*remote sensing*

MDPI

*Article*

# Pansharpening Techniques: Optimizing the Loss Function for Convolutional Neural Networks

**Rocco Restaino**

Department of Information Engineering, Electrical Engineering and Applied Mathematics, University of Salerno, 84084 Fisciano, Italy; restaino@unisa.it; Tel.: +39-089-96-4350

**Abstract:** Pansharpening is a traditional image fusion problem where the reference image (or ground truth) is not accessible. Machine-learning-based algorithms designed for this task require an extensive optimization phase of network parameters, which must be performed using unsupervised learning techniques. The learning phase can either rely on a companion problem where ground truth is available, such as by reproducing the task at a lower scale or using a pretext task, or it can use a reference-free cost function. This study focuses on the latter approach, where performance depends not only on the accuracy of the quality measure but also on the mathematical properties of these measures, which may introduce challenges related to computational complexity and optimization. The evaluation of the most recognized no-reference image quality measures led to the proposal of a novel criterion, the Regression-based QNR (RQNR), which has not been previously used. To mitigate computational challenges, an approximate version of the relevant indices was employed, simplifying the optimization of the cost functions. The effectiveness of the proposed cost functions was validated through the reduced-resolution assessment protocol applied to a public dataset (PairMax) containing images of diverse regions of the Earth's surface.

**Keywords:** image fusion; multispectral images; panchromatic images; convolutional neural networks; unsupervised learning; loss functions

check for updates

Academic Editor: Silvia Liberata Ullo

Received: 19 September 2024
Revised: 20 December 2024
Accepted: 21 December 2024
Published: 25 December 2024

**Citation:** Restaino, R. Pansharpening Techniques: Optimizing the Loss Function for Convolutional Neural Networks. *Remote Sens.* **2025**, *17*, 16. https://doi.org/10.3390/rs17010016

**Copyright:** © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

The production of high-resolution synthetic images is a major focus in remote sensing image processing. It serves as a critical preprocessing step for the subsequent extraction of information, whether through human interpretation or numerical algorithms. These techniques generate detailed images by leveraging image redundancies, as in super-resolution [1], or by combining complementary data from multiple images of the same area, as in data fusion [2,3]. This enables the creation of images with higher detail than what state-of-the-art sensors can directly capture.

One of the most widely studied data fusion techniques is pansharpening. Historically, it aims to merge a low-resolution (LR) multispectral (MS) image with a high-resolution (HR) panchromatic (PAN) image to create an HR version of the MS image [4–6]. The good results achieved in pansharpening can be attributed to the fact that many satellites carry both types of sensors, allowing near-simultaneous, co-located acquisition of surface images. This enables applications ranging from enhanced visualization in virtual globe software, like Google Earth and Bing Maps, to improved scene classification [7] and change detection [8].

Numerous pansharpening solutions have been proposed, with the most prominent methodologies evolving rapidly alongside advances in signal processing [5,9]. The objective

is effectively described by Wald's protocol, which provides a guiding framework for developing pansharpening solutions [10]. Given a multispectral sensor, the goal is to generate an image that mimics what the same sensor would capture at a higher resolution. To achieve this, a panchromatic image of the same area, acquired at the target resolution, is utilized to extract the missing high-resolution details.

Early developments were marked by the competition between spatial domain methods and spectral domain techniques [11]. Spatial domain methods, often called Multiresolution Analysis (MRA) techniques, rely on the multiscale decomposition of images to be fused [12]. These can be achieved through linear systems like Gaussian filters [13], wavelets [14,15], curvelets [16], and contourlets [17], or nonlinear schemes [18]. Spectral domain methods, typically denoted as Component Substitution (CS) techniques, on the other hand, perform image fusion in a transformed domain, where the spatial component is enhanced using the PAN image. Examples include Principal Component Analysis (PCA) [19], Gram–Schmidt orthogonalization [20], Brovey transform (BT) [21], and the band-dependent spatial-detail-based (BDSD) approach [22,23].

Classical methods remain relevant due to their simplicity and adaptability to various image types. Their continued use is further supported by the high performance achieved through ongoing refinements. In particular, detail injection techniques in both CS and MRA methods have seen significant improvements. For example, the optimization of high-pass modulation (HPM) [24] has resulted in state-of-the-art performance [25,26], while regression-based projective methods, like the Gram–Schmidt adaptive (GSA) algorithm [27] and the MTF-based generalized Laplacian pyramid with context-based decision (MTF-GLP-CBD) [28], have advanced the field. Physical considerations have also contributed, such as the introduction of haze correction in the Brovey transform (BT-H) and MTF-GLP with HPM injection model (MTF-GLP-HPM-H) [29], as well as the BDSD with physical constraints (BDSD-PC) [30]. Contextual adaptive techniques, often indicated by the prefix "C", such as C-BDSD [23], C-GSA, and C-MTF-GLP-CBD [31], also compete with the most innovative pansharpening methods, including variational optimization (VO) and machine learning (ML) approaches [9].

Before ML techniques gained popularity, VO methods attracted significant attention. They delivered high performance but required dataset-specific parameter tuning and complex hyperparameter estimation [9]. Despite these obstacles, some VO approaches, such as semiblind deconvolution-based filter estimation [32], total variation [33], and low-rank representation [34,35], have achieved success. Sparse representation methods have achieved noteworthy results, beginning with seminal works where this representation was applied to images [36,37]. These methods have evolved to include techniques focused on details (SR-D) [38,39] and approaches utilizing dictionaries optimized for the spatial and spectral information of the images being combined [40].

Machine learning techniques, the main focus of this study, have revolutionized pansharpening. Two pioneering methods are the autoencoder-based approach [41] and the pansharpening neural network (PNN) based on a convolutional neural network (CNN) [42]. These approaches remain the foundation of many current implementations. A key innovation in recent years has been the introduction of residual learning, which accelerates and stabilizes the parameter learning process [43,44]. Deeper networks have been proposed to further enhance performance [43], though an increasing number of parameters can complicate the learning process [45]. Attention mechanisms [46], later integrated into more sophisticated transformer-based networks [47], offer a promising approach to improving performance without excessive parameter growth by focusing on the most relevant regions in the image. Generative methods have also evolved in parallel. Pansharpening works well with generative techniques. It was first implemented using generative adversarial net-

works (GANs). In this approach, a generator creates the fused image, and a discriminator evaluates its quality [48–50]. Recently, diffusion models have emerged as an alternative, improving stability in the training process, which is often a challenge with GANs [51].

Current research emphasizes methods that rely solely on available data, eliminating the need for reference images. These techniques, often categorized as nonsupervised, can be further divided into various subgroups [52]. While adaptable networks for different image types show promise [53], the most effective techniques today are either fully unsupervised (USL) [54,55] or involve fine-tuning models through transfer learning (TL) [45]. The fields of weakly supervised learning (WSL) and semi-supervised learning (Semi-SL) are still underexplored but have begun to see practical applications. Semi-supervised learning (Semi-SL) often starts with supervised training at reduced resolution. This is followed by unsupervised fine-tuning [56,57]. Self-supervised learning (SSL), on the other hand, begins with a pretext task. It then fine-tunes the model using pansharpening-specific data [58,59].

The primary challenge in pansharpening remains the difficulty of developing ready-to-use solutions for unseen datasets. Although current research is actively seeking ways to enhance the generalizability of pansharpening methods across diverse images, the performance of ML-based algorithms, including shallow networks, often falls short in this regard [44,45]. The most effective strategy continues to be a brief, dataset-specific training phase, where pre-trained networks are fine-tuned using the target images through an effectively implementable technique. This paper contributes to the ongoing research on unsupervised methods by focusing on the optimization of network parameters using reference-free cost functions. It examines key cost functions derived from well-established distortion metrics that rely solely on the images to be fused. Most of them had never been used for this purpose before. Special emphasis is placed on the simplicity of implementation, pursued through appropriate approximations. This aims to ensure that the optimization process remains efficient, both in terms of speed and solution stability.

The article is organized as follows. Section 2 outlines the motivations behind this study and highlights its contributions to addressing gaps in the existing literature. Section 3 formalizes the pansharpening problem, describing the architectures employed in this work and the cost functions investigated. Section 4 details the experiments conducted, including descriptions of the datasets, quality metrics, algorithms used, and the configuration of the networks being compared. The experimental results are presented in Section 5 and discussed in Section 6. Finally, Section 7 summarizes the key findings and conclusions drawn from this study.

## 2. Motivation and Contribution of the Work

This paper evaluates the practical performance of several deep learning algorithms within a framework suitable for real-world applications and highly reproducible settings. This section presents the addressed problem, experimental techniques employed, and key findings from the study.

### *2.1. Specificity of Pansharpening*

Pansharpening is a specialized form of data fusion that presents unique challenges, making it difficult to apply techniques developed for other domains directly. Beyond the complexities of data fusion, pansharpening requires the integration of two images with different spatial resolutions.

A significant challenge in remote sensing data fusion is the absence of reference images, complicating the development and evaluation of algorithms. For tasks involving resolution enhancement, finding an image at the target resolution can be difficult. In some cases, reference images can be obtained by adjusting the distance between the imaging device

and the scene or using higher-resolution devices. However, in satellite and aerial remote sensing, acquiring images with a resolution higher than the sensor's capacity is practically impossible [2,4,60].

This absence of reference images has long hindered the evaluation of pansharpening algorithms, leading to the development of specific methodologies [10]. Two commonly used protocols are the reduced-resolution (RR) protocol and the full-resolution (FR) protocol [5]. The RR protocol uses the available MS image as a reference, and the algorithms are tested on degraded versions of the images, with results compared against the reference. While this protocol allows for robust evaluation due to the availability of a reference, it assumes scale invariance as it evaluates the techniques at a lower resolution. The validity of this assumption cannot be guaranteed in pansharpening, as the high resolution of the images reveals details that are only visible at full scale. These details may impair the performance of the algorithms in practical applications. Conversely, the FR protocol tests algorithms on full-resolution images without any reference, resulting in less consistent quality metrics [60].

Before the rise of ML-based methods, the lack of reference images had a limited impact on tuning pansharpening algorithms. Classical approaches were often derived from general-purpose image processing techniques adapted to the problem [5]. However, for ML-based algorithms, the learning phase is crucial, and the absence of reference images complicates network training [45]. As with performance evaluation, ML algorithms must either be trained at a reduced scale or use reference-free cost functions.

Few studies have directly addressed this problem. Most research focuses on algorithms trained and tested on images of the same resolution, often using supervised learning where images are split into training and test sets from the same dataset. However, these approaches neglect the issue of resolution differences between the images to be fused and the available labels.

### *2.2. The Choice of the Evaluation Framework*

The main goal of this research is to evaluate basic ML-based techniques by comparing them to classical algorithms. We use a well-established benchmark dataset and reliable evaluation methods. This study focuses on cost functions that do not require reference images, making them suitable for real-world applications. We examine common feed-forward networks and analyze the results of fine-tuning them using various feasible approaches.

One of the key decisions in this study is the choice of evaluation protocol. We used the RR protocol, where the available MS image serves as a ground truth reference, and the fusion techniques are applied to degraded versions of the original MS and PAN images. This approach has advantages, such as allowing the use of robust metrics to assess spatial and spectral quality. It is based on the premise that the degradation introduced by the acquisition system can be accurately approximated [61]. Additionally, since the study focuses on cost functions that do not require ground-truth data, using reference-based metrics during testing ensures independence from the cost functions used during training.

However, this approach also introduces limitations, including the assumption of scale invariance, which may not always hold, and potential biases favoring algorithms that incorporate the same degradation technique used in training. To confirm the findings