLE 7. THE AVERAGE RESULTS OF THE APPROACHES BELONGING TO THE BENCHMARK FOR THE REDUCED-RESOLUTION WV3 RIO TESTING DATA SET, I.E., ON THE FOUR WV3 TESTING DATA SETS IN TABLE 3 (B).

	Q8 (±STANDARD DEVIATION)	SAM (±STANDARD DEVIATION)	ERGAS (±STANDARD DEVIATION)
		CS/MRA/VO	
GT	1 ± 0	0 ± 0	0 ± 0
EXP	0.5974 ± 0.0571	9.2031 ± 0.7655	9.3369 ± 0.4756
BT-H	0.8898 ± 0.0323	7.67 ± 0.7613	4.6132 ± 0.1695
BDSD-PC	0.8454 ± 0.0608	8.9376 ± 0.8568	5.0893 ± 0.2015
C-GSA	0.8695 ± 0.0436	8.8042 ± 0.8652	5.0183 ± 0.1327
SR-D	0.8693 ± 0.032	7.9449 ± 0.4946	5.0739 ± 0.2807
MTF-GLP-HPM-R	0.8625 ± 0.0499	9.4911 ± 1.1386	5.2141 ± 0.2881
MTF-GLP-FS	0.8533 ± 0.0526	9.1442 ± 0.9443	5.2496 ± 0.2077
TV	0.8031 ± 0.0929	8.9863 ± 0.8592	5.3569 ± 0.1948
		ML	
PanNet	0.9232 ± 0.0324	5.1447 ± 0.3995	3.1906 ± 0.2192
DRPNN	0.9203 ± 0.0323	5.1492 ± 0.35	3.2171 ± 0.2067
MSDCNN	0.9154 ± 0.0351	5.5887 ± 0.3471	3.3978 ± 0.1715
BDPN	0.9137 ± 0.0327	6.0121 ± 0.4713	3.6072 ± 0.2425
DiCNN	0.9265 ± 0.0283	5.1285 ± 0.3217	3.1894 ± 0.2106
PNN	0.9068 ± 0.0425	5.9259 ± 0.4544	3.4998 ± 0.1341
A-PNN-FT	**0.9327 ± 0.0255**	4.9125 ± 0.3794	3.088 ± 0.2312
FusionNet	**0.9327 ± 0.0272**	**4.6482 ± 0.3508**	**2.9028 ± 0.1967**

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

testing data set. As evident from the figure, the traditional TV method that achieves the best HQNR does not have the clearest fused image compared with those of the ML-based techniques (see the blue and green close-ups). The BDSD-PC seems to have a blur effect and a clear spectral distortion (mainly due to a color contrast change). A relevant spectral distortion also happens in the case of the pansharpened SR-D product. Furthermore, the BT-H seems to produce precise spatial details even though its quantitative outcomes are not so promising. The visual products of the ML-based techniques are quite competitive without significant spectral distortion. However, some methods, such as the DiCNN and PNN, generate significant blur effects and artifacts (such as outliers), indicating a weak visual appearance. Finally, the BDPN has the clearest spatial details without any artifacts. Note that the rest of the ML-based approaches yield similar visual performance, showing clear spatial details and good spectral preservation.

TABLE 8. A QUANTITATIVE COMPARISON OF THE OUTCOMES OF THE BENCHMARK ON THE REDUCED-RESOLUTION WV3 NEW YORK DATA SET (SEE FIGURE 11).

	Q8	SAM	ERGAS
	CS/MRA/VO		
GT	1	0	0
EXP	0.6513	7.2118	8.1106
BT-H	0.9241	**6.453**	3.9714
BDSD-PC	**0.9327**	6.8388	**3.8905**
C-GSA	0.9213	6.6966	4.0503
SR-D	0.9113	6.6269	4.3472
MTF-GLP-HPM-R	0.9228	7.0038	4.0692
MTF-GLP-FS	0.9228	6.7650	4.0434
TV	0.9277	6.6213	4.063
	ML		
PanNet	0.9238	6.905	4.2365
DRPNN	0.9205	7.3887	4.2504
MSDCNN	0.9087	7.5139	4.4214
BDPN	0.918	7.7148	4.4522
DiCNN	0.8567	8.0256	5.5124
PNN	0.8849	12.6019	6.7233
A-PNN-FT	0.9132	7.6201	4.4536
FusionNet	0.8499	8.3823	6.0458

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

TABLE 9. A QUANTITATIVE COMPARISON OF THE OUTCOMES OF THE BENCHMARK ON THE FULL-RESOLUTION WV3 NEW YORK DATA SET (SEE FIGURE 11).

	D_λ	D_S	HQNR
	CS/MRA/VO		
EXP	0.0562	0.1561	0.7964
BT-H	0.0983	0.0829	0.8269
BDSD-PC	0.1554	0.0251	0.8234
C-GSA	0.1022	0.0747	0.8307
SR-D	**0.0199**	0.0369	0.944
MTF-GLP-HPM-R	0.0356	0.0679	0.8989
MTF-GLP-FS	0.0347	0.074	0.8939
TV	0.0234	0.0252	**0.952**
	ML		
PanNet	0.0376	**0.0162**	0.9468
DRPNN	0.1207	0.0392	0.8449
MSDCNN	0.1583	0.0557	0.7948
BDPN	0.1338	0.0563	0.8175
DiCNN	0.1023	0.0979	0.8098
PNN	0.1465	0.0835	0.7823
A-PNN-FT	0.051	0.0198	0.9302
FusionNet	0.0941	0.0882	0.826

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

ASSESSMENT ON WORLDVIEW-4 DATA

In this section, we repeat the same three tests but use WV4 data. Multiple reduced-resolution testing data sets are first evaluated. Then, another data set is used to assess the performance at reduced resolution and at full resolution.

PERFORMANCE ON EIGHT REDUCED-RESOLUTION TESTING DATA SETS

After evaluating the performance of the eight-band WV2 and WV3 data sets, this section mainly focuses on comparing the performance of the four-band WV4 data set acquired over Acapulco. Although we have a different spectral band number with respect to the eight-band data sets in the "Assessment on WorldView 2 Data" and "Assessment on WorldView-3 Data" sections, the testing procedure is the same. Indeed, all the compared pansharpening methods are evaluated on eight reduced-resolution samples extracted from the WV4 Acapulco testing data set in Table 3 (C) and Figure 10. These testing data sets share similar features with the training data.

Table 10 presents a quantitative comparison showing that the ML-based approaches yield better performance than the traditional techniques. The PanNet, belonging to the ML class, achieves the best indicators among all the methods. The rest of ML-based methods, i.e., the DRPNN, MSDCNN, BDPN, DiCNN, PNN, A-PNN-FT, and FusionNet, have similar performance, with small gaps among the three metrics exploited at reduced resolution. The DRPNN obtained the second-best Q4 and ERGAS indicators. The FusionNet got the second-best SAM. Among the traditional approaches, the MTF-GLP-HPM-R produced the best Q4, and the SR-D had the best SAM and ERGAS. The same conclusion as in the "Performance on 12 Reduced-Resolution Testing Data Sets" section about the relationship between the performance of ML-based approaches and traditional methods can be drawn.

PERFORMANCE ON THE REDUCED-RESOLUTION WORLDVIEW-4 ALICE SPRINGS DATA SET

This section investigates the performance of all the methods on a different reduced-resolution WV4 data set acquired over Alice Springs (see Figure 11). Table 11 reveals that the DRPNN and PanNet ML-based methods achieve the best Q4 and SAM, respectively, while the traditional SR-D approach has the best ERGAS. Overall, the quantitative performance of all the methods is similar. No approach obtains the best outcome on all the indexes. For example, some ML-based techniques, e.g., the A-PNN-FT, PanNet, and PNN, have a better SAM than several traditional methods, e.g., the BDSD-PC, C-GSA, and MTF-GLP-FS, whereas some traditional methods, e.g., the BT-H and SR-D, obtain a better

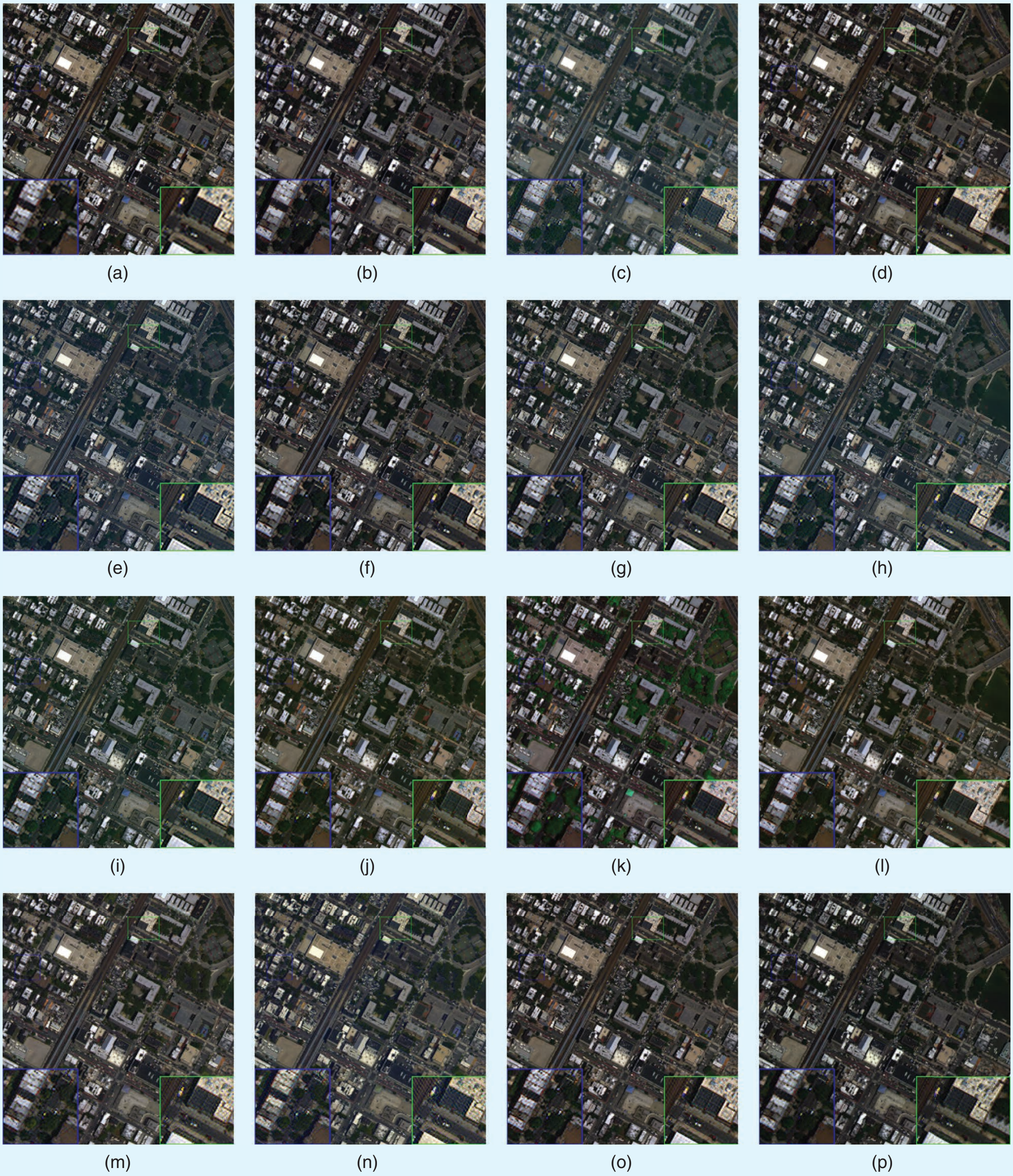


FIGURE 14. Visual comparisons in natural colors of the evaluated approaches on the full-resolution WV3 New York data set (see Figure 11). The (a) EXP, (b) BT-H, (c) BDSD-PC, (d) C-GSA, (e) SR-D, (f) MTF-GLP-HPM-R, (g) MTF-GLP-FS, (h) TV, (i) PanNet, (j) DRPNN, (k) MSDCNN, (l) BDPN, (m) DiCNN, (n) PNN, (o) A-PNN-FT, and (p) FusionNet.

SAM than some ML-based methods, such as the DRPNN, DiCNN, and FusionNet. Among the ML-based methods, although the DRPNN achieves the best Q4, its SAM value is significantly lower than that of the PanNet and A-PNN-FT. The FusionNet yields the worst metrics among all the ML-based methods. Figure 15 provides a visual comparison of the pansharpening approaches, showing that all the methods obtain excellent results, with high spatial fidelity in the urban area. In particular, traditional methods, such as the BT-H, C-GSA, MTF-GLP-HPM-R, and MTF-GLP-FS, display products with clearer spatial details than the ML-based methods (see the close-ups in Figure 15). Some other traditional methods, such as the SR-D and TV, show significant blur (see the blur and green close-ups in Figure 15).

TABLE 10. THE AVERAGE RESULTS OF THE APPROACHES BELONGING TO THE BENCHMARK FOR THE REDUCED-RESOLUTION WV4 ACAPULCO TESTING DATA SET, I.E., ON THE EIGHT WV4 TESTING DATA SETS IN TABLE 3 (C).

	Q4 (±STANDARD DEVIATION)	SAM (±STANDARD DEVIATION)	ERGAS (±STANDARD DEVIATION)
	CS/MRA/VO		
GT	1 ± 0	0 ± 0	0 ± 0
EXP	0.2638 ± 0.1579	3.9822 ± 0.5496	4.799 ± 0.9197
BT-H	0.6499 ± 0.0734	4.3582 ± 0.5607	4.633 ± 0.8555
BDSD-PC	0.6512 ± 0.068	3.6968 ± 0.5468	4.1906 ± 0.9251
C-GSA	0.6528 ± 0.0638	3.753 ± 0.5137	4.4599 ± 0.8186
SR-D	0.6564 ± 0.0895	3.6514 ± 0.4579	3.9887 ± 0.7342
MTF-GLP-HPM-R	0.6698 ± 0.0606	3.7980 ± 0.6864	4.2282 ± 0.8625
MTF-GLP-FS	0.6666 ± 0.0598	3.7776 ± 0.6527	4.2159 ± 0.8816
TV	0.5125 ± 0.1459	4.0344 ± 0.5386	4.26 ± 0.6768
	ML		
PanNet	**0.6963 ± 0.0842**	**3.371 ± 0.4221**	**3.6088 ± 0.6313**
DRPNN	0.681 ± 0.0845	3.4778 ± 0.4499	3.6706 ± 0.6433
MSDCNN	0.6739 ± 0.0849	3.4837 ± 0.4601	3.7052 ± 0.6661
BDPN	0.6535 ± 0.0834	3.5222 ± 0.4612	3.8269 ± 0.7144
DiCNN	0.6767 ± 0.0832	3.4555 ± 0.4453	3.7087 ± 0.6629
PNN	0.6793 ± 0.0822	3.4777 ± 0.4589	3.6894 ± 0.6613
A-PNN-FT	0.6787 ± 0.082	3.4271 ± 0.4425	3.6995 ± 0.6705
FusionNet	0.6759 ± 0.0805	3.3979 ± 0.4442	3.6842 ± 0.676

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

TABLE 11. A QUANTITATIVE COMPARISON OF THE OUTCOMES OF THE BENCHMARK ON THE REDUCED-RESOLUTION WV4 ALICE SPRINGS DATA SET (SEE FIGURE 11).

	Q4	SAM	ERGAS
	CS/MRA/VO		
GT	1	0	0
EXP	0.7901	4.588	5.8077
BT-H	0.9444	4.1988	3.2757
BDSD-PC	0.9431	4.7527	3.2996
C-GSA	0.9417	4.8812	3.3223
SR-D	0.9493	4.1597	**3.0647**
MTF-GLP-HPM-R	0.9432	5.1721	3.2724
MTF-GLP-FS	0.9432	4.9296	3.2437
TV	0.925	4.7857	3.6899
	ML		
PanNet	0.9486	**3.8737**	3.4154
DRPNN	**0.9521**	4.7059	3.2533
MSDCNN	0.9266	4.517	3.9181
BDPN	0.9439	4.4687	3.557
DiCNN	0.9347	4.8219	3.6978
PNN	0.9364	4.4324	3.5983
A-PNN-FT	0.9511	3.9217	3.1598
FusionNet	0.9261	4.9779	3.9561

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

PERFORMANCE ON THE FULL-RESOLUTION WORLDVIEW-4 ALICE SPRINGS DATA SET

Table 12 reports the quantitative results for the WV4 Alice Spring data set using data at the original (full) resolution (see Figure 11). Note that due to the absence of a GT image, we employ no reference indicators, such as the HQNR, D_λ, and D_S, to evaluate the quantitative performance. From the table, it is clear that some traditional and ML-based methods, such as the SR-D, TV, and A-PNN-FT, achieve the highest HQNR index values. Most of the ML-based approaches have better indexes than the rest of the traditional techniques, i.e., the BT-H, BDSD-PC, and C-GSA. Among all the ML-based methods, the A-PNN-FT, PanNet, and DRPNN belong to the best performance class. Moreover, the MSDCNN, BDPN, and DiCNN represent the second-best class, while the rest of the ML-based approaches (i.e., the PNN and FusionNet) achieve the lowest performance. Finally, the A-PNN-FT obtains the best ML-based quantitative outcome, corroborating the effectiveness of the fine-tuning strategy.

ASSESSMENT ON QUICKBIRD DATA

This section first investigates the performance on reduced-resolution and full-resolution testing sets, similar to the analysis conducted previously. Then, it evaluates the ability of the compared networks to generalize with respect to the acquisition sensor. We exploit ML-based methods trained on the QB training set but evaluate them on another four-band data set acquired by the Ikonos sensor.

PERFORMANCE ON SEVEN REDUCED-RESOLUTION TESTING DATA SETS

This section focuses on testing on seven reduced-resolution QB Indianapolis data sets that can be found in Figure 10. These testing data sets have a similar area and the same acquisition time as the training data set (see data ① in Table 3). Due to this, the ML-based approaches achieve better quantitative results than the compared traditional methods (see Table 13). The FusionNet produces the best Q4, SAM, and ERGAS indicators, and the PanNet, DRPNN, MSDCNN, DiCNN, and A-PNN-FT represent the second-best class. Comparing the ML-based approaches, the BDPN has relatively worse performance than the others but still outperformed the traditional techniques.

PERFORMANCE ON THE REDUCED-RESOLUTION QUICKBIRD SAN FRANCISCO DATA SET

These results assess all the methods on another reduced-resolution data set, acquired by the QB sensor over San Francisco (see Figure 11). In Table 14, we observe that the traditional approaches have better quantitative results than the ML-based methods (except for the PanNet). The C-GSA and BT-H approaches yield the lowest and second-lowest ERGAS, respectively. Among the ML-based methods, the PanNet has the best Q4, SAM, and ERGAS indicators, even better than those of all the traditional approaches. None of the other ML-based methods generates the best outcomes on all the indexes. The quantitative results for the rest of the ML-based approaches are not stable. For instance, the DRPNN yields the second-best Q4 among all the ML-based methods, but its SAM value is clearly larger than that of the FusionNet.

PERFORMANCE ON THE FULL-RESOLUTION QUICKBIRD SAN FRANCISCO DATA SET

The QB San Francisco data set in Figure 11 is also used at full resolution. From Table 15, reporting all the no-reference indexes, it is easy to see that the PanNet method obtains the best no-reference index, i.e., the HQNR, which means the best quantitative outcome. The SR-D traditional method and the ML-based FusionNet rank in second and third place, respectively. Overall, the traditional methods (except for the SR-D and TV) obtain poorer performance than most of the ML-based methods. The HQNR achieved by the DiCNN method is the lowest, demonstrating that the learned weights of the DiCNN network cannot fit the problem presented during the testing phase.

Figure 16 visually compares all the methods on the full-resolution QB San Francisco data set. From the figure, the ML-based methods, i.e., the PanNet and FusionNet, retain the clearest details, consistent with their HQNR performance in Table 15. Most of the other ML-based methods, such as the DRPNN, MSDCNN, BDPN, and A-PNN-FT, preserve the spatial content. Only the DiCNN and PNN seem to have relatively noticeable blur effects and artifacts. Among the traditional methods, the BDSD-PC shows a significant spectral distortion. The two MTF-based techniques, i.e., the MTF-GLP-HPM-R and MTF-GLP-FS, obtain high spatial fidelity, although they fail to generate promising HQNR values. Finally, TV and the SR-D have a spatial preservation similar to that of the A-PNN-FT.

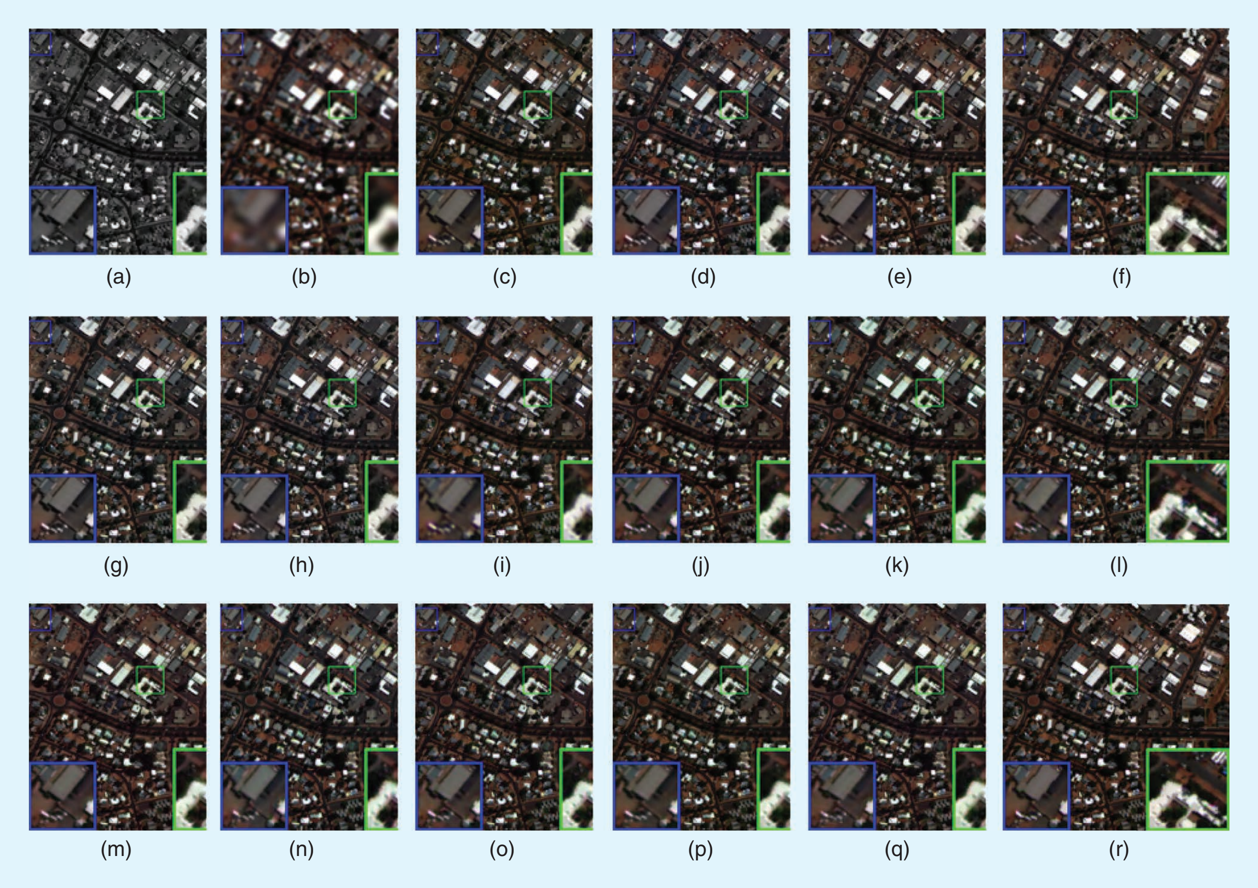


FIGURE 15. Visual comparisons in natural colors of the evaluated approaches on the reduced-resolution WV4 Alice Springs data set (see Figure 11). The (a) PAN, (b) EXP, (c) BT-H, (d) BDSD-PC, (e) C-GSA, (f) SR-D, (g) MTF-GLP-HPM-R, (h) MTF-GLP-FS, (i) TV, (j) PanNet, (k) DRPNN, (l) MSDCNN, (m) BDPN, (n) DiCNN, (o) PNN, (p) A-PNN-FT, (q) FusionNet, and (r) GT.

TABLE 12. A QUANTITATIVE COMPARISON OF THE OUTCOMES OF THE BENCHMARK ON THE FULL-RESOLUTION WV4 ALICE SPRINGS DATA SET (SEE FIGURE 11).

	D_λ	D_S	HQNR
	CS/MRA/VO		
EXP	0.0362	0.0322	0.9328
BT-H	0.0585	0.0625	0.8826
BDSD-PC	0.0644	0.0435	0.895
C-GSA	0.0668	0.0796	0.8589
SR-D	**0.0109**	0.0331	**0.9564**
MTF-GLP-HPM-R	0.0229	0.0608	0.9177
MTF-GLP-FS	0.023	0.0623	0.9161
TV	0.0251	**0.0237**	0.9518
	ML		
PanNet	0.012	0.0429	0.9456
DRPNN	0.0223	0.0333	0.9452
MSDCNN	0.0221	0.0641	0.9152
BDPN	0.026	0.0498	0.9255
DiCNN	0.0455	0.0358	0.9203
PNN	0.0195	0.0722	0.9097
A-PNN-FT	0.0195	0.0306	0.9505
FusionNet	0.0668	0.0274	0.9076

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

TABLE 13. THE AVERAGE RESULTS OF THE APPROACHES BELONGING TO THE BENCHMARK FOR THE REDUCED-RESOLUTION QB INDIANAPOLIS TESTING DATA SET, I.E., ON THE SEVEN QB TESTING DATA SETS IN TABLE 3 (D).

	Q4 (±STANDARD DEVIATION)	SAM (±STANDARD DEVIATION)	ERGAS (±STANDARD DEVIATION)
	CS/MRA/VO		
GT	1 ± 0	0 ± 0	0 ± 0
EXP	0.749 ± 0.017	4.5865 ± 0.4136	4.0991 ± 0.166
BT-H	0.8729 ± 0.0102	3.7376 ± 0.4094	3.0172 ± 0.1599
BDSD-PC	0.8643 ± 0.0107	4.0724 ± 0.5258	3.2261 ± 0.1212
C-GSA	0.8307 ± 0.0367	4.4207 ± 0.7011	3.5343 ± 0.4436
SR-D	0.8789 ± 0.0091	3.6989 ± 0.3694	2.9774 ± 0.1687
MTF-GLP-HPM-R	0.8628 ± 0.0151	3.9175 ± 0.7783	3.2746 ± 0.4262
MTF-GLP-FS	0.8513 ± 0.0152	4.0604 ± 0.7747	3.3176 ± 0.105
TV	0.8049 ± 0.0371	4.8419 ± 0.3162	3.9387 ± 0.4611
	ML		
PanNet	0.9575 ± 0.0072	1.9853 ± 0.1919	1.7365 ± 0.088
DRPNN	0.951 ± 0.0086	2.0873 ± 0.1875	1.8378 ± 0.0916
MSDCNN	0.9509 ± 0.0088	2.0771 ± 0.181	1.8565 ± 0.1025
BDPN	0.9238 ± 0.0113	2.5859 ± 0.1981	2.3305 ± 0.1474
DiCNN	0.951 ± 0.0088	2.0704 ± 0.1793	1.8764 ± 0.1086
PNN	0.9487 ± 0.0085	2.1556 ± 0.185	1.9054 ± 0.1048
A-PNN-FT	0.9585 ± 0.0074	1.8825 ± 0.1676	1.7086 ± 0.0963
FusionNet	**0.96 ± 0.0082**	**1.8298 ± 0.1391**	**1.647 ± 0.0918**

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

SENSOR GENERALIZATION ABILITY ASSESSED ON THE REDUCED-RESOLUTION IKONOS DATA SET

This section evaluates the network generalization ability for all the compared ML-based methods. The latter are trained on the four-band QB training set used in the previous sections. Then, we directly test the ML-based approaches, running them on a four-band Ikonos data set acquired over Toulouse. We also compare the ML-based methods with some traditional techniques. From Table 16, it is clear that the BT-H, TV, and the SR-D achieve the best Q4, SAM, and ERGAS, respectively. In contrast, the ML-based methods have poor performance, demonstrating weak network generalization. Overall, the traditional methods outperform all the ML-based approaches. Among the ML-based techniques, the PanNet and A-PNN-FT yield the best quantitative results on the three quality metrics. The other ML-based methods obtain worse performance.

Figure 17 depicts the fused products, showing the competitive performance for some traditional methods, i.e., the BT-H, C-GSA, MTF-GLP-HPM-R, and MTF-GLP-FS. Although the SR-D has the best ERGAS, some artifacts appear in the related outcome (see the blue close-up in Figure 17). Among the ML-based methods, all the compared approaches have similar spatial details. However, some of them, such as the DRPNN, DiCNN, and FusionNet, have a significant spectral distortion (see the color of the river in Figure 17). This is also corroborated by the SAM values in Table 16.

SENSOR GENERALIZATION ABILITY ASSESSED ON THE FULL-RESOLUTION IKONOS DATA SET

The analysis in the previous section is performed at full resolution, exploiting the Ikonos Toulouse data set. The A-PNN-FT yields the best overall performance (see Table 17). Indeed, thanks to the use of the fine-tuning strategy, the A-PNN-FT has a better network generalization ability than the other ML techniques. This is a good hint at future developments that could include this strategy to increase the generalization ability. Another ML approach achieving competitive performance with respect to traditional methods is the PanNet. Among the traditional methods, the SR-D obtains the highest performance. The C-GSA and TV also achieve promising results. Finally, it is worth pointing out that despite the HQNR index representing a state-of-the-art quality index, more research is needed on this topic [4]. The difficulty of ranking approaches belonging to different philosophies (e.g., classical against ML methods) is evident. Thus, results at reduced and full resolution can hardly be compared when referring to methods in different classes.

DISCUSSIONS

This section is devoted to final discussions about the ML-based approaches. Some aspects, such as convergence, testing and training times, the number of parameters, and so forth, are detailed in the following.

CONVERGENCE

Figure 18 exhibits the training loss and validation loss of all the compared ML-based approaches. The goal of this analysis is to show that the ML approaches converge but avoid the overfitting phenomenon. Observing the curves in Figure 18, we state that the goal is achieved by all the compared methods.

TESTING TIME

To evaluate the testing time of all the compared pansharpening methods, we employ four reduced-resolution WV3 testing data sets (see the "Performance on Four Reduced-Resolution Testing Data Sets" section for more details). Table 18 reports the average testing time for all the compared methods. Note that the traditional approaches are implemented on the CPU, while the ML-based methods exploit the GPU. From the table, it is easy to note that some traditional methods, such as the BT-H, BDSD-PC, MTF-GLP-HPM-R, and MTF-GLP-FS, run very fast, even though they are tested on the CPU. Other traditional approaches, i.e., the SR-D and TV, take more time (in particular, TV). The testing times of the ML-based methods are quite close (less than 1 s) to the very fast traditional techniques. This is because ML approaches use the GPU.

TRAINING TIME, NUMBER OF PARAMETERS, AND GIGA FLOATING-POINT OPERATIONS PER SECOND

We also investigate the training times of all the ML-based methods to evaluate the cost of the training. From the first row in Table 19, it is clear that the slowest method, i.e., the BDPN, needs almost one day to train the network on the WV3 training data set, while the fastest approach, i.e., the PanNet, can complete the training phase in 2 h. Looking at the number of parameters (the second row in Table 19), the BDPN has the highest value, while the DiCNN has the lowest. Finally, by evaluating the giga floating-point operations per second, the BDPN and DiCNN show extreme values.

HISTOGRAM COMPARISON OF ERROR MAPS

Figure 19 displays histograms of the errors between each fused image and the GT evaluated on the four reduced-resolution WV3 data sets used in the "Performance on Four Reduced-Resolution Testing Data Sets" section. From the figure, we see that the A-PNN-FT and FusionNet have smaller standard deviations, indicating better overall results for this test case. Moreover, the range proportion (RP) within $[-0.02, 0.02]$ (the larger the RP, the better the performance) has been reported in Figure 19. Again, the best values are obtained by the FusionNet and A-PNN-FT.

PERFORMANCE VERSUS THE NUMBER OF PARAMETERS

Figure 20 investigates the relationship between quantitative performance and the number of parameters, aiming to illustrate the effectiveness of the compared ML-based methods. Again, four reduced-resolution data sets acquired by the WV3 sensor, which were also used in the "Performance on Four Reduced-Resolution Testing Data Sets" section, are exploited. The quality is measured using the three

TABLE 14. A QUANTITATIVE COMPARISON OF THE OUTCOMES OF THE BENCHMARK ON THE REDUCED-RESOLUTION QB SAN FRANCISCO DATA SET (SEE FIGURE 11).

	Q4	SAM	ERGAS
	CS/MRA/VO		
GT	1	0	0
EXP	0.5759	9.1351	10.9039
BT-H	0.8942	7.5545	5.2697
BDSD-PC	0.8788	8.745	5.7536
C-GSA	0.8961	7.4711	**5.2337**
SR-D	0.8831	7.8766	5.5558
MTF-GLP-HPM-R	0.8919	8.489	5.4754
MTF-GLP-FS	0.877	8.7026	5.7855
TV	0.8802	8.4317	6.0476
	ML		
PanNet	**<u>0.9074</u>**	**<u>6.9841</u>**	<u>5.3314</u>
DRPNN	0.8969	8.253	5.9467
MSDCNN	0.8768	7.5988	5.6965
BDPN	0.883	8.4378	5.9962
DiCNN	0.8062	11.211	8.7013
PNN	0.8301	10.1118	6.8375
A-PNN-FT	0.8586	7.8767	6.2049
FusionNet	0.8614	7.3459	6.342

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

TABLE 15. A QUANTITATIVE COMPARISON OF THE OUTCOMES OF THE BENCHMARK ON THE FULL-RESOLUTION QB SAN FRANCISCO DATA SET (SEE FIGURE 11).

	D_λ	D_S	HQNR
	CS/MRA/VO		
EXP	0.047	0.1571	0.8033
BT-H	0.0925	0.0925	0.8236
BDSD-PC	0.1383	0.0476	0.8207
C-GSA	0.0818	0.1114	0.8159
SR-D	**0.0144**	0.0362	0.9499
MTF-GLP-HPM-R	0.0343	0.1126	0.857
MTF-GLP-FS	0.0372	0.1323	0.8354
TV	0.0269	0.0513	0.9233
	ML		
PanNet	<u>0.0224</u>	0.0264	**<u>0.9518</u>**
DRPNN	0.0662	0.021	0.9142
MSDCNN	0.0771	0.0268	0.8982
BDPN	0.0621	0.0708	0.8715
DiCNN	0.0939	0.1244	0.7933
PNN	0.1071	0.0671	0.833
A-PNN-FT	0.0364	0.0303	0.9344
FusionNet	0.0388	**<u>0.0139</u>**	0.9479

Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

TABLE 16. A QUANTITATIVE COMPARISON OF THE OUTCOMES OF THE BENCHMARK ON THE REDUCED-RESOLUTION IKONOS TOULOUSE DATA SET (SEE FIGURE 11).

	Q4	SAM	ERGAS
	CS/MRA/VO		
GT	1	0	0
EXP	0.4795	5.1823	6.3953
BT-H	**0.912**	3.4491	2.9962
BDSD-PC	0.9094	2.9576	2.9309
C-GSA	0.9006	2.9667	3.1751
SR-D	0.9108	2.9571	**2.8708**
MTF-GLP-HPM-R	0.9105	3.1454	2.9727
MTF-GLP-FS	0.9076	3.0906	3.0104
TV	0.9023	**2.8455**	2.9508
	ML		
PanNet	0.8826	3.901	3.6584
DRPNN	0.8884	5.9745	4.2175
MSDCNN	0.8736	4.1837	3.5057
BDPN	0.8783	4.0874	3.7266
DiCNN	0.8143	6.2024	5.5863
PNN	0.8406	4.6105	3.9881
A-PNN-FT	0.8838	3.6224	3.3742
FusionNet	0.8159	4.2536	4.071

The ML-based approaches are trained on the QB data set.
Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

TABLE 17. A QUANTITATIVE COMPARISON OF THE OUTCOMES OF THE BENCHMARK ON THE FULL-RESOLUTION IKONOS TOULOUSE DATA SET (SEE FIGURE 11).

	D_λ	D_S	HQNR
	CS/MRA/VO		
EXP	0.056	0.1723	0.7813
BT-H	0.069	0.0812	0.8554
BDSD-PC	0.088	0.0617	0.8557
C-GSA	0.0641	0.0371	0.9012
SR-D	**0.0163**	0.0522	0.9323
MTF-GLP-HPM-R	0.0275	0.0853	0.8896
MTF-GLP-FS	0.0285	0.0908	0.8834
TV	0.0472	0.0307	0.9235
	ML		
PanNet	0.0239	0.0344	0.9425
DRPNN	0.0723	0.0197	0.9095
MSDCNN	0.083	0.0261	0.893
BDPN	0.0699	0.0405	0.8924
DiCNN	0.1507	**0.0156**	0.8361
PNN	0.0823	0.0453	0.8761
A-PNN-FT	0.0282	0.0194	**0.9529**
FusionNet	0.0662	0.0263	0.9093

The ML-based approaches are trained on the QB data set.
Bold: the best among all the compared methods; underline: the best among all the ML-based methods.

quality indexes at reduced resolution (i.e., the Q8, ERGAS, and SAM). Optimal results are plotted in the top-left area for the Q8, indicating high Q8 values with few parameters. For the ERGAS and SAM, the optimal area is located in the bottom-left part of the plot. The closer the methods are to the optimal areas, the better the tradeoff between quality and computational burden. Examining Figure 20, we note that the A-PNN-FT and FusionNet achieve excellent performance on the data used in this analysis for all three quality metrics.

CONCLUDING REMARKS

In this article, we presented the first critical comparison of pansharpening approaches based on the ML paradigm. A complete review of the ML literature was conducted. Then, eight state-of-the-art solutions for sharpening MS images using PAN data were compared. To this end, a toolbox exploiting a common software platform and open source library for all the ML approaches was developed. All the ML approaches were implemented, exploiting the common software platform (we selected PyTorch for this). The developed toolbox will be distributed free to the community. A careful tuning phase was performed to ensure the highest performance for each of the compared approaches.

A broad experimental analysis, exploiting different test cases, was conducted with the aim of assessing the performance of each ML-based state-of-the-art approach. Widely used sensors for pansharpening were involved (i.e., WV2, WV3, WV4, QB, and Ikonos). Assessments at reduced resolution and full resolution were considered. The comparison of ML-based approaches was enlarged to state-of-the-art methods belonging to different paradigms (i.e., CS, MRA, and VO). The generalization ability of the networks with respect to changes of the acquisition sensor and scenario was also reported. Finally, a wide computational analysis was presented in the "Discussions" section.

ML-based approaches demonstrated outstanding performance in scenarios close to those during the training phase. Reduced performance (in particular, in comparison with recent state-of-the-art traditional methods) was observed when a completely different scenario was used in the testing phase, thus showing a limited generalization ability of these approaches. However, fine-tuning proved to be of value in remedying the issue, guaranteeing high performance even in these challenging test cases. The computational burden, measured during the testing phase, of the compared ML approaches can be considered adequate, even in comparison with the fastest traditional methods. At any rate, the training phase is still time-consuming for several approaches, even requiring one day (see the BDPN case) for training with a relatively small number of samples.

Finally, we want to draw some guidelines for the development of new ML-based pansharpening approaches. Indeed, focusing on the analyzed ML-based pansharpening approaches, it can be remarked that skip connections can help ML-based methods obtain faster convergence. The design of multiscaled architectures (including bidirectional structures) can support better extraction and learning of features. Furthermore, fine-tuning and learning in a specific domain (i.e., not in the original image domain) can increase the generalization ability of the networks.

However, challenges still exist, representing room for improvement for researchers in the future. Specifically, as pointed out, the computational burden is still an open

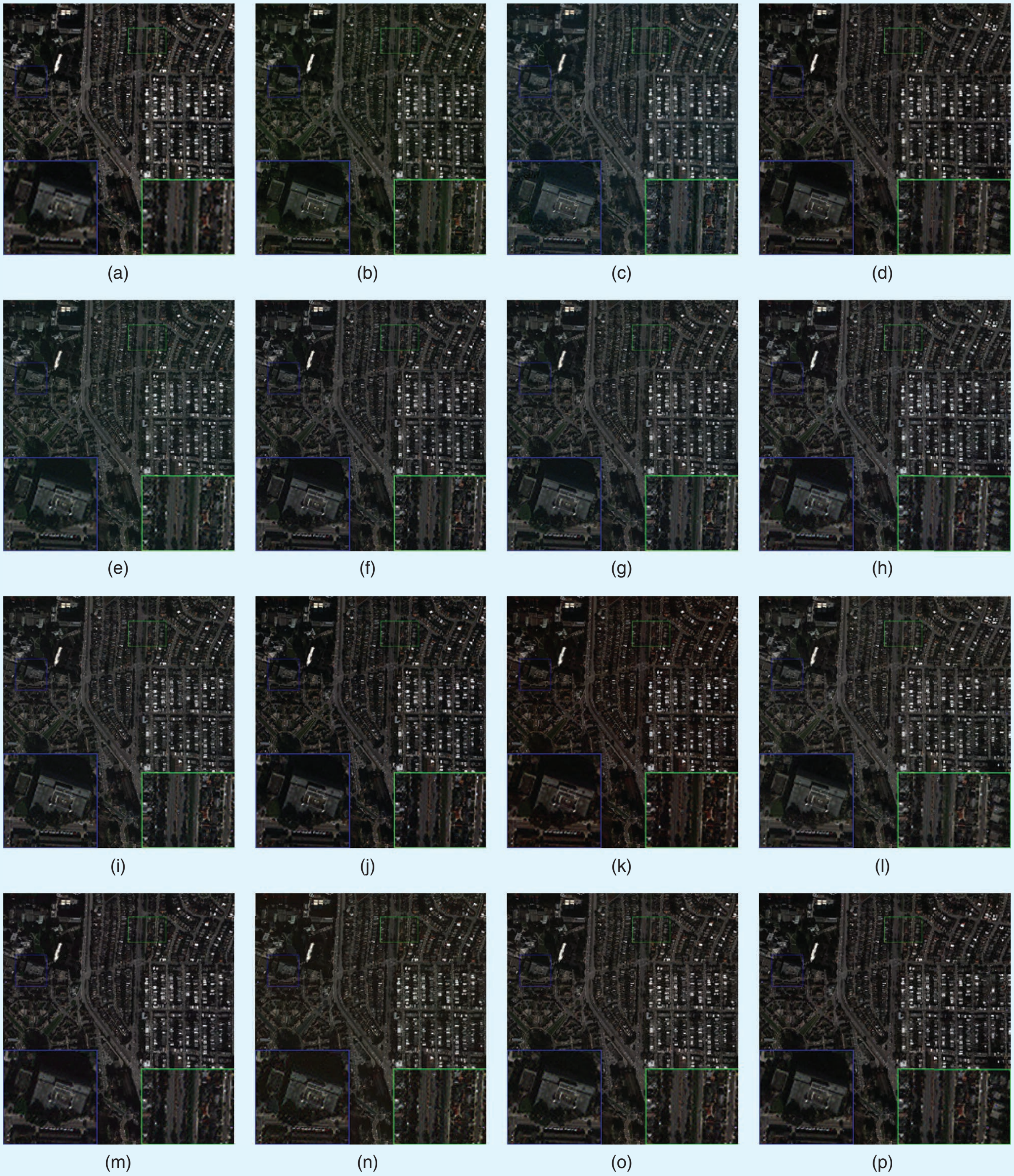


FIGURE 16. Visual comparisons in natural colors of the evaluated approaches on the full-resolution San Francisco data set (see Figure 11). The (a) EXP, (b) BT-H, (c) BDSD-PC, (d) C-GSA, (e) SR-D, (f) MTF-GLP-HPM-R, (g) MTF-GLP-FS, (h) TV, (i) PanNet, (j) DRPNN, (k) MSDCNN, (l) BDPN, (m) DiCNN, (n) PNN, (o) A-PNN-FT, and (p) FusionNet.

issue pushing researchers to develop networks with fewer parameters (even getting fast convergence) while ensuring a network's effectiveness. Generalization is limited for most new developments in ML for pansharpening. This is a crucial point to be addressed to move toward ML products for remote sensing image fusion in a commercial environment.

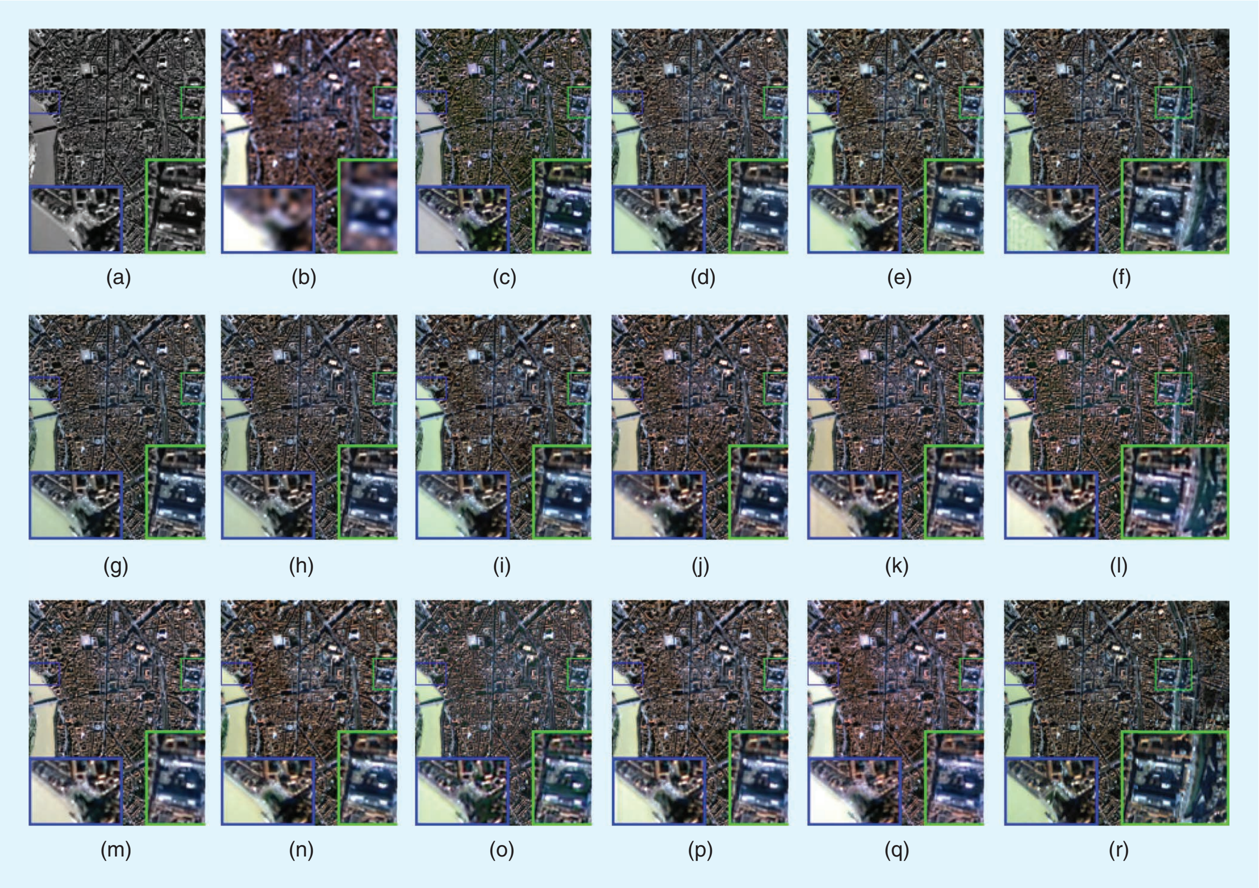


FIGURE 17. Visual comparisons in natural colors of the evaluated approaches on the reduced-resolution Ikonos Toulouse data set (see Figure 11). The (a) PAN, (b) EXP, (c) BT-H, (d) BDSD-PC, (e) C-GSA, (f) SR-D, (g) MTF-GLP-HPM-R, (h) MTF-GLP-FS, (i) TV, (j) PanNet, (k) DRPNN, (l) MSDCNN, (m) BDPN, (n) DiCNN, (o) PNN, (p) A-PNN-FT, (q) FusionNet, and (r) GT.

TABLE 18. THE COMPARISON OF TESTING TIMES (IN SECONDS) FOR ALL THE COMPARED METHODS.

	EXP	BT-H	BDSD-PC	C-GSA	SR-D	MTF-GLP-HPM-R	MTF-GLP-FS	TV
Testing time	**0.007**	0.092	0.234	1.305	7.138	0.246	0.314	31.232
	PanNet	DRPNN	MSDCNN	BDPN	DiCNN	PNN	A-PNN-FT	FusionNet
Testing time	0.339	0.337	0.442	0.493	0.37	0.456	0.921	0.376

Note that the traditional methods (first row) are implemented on the CPU, and the ML-based approaches (second row) exploit the GPU. The times are computed on four reduced-resolution WV3 testing data sets.

TABLE 19. THE COMPARISON OF THE TRAINING TIMES (HOURS: MINUTES), NUMBER OF PARAMETERS, AND GIGA FLOATING-POINT OPERATIONS PER SECOND (GFLOPS) FOR ALL THE COMPARED ML-BASED METHODS.

	PanNet	DRPNN	MSDCNN	BDPN	DiCNN	PNN	A-PNN-FT	FusionNet
Training time	**1:46**	4:42	3:08	23:22	8:21	8:4	7:55	3:01
Number of parameters	78,504	433,465	228,556	1,484,412	**47,369**	104,36	104,36	76,308
GFlops	0.32	1.78	0.91	3.8	**0.19**	0.29	0.22	0.32

The WV3 training data set is used as the reference for this evaluation.

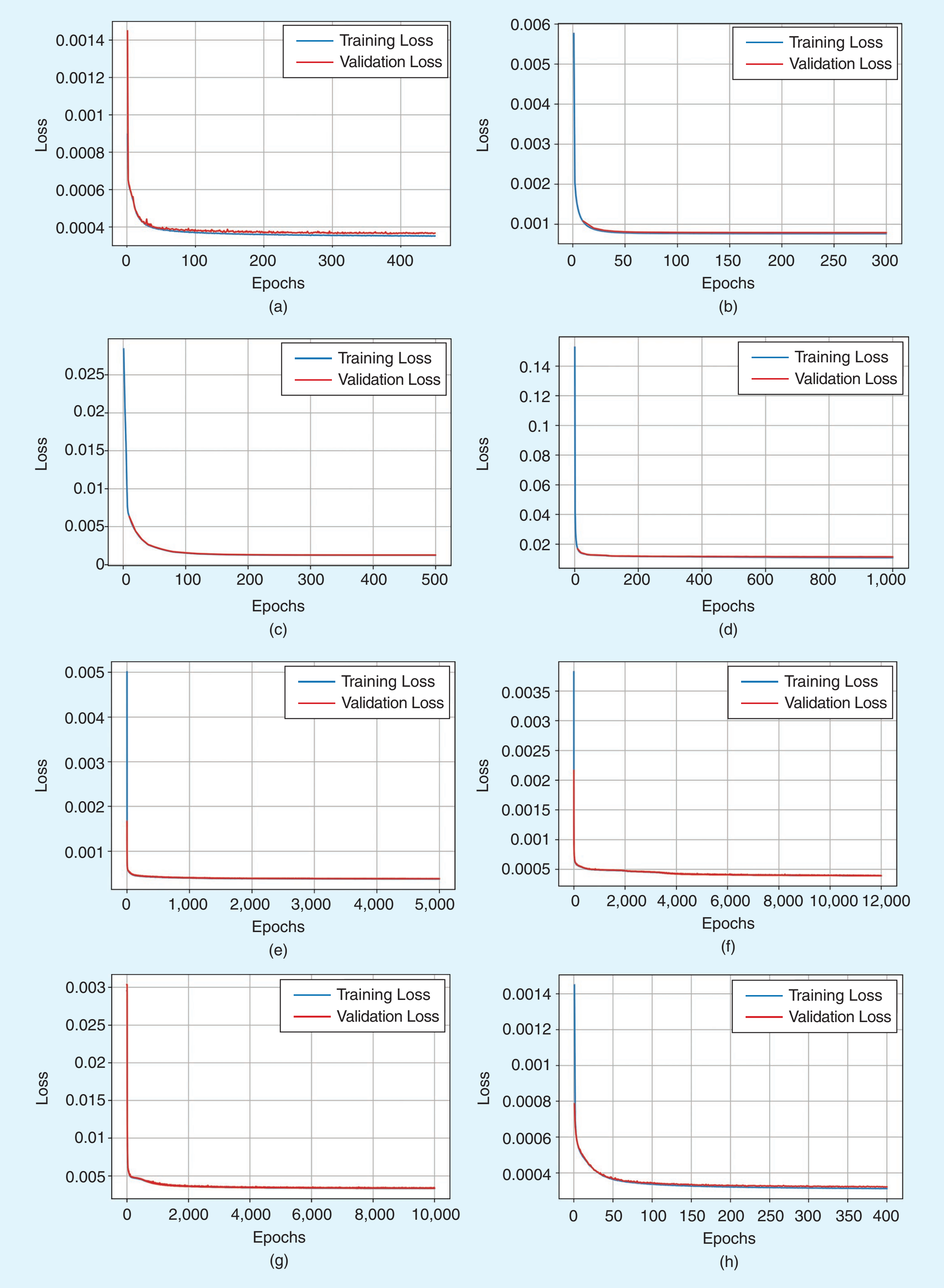


FIGURE 18. The convergence curves for all the compared ML-based methods. The corresponding loss functions are reported in Table 2. The (a) PanNet, (b) DRPNN, (c) MSDCNN, (d) BDPN, (e) DiCNN, (f) PNN, (g) A-PNN-FT, and (h) FusionNet.

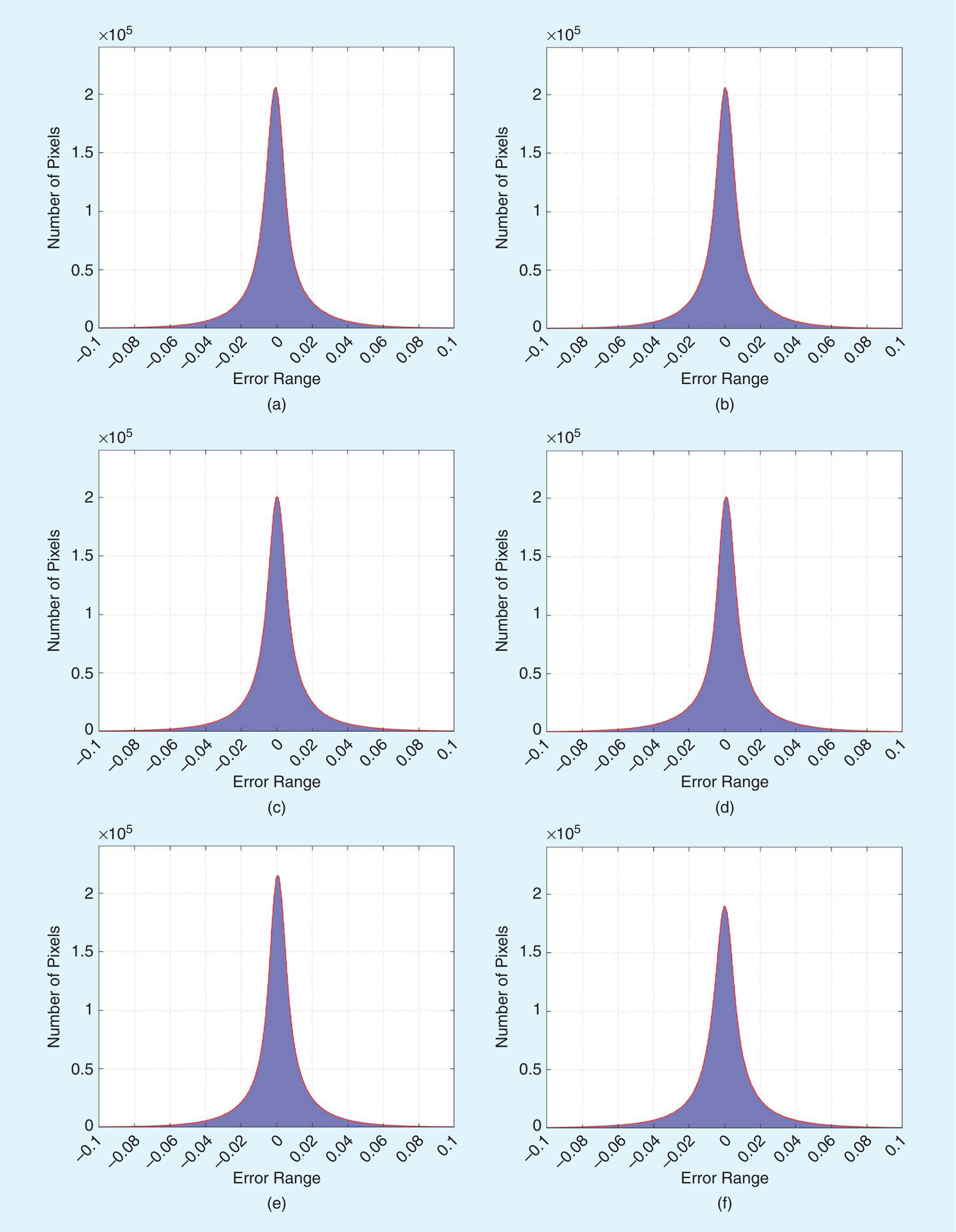
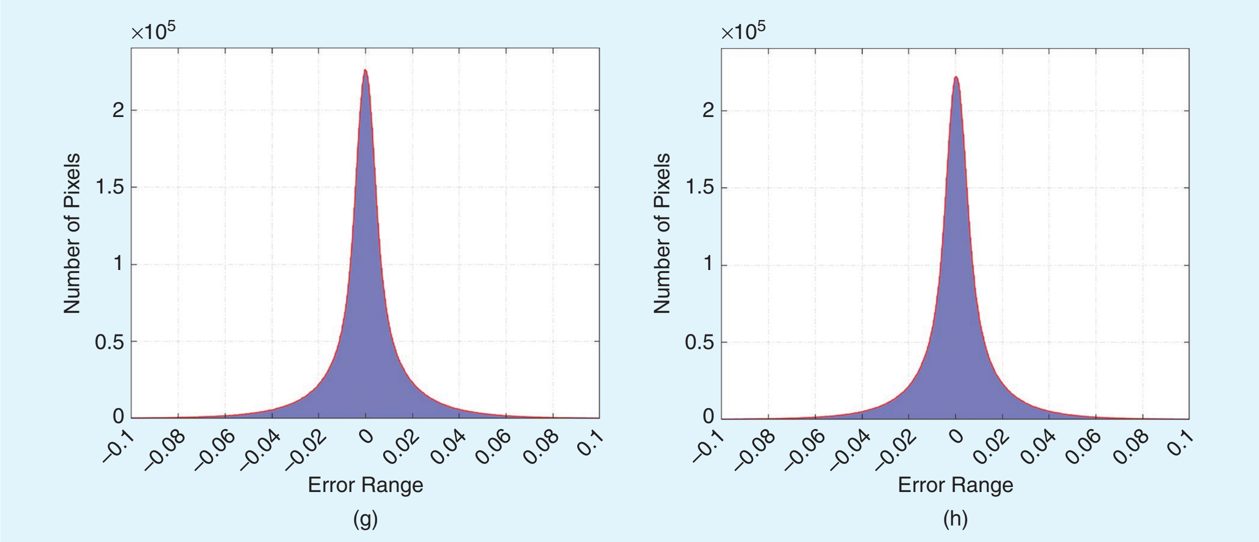
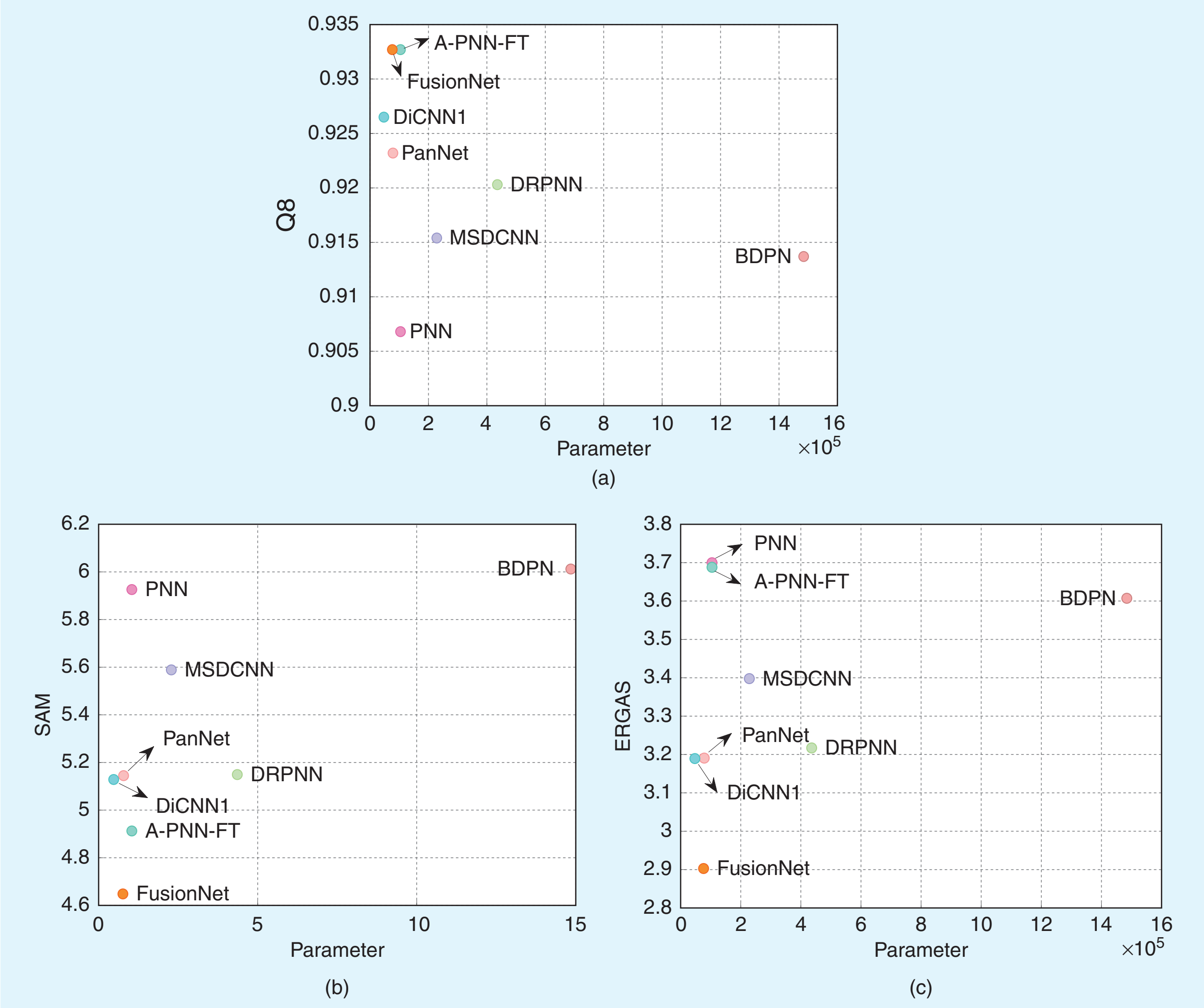


FIGURE 19. The comparison of the error histograms for all the ML-based methods. The error is computed between each fused image and the GT on four reduced-resolution WV3 data sets, which were also used in the "Performance on Four Reduced-Resolution Training Data Sets" section. Synthetic indexes, such as the standard deviation and range proportion (RP), are reported. The best results are in boldface. (a) The PanNet (standard deviation/RP = 0.019/0.827). (b) The DRPNN (0.019/0.829). (c) The MSDCNN (0.020/0.818). (d) The BDPN (0.022/0.804). (e) The DiCNN (0.019/0.831). (f) The PNN (0.021/0.807). (*continued*)



(g) (h)

FIGURE 19. (*continued*) (g) The A-PNN-FT (0.018/0.836). (h) The FusionNet (**0.017/0.849**).



(a)

(b) (c)

FIGURE 20. The comparison of the quantitative performance versus the number of parameters on four reduced-resolution WV3 data sets, which were also used in the "Performance on Four Reduced-Resolution Training Data Sets" section. (a) Q8, (b) SAM, and (c) ERGAS measurements are provided.

The original idea of working at reduced resolution to obtain labels to train networks is helpful. However, it is based on the hypothesis of "invariance among scales," which not be valid. Thus, as noted in our literature review, new (unsupervised) approaches based on loss functions measuring similarities at full resolution have been developed. This is an interesting research line, but developments are still required, even considering the need for new studies about more accurate quality metrics at full resolution.

ACKNOWLEDGMENT

We thank Xiao Wu and Ran Ran, graduate students from the University of Electronic Science and Technology of China, who provided great support to the realization of the toolbox presented in this article.

AUTHOR INFORMATION

Liang-Jian Deng (liangjian.deng@uestc.edu.cn) is with the School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, 611731, China. He is a Member of IEEE.

Gemine Vivone (gemine.vivone@imaa.cnr.it) is with the Institute of Methodologies for Environmental Analysis, Tito Scalo, 85050, Italy. He is a Senior Member of IEEE.

Mercedes E. Paoletti (mpaolett@ucm.es) is with the Department of Computer Architecture and Automatics, Faculty of Computer Science, Complutense University of Madrid, Madrid, 28040, Spain. She is a Senior Member of IEEE.

Giuseppe Scarpa (giscarpa@unina.it) is with the Department of Electrical Engineering and Information Technology, University Federico II, Naples, 80125, Italy. He is a Senior Member of IEEE.

Jiang He (jiang_he@whu.edu.cn) is with the School of Geodesy and Geomatics, Wuhan University, Wuhan, 430079, China.

Yongjun Zhang (zhangyj@whu.edu.cn) is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, 430079, China. He is a Member of IEEE.

Jocelyn Chanussot (jocelyn.chanussot@gipsa-lab.grenoble-inp.fr) is with the University of Grenoble Alpes, Grenoble, 38000, France. He is a Fellow of IEEE.

Antonio Plaza (aplaza@unex.es) is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, 10003, Spain. He is a Fellow of IEEE.

REFERENCES

[1] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli, *Remote Sensing Image Fusion*. Boca Raton, FL, USA: CRC Press, Jan. 2015.

[2] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015, doi: 10.1109/TGRS.2014.2361734.

[3] P. Ghamisi *et al.*, "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 6–39, Mar. 2019, doi: 10.1109/MGRS.2018.2890023.

[4] G. Vivone *et al.*, "A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 1, pp. 53–81, Mar. 2021, doi: 10.1109/MGRS.2020.3019315.

[5] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, Nov. 2016, doi: 10.1016/j.inffus.2016.03.003.

[6] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, Jan. 2017, doi: 10.1016/j.inffus.2016.05.004.

[7] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu, "Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges," *Inf. Fusion*, vol. 46, pp. 102–113, Mar. 2019, doi: 10.1016/j.inffus.2018.05.006.

[8] W. Carper, T. Lillesand, and R. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogrammetric Eng. Remote Sens.*, vol. 56, pp. 459–467, Apr. 1990.

[9] P. S. Chavez Jr., S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogrammetric Eng. Remote Sens.*, vol. 57, pp. 295–303, Mar. 1991.

[10] P. S. Chavez Jr. and A. W. Kwarteng, "Extracting spectral contrast in Landsat thematic mapper image data using selective principal component analysis," *Photogrammetric Eng. Remote Sens.*, vol. 55, pp. 339–348, Jan. 1989.

[11] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogrammetric Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, 1992.

[12] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," *U.S. Patent 6 011 875*, 2000.

[13] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007, doi: 10.1109/TGRS.2007.901007.

[14] G. A. Licciardi, M. M. Khan, J. Chanussot, A. Montanvert, L. Condat, and C. Jutten, "Fusion of hyperspectral and panchromatic images using multiresolution analysis and nonlinear PCA band reduction," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 207, Sep. 2012, doi: 10.1186/1687-6180-2012-207.

[15] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, Jan. 2011, doi: 10.1109/TGRS.2010.2051674.

[16] R. Restaino, M. Dalla Mura, G. Vivone, and J. Chanussot, "Context-adaptive pansharpening based on image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 753–766, Feb. 2017, doi: 10.1109/TGRS.2016.2614367.

[17] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images,"

IEEE Trans. Geosci. Remote Sens., vol. 46, no. 1, pp. 228–236, Jan. 2008, doi: 10.1109/TGRS.2007.907604.

[18] G. Vivone, "Robust band-dependent spatial-detail approaches for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6421–6433, Sep. 2019, doi: 10.1109/TGRS.2019.2906073.

[19] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Commun. Lett.*, vol. 31, no. 4, pp. 532–540, Apr. 1983, doi: 10.1109/TCOM.1983.1095851.

[20] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*, vol. 103, A. Antoniadis and G. Oppenheim, Eds. New York, NY, USA: Springer-Verlag, 1995, pp. 281–299.

[21] J. L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002, doi: 10.1109/TIP.2002.1014998.

[22] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005, doi: 10.1109/TIP.2005.859376.

[23] G. Vivone, R. Restaino, G. Licciardi, M. D. Mura, and J. Chanussot, "Multiresolution analysis and component substitution techniques for hyperspectral pansharpening," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 2649–2652, doi: 10.1109/IGARSS.2014.6947018.

[24] L. Alparone, S. Baronti, B. Aiazzi, and A. Garzelli, "Spatial methods for multispectral pansharpening: Multiresolution analysis demystified," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2563–2576, May 2016, doi: 10.1109/TGRS.2015.2503045.

[25] S. Zheng, W. Shi, J. Liu, and J. Tian, "Remote sensing image fusion using multiscale mapped LS-SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1313–1322, May 2008, doi: 10.1109/TGRS.2007.912737.

[26] 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, Jun. 2016, doi: 10.1109/TIP.2016.2556944.

[27] G. Vivone, L. Alparone, A. Garzelli, and S. Lolli, "Fast reproducible pansharpening based on instrument and acquisition modeling: AWLP revisited," *Remote Sens.*, vol. 11, no. 19, pp. 2315:1–2315:23, Oct. 2019, doi: 10.3390/rs11192315.

[28] R. Restaino, G. Vivone, P. Addesso, and J. Chanussot, "A pansharpening approach based on multiple linear regression estimation of injection coefficients," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 102–106, Jan. 2020, doi: 10.1109/LGRS.2019.2914093.

[29] Y. Zhang, "A new merging method and its spectral and spatial effects," *Int. J. Remote Sens.*, vol. 20, no. 10, pp. 2003–2014, 1999, doi: 10.1080/014311699212317.

[30] 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, Dec. 2017, doi: 10.1109/LGRS.2017.2761021.

[31] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogrammetric Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006, doi: 10.14358/PERS.72.5.591.

[32] G. Vivone *et al.*, "Pansharpening based on semiblind deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1997–2010, Apr. 2015, doi: 10.1109/TGRS.2014.2351754.

[33] G. Vivone, P. Addesso, R. Restaino, M. Dalla Mura, and J. Chanussot, "Pansharpening based on deconvolution for multiband filter estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 540–553, Jan. 2019, doi: 10.1109/TGRS.2018.2858288.

[34] G. Vivone and J. Chanussot, "Fusion of short-wave infrared and visible near-infrared WorldView-3 data," *Inf. Fusion*, vol. 61, pp. 71–83, Sep. 2020, doi: 10.1016/j.inffus.2020.03.012.

[35] L. Alparone, A. Garzelli, and G. Vivone, "Inter-sensor statistical matching for pansharpening: Theoretical issues and practical solutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4682–4695, Aug. 2017, doi: 10.1109/TGRS.2017.2697943.

[36] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005, doi: 10.1109/TGRS.2005.856106.

[37] M. Ghahremani and H. Ghassemian, "Remote-sensing image fusion based on curvelets and ICA," *Int. J. Remote Sens.*, vol. 36, no. 16, pp. 4131–4143, Nov. 2015, doi: 10.1080/01431161.2015.1071897.

[38] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pansharpening method via a combined adaptive-PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008, doi: 10.1109/TGRS.2008.916211.

[39] W. Liao *et al.*, "Two-stage fusion of thermal hyperspectral and visible RGB image by PCA and guided filter," in *Proc. 7th IEEE Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2015, pp. 1–4, doi: 10.1109/WHISPERS.2015.8075405.

[40] P. Liu, L. Xiao, and T. Li, "A variational pan-sharpening method based on spatial fractional-order geometry and spectral–Spatial low-rank priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1788–1802, Mar. 2018, doi: 10.1109/TGRS.2017.2768386.

[41] L. J. Deng, G. Vivone, W. Guo, M. D. Mura, and J. Chanussot, "A variational pansharpening approach based on reproducible kernel Hilbert space and Heaviside function," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4330–4344, Sep. 2018, doi: 10.1109/TIP.2018.2839531.

[42] C. H. Wang, C.-H. Lin, J. Bioucas-Dias, W. C. Zheng, and K. H. Tseng, "Panchromatic sharpening of multispectral satellite imagery via an explicitly defined convex self-similarity regularization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Nov. 2019, pp. 3129–3132, doi: 10.1109/IGARSS.2019.8900610.

[43] T. Wang, F. Fang, F. Li, and G. Zhang, "High-quality Bayesian pansharpening," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 227–239, Aug. 2019, doi: 10.1109/TIP.2018.2866954.

[44] L. J. Deng, M. Feng, and X. C. Tai, "The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior," *Inf. Fusion*, vol. 52, pp. 76–89, Dec. 2019, doi: 10.1016/j.inffus.2018.11.014.

[45] R. Dian, S. Li, B. Sun, and A. Guo, "Recent advances and new guidelines on hyperspectral and multispectral image fusion," *Inf. Fusion*, vol. 69, pp. 40–51, May 2021, doi: 10.1016/j.inffus.2020.11.001.

[46] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, Oct. 2019, doi: 10.1109/TIP.2019.2916734.

[47] Z.-C. Wu *et al.*, "A new variational approach based on proximal deep injection and gradient intensity similarity for spatio-spectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6277–6290, Oct. 2020, doi: 10.1109/JSTARS.2020.3030129.

[48] J. Duran, A. Buades, B. Coll, C. Sbert, and G. Blanchet, "A survey of pansharpening methods with a new band-decoupled variational model," *ISPRS J. Photogrammetry Remote Sens.*, vol. 125, pp. 78–105, Mar. 2017, doi: 10.1016/j.isprsjprs.2016.12.013.

[49] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for P+XS image fusion," *Int. J. Comput. Vis.*, vol. 69, no. 1, pp. 43–58, Aug. 2006, doi: 10.1007/s11263-006-6852-x.

[50] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "A new pansharpening algorithm based on total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 318–322, Jan. 2014, doi: 10.1109/LGRS.2013.2257669.

[51] X. He, L. Condat, J. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4160–4174, Sep. 2014, doi: 10.1109/TIP.2014.2333661.

[52] H. A. Aly and G. Sharma, "A regularized model-based optimization framework for pan-sharpening," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2596–2608, Apr. 2014, doi: 10.1109/TIP.2014.2316641.

[53] M. Möller, T. Wittman, A. L. Bertozzi, and M. Burger, "A variational approach for sharpening high dimensional images," *SIAM J. Imag. Sci.*, vol. 5, no. 1, pp. 150–178, Jan. 2012, doi: 10.1137/100810356.

[54] G. Zhang, F. Fang, A. Zhou, and F. Li, "Pan-sharpening of multi-spectral images using a new variational model," *Int. J. Remote Sens.*, vol. 36, no. 5, pp. 1484–1508, Mar. 2015, doi: 10.1080/01431161.2015.1014973.

[55] F. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, "Model-based reduced-rank pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 656–660, Apr. 2020, doi: 10.1109/LGRS.2019.2926681.

[56] D. Fasbender, J. Radoux, and P. Bogaert, "Bayesian data fusion for adaptable image pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1847–1857, Jun. 2008, doi: 10.1109/TGRS.2008.917131.

[57] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3834–3843, Nov. 2009, doi: 10.1109/TGRS.2009.2017737.

[58] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model-based fusion of multi- and hyperspectral images using PCA and wavelets," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2652–2663, May 2015, doi: 10.1109/TGRS.2014.2363477.

[59] Y. Zhang, A. Duijster, and P. Scheunders, "A Bayesian restoration approach for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3453–3462, Sep. 2012, doi: 10.1109/TGRS.2012.2184122.

[60] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011, doi: 10.1109/TGRS.2010.2067219.

[61] C. Jiang, H. Zhang, H. Shen, and L. Zhang, "A practical compressed sensing-based pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 629–633, Jul. 2015, doi: 10.1109/LGRS.2011.2177063.

[62] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4779–4789, Sep. 2013, doi: 10.1109/TGRS.2012.2230332.

[63] M. Cheng, C. Wang, and J. Li, "Sparse representation based pansharpening using trained dictionary," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 293–297, Jan. 2014, doi: 10.1109/LGRS.2013.2256875.

[64] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, May 2013, doi: 10.1109/TGRS.2012.2213604.

[65] X. X. Zhu, C. Grohnfeld, and R. Bamler, "Exploiting joint sparsity for pansharpening: The J-sparseFI algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2664–2681, May 2016, doi: 10.1109/TGRS.2015.2504261.

[66] M. R. Vicinanza, R. Restaino, G. Vivone, M. Dalla Mura, G. Licciardi, and J. Chanussot, "A pansharpening method based on the sparse representation of injected details," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 180–184, Jan. 2015, doi: 10.1109/LGRS.2014.2331291.

[67] X. Tian, Y. Chen, C. Yang, X. Gao, and J. Ma, "A variational pansharpening method based on gradient sparse representation," *IEEE Signal Process. Lett.*, vol. 27, pp. 1180–1184, Jul. 2020, doi: 10.1109/LSP.2020.3007325.

[68] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang, "A new pansharpening method with deep neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1037–1041, May 2015, doi: 10.1109/LGRS.2014.2376034.

[69] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, p. 594, Jul. 2016, doi: 10.3390/rs8070594.

[70] J. Zhong, B. Yang, G. Huang, F. Zhong, and Z. Chen, "Remote sensing image fusion with convolutional neural network," *Sens. Imag.*, vol. 17, no. 1, pp. 1–16, Jun. 2016, doi: 10.1007/s11220-016-0135-6.

[71] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017, doi: 10.1109/LGRS.2017.2736020.

[72] Y. Rao, L. He, and J. Zhu, "A residual convolutional neural network for pan-sharpening," in *Proc. Int. Workshop Remote Sens. Intell. Process.*, May 2017, pp. 1–4, doi: 10.1109/RSIP.2017.7958807.

[73] J. Yang, X. Fu, Y. Hu, Y. Huang, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1753–1761, doi: 10.1109/ICCV.2017.193.

[74] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018, doi: 10.1109/TGRS.2018.2817393.

[75] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 978–989, Mar. 2018, doi: 10.1109/JSTARS.2018.2794888.

[76] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018, doi: 10.1109/JSTARS.2018.2805923.

[77] W. Yao, Z. Zeng, C. Lian, and H. Tang, "Pixel-wise regression using U-Net and its application on pansharpening," *Neurocomputing*, vol. 312, pp. 364–371, Oct. 2018, doi: 10.1016/j.neucom.2018.05.103.

[78] X. Liu, Y. Wang, and Q. Liu, "PSGAN: A generative adversarial network for remote sensing image pan-sharpening," in *Proc. 25th IEEE Int. Conf. Image Process.*, Sep. 2018, pp. 873–877, doi: 10.1109/ICIP.2018.8451049.

[79] Y. Zhang, C. Liu, M. Sun, and Y. Ou, "Pan-sharpening using an efficient bidirectional pyramid network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, pp. 1–15, Aug. 2019, doi: 10.1109/TGRS.2019.2900419.

[80] K. Li, W. Xie, Q. Du, and Y. Li, "DDLPS: Detail-based deep Laplacian pansharpening for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8011–8025, Oct. 2019, doi: 10.1109/TGRS.2019.2917759.

[81] H. Zhang and J. Ma, "GTP-PNet: A residual learning network based on gradient transformation prior for pansharpening," *ISPRS J. Photogrammetry Remote Sens.*, vol. 172, pp. 223–239, Feb. 2021, doi: 10.1016/j.isprsjprs.2020.12.014.

[82] J. Liu, Y. Feng, C. Zhou, and C. Zhang, "PWNet: An adaptive weight network for the fusion of panchromatic and multispectral images," *Remote Sens.*, vol. 12, no. 17, p. 2804, Aug. 2020, doi: 10.3390/rs12172804.

[83] J. Liu, C. Zhou, R. Fei, C. Zhang, and J. Zhang, "Pansharpening via neighbor embedding of spatial details," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4028–4042, Mar. 2021, doi: 10.1109/JSTARS.2021.3067877.

[84] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3D full convolutional neural network," *Remote Sens.*, vol. 9, no. 11, pp. 1139:1–1139:22, Nov. 2017, doi: 10.3390/rs9111139.

[85] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 146, pp. 305–319, Dec. 2018, doi: 10.1016/j.isprsjprs.2018.09.018.

[86] M. Gargiulo, A. Mazza, R. Gaetano, G. Ruello, and G. Scarpa, "Fast super-resolution of 20 m Sentinel-2 bands using convolutional neural networks," *Remote Sens.*, vol. 11, no. 22, pp. 2635:1–2635:18, Nov. 2019, doi: 10.3390/rs11222635.

[87] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse Dirichlet-Net for hyperspectral image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2511–2520, doi: 10.1109/CVPR.2018.00266.

[88] Q. Xie, M. Zhou, Q. Zhao, Z. Xu, and D. Meng, "MHF-Net: An interpretable deep network for multispectral and hyperspectral image fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1457–1473, 2020, doi: 10.1109/TPAMI.2020.3015691.

[89] J. F. Hu, T. Z. Huang, L. J. Deng, T. X. Jiang, G. Vivone, and J. Chanussot, "Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, 2021, doi: 10.1109/TNNLS.2021.3084682.

[90] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[91] L. He *et al.*, "Pansharpening via detail injection based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1188–1204, Apr. 2019, doi: 10.1109/JSTARS.2019.2898574.

[92] 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, doi: 10.1109/TGRS.2020.3031366.

[93] R. Dian, S. Li, and X. Kang, "Regularizing hyperspectral and multispectral image fusion by CNN denoiser," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1124–1135, Mar. 2021, doi: 10.1109/TNNLS.2020.2980398.

[94] H. Shen, M. Jiang, J. Li, Q. Yuan, Y. Wei, and L. Zhang, "Spatial–spectral fusion by combining deep learning and variational model," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6169–6181, 2019, doi: 10.1109/TGRS.2019.2904659.

[95] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1529–1543, 2020, doi: 10.1109/TNNLS.2019.2920857.

[96] Z. C. Wu, T. Z. Huang, L. J. Deng, J. F. Hu, and G. Vivone, "VO+Net: An adaptive approach using variational optimization and deep learning for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Mar. 2021, doi: 10.1109/TGRS.2021.3066425.

[97] Y. Feng, J. Liu, K. Chen, B. Wang, and Z. Zhao, "Optimization algorithm unfolding deep networks of detail injection model for pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, Art no. 5001305, doi: 10.1109/LGRS.2021.3077183.

[98] X. Shuang, J. Zhang, Z. Zhao, K. Sun, J. Liu, and C. Zhang, "Deep gradient projection networks for pan-sharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 1366–1375, doi: 10.1109/CVPR46437.2021.00142.

[99] X. Cao, X. Fu, D. Hong, Z. Xu, and D. Meng, "PanCSC-net: A model-driven deep unfolding method for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Oct. 2021, doi: 10.1109/TGRS.2021.3115501.

[100] H. Yin, "PSCSC-net: A deep coupled convolutional sparse coding network for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022, Art. no. 5402016, doi: 10.1109/TGRS.2021.3088313.

[101] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Inf. Fusion*, vol. 62, pp. 110–120, Oct. 2020, doi: 10.1016/j.inffus.2020.04.006.

[102] S. Luo, S. Zhou, Y. Feng, and J. Xie, "Pansharpening via unsupervised convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4295–4310, Jul. 2020, doi: 10.1109/JSTARS.2020.3008047.

[103] Y. Qu, R. Baghbaderani, H. Qi, and C. Kwan, "Unsupervised pansharpening based on self-attention mechanism," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3192–3208, 2021, doi: 10.1109/TGRS.2020.3009207.

[104] M. Ciotola, S. Vitale, A. Mazza, G. Poggi, and G. Scarpa, "Pansharpening by convolutional neural networks in the full resolution framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022, Art no. 5408717, doi: 10.1109/TGRS.2022.3163887.

[105] I. J. Goodfellow *et al.*, "Generative adversarial networks," 2014, *arXiv:1406.2661*.

[106] Z. Shao, Z. Lu, M. Ran, L. Fang, J. Zhou, and Y. Zhang, "Residual encoder–decoder conditional generative adversarial network for pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 9, pp. 1573–1577, 2020, doi: 10.1109/LGRS.2019.2949745.

[107] W. Dong, S. Hou, S. Xiao, J. Qu, Q. Du, and Y. Li, "Generative dual-adversarial network with spectral fidelity and spatial enhancement for hyperspectral pansharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, 2021, doi: 10.1109/TNNLS.2021.3084745.

[108] Z. Zhao *et al.*, "FGF-GAN: A lightweight generative adversarial network for pansharpening via fast guided filter," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2021, pp. 1–6, doi: 10.1109/ICME51207.2021.9428272.

[109] W. Xie, Y. Cui, Y. Li, J. Lei, Q. Du, and J. Li, "HPGAN: Hyperspectral pansharpening using 3-D generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 463–477, 2021, doi: 10.1109/TGRS.2020.2994238.

[110] A. Gastineau, J.-F. Aujol, Y. Berthoumieu, and C. Germain, "Generative adversarial network for pansharpening with spectral and spatial discriminators," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022, Art no. 4401611, doi: 10.1109/TGRS.2021.3060958.

[111] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008, doi: 10.1109/TGRS.2007.912448.

[112] T. M. Tu, S. C. Su, H. C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001, doi: 10.1016/S1566-2535(01)00036-7.

[113] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images-II. Channel ratio and "Chromaticity" transform techniques," *Remote Sens. Environ.*, vol. 22, no. 3, pp. 343–365, Aug. 1987, doi: 10.1016/0034-4257(87)90088-5.

[114] 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, May 2014, doi: 10.1109/LGRS.2013.2281996.

[115] G. Vivone, R. Restaino, and J. Chanussot, "A regression-based high-pass modulation pansharpening approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 984–996, Feb. 2018, doi: 10.1109/TGRS.2017.2757508.

[116] G. Vivone, R. Restaino, and J. Chanussot, "Full scale regression-based injection coefficients for panchromatic sharpening," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3418–3431, Jul. 2018, doi: 10.1109/TIP.2018.2819501.

[117] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[118] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002, doi: 10.1109/TGRS.2002.803623.

[119] T.-W. Hui, C. C. Loy, and X. Tang, "Depth map super-resolution by deep multi-scale guidance," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 353–369, doi: 10.1007/978-3-319-46487-9_22.

[120] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654, doi: 10.1109/CVPR.2016.182.

[121] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetric Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, Jun. 1997.

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

[123] L. Wald, *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*. Paris, France: Les Presses de l'École des Mines, 2002.

[124] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 6, no. 4, pp. 662–665, Oct. 2009, doi: 10.1109/LGRS.2009.2022650.

[125] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogrammetric Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008, doi: 10.14358/PERS.74.2.193.

[126] M. M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3880–3891, Nov. 2009, doi: 10.1109/TGRS.2009.2029094.

GRS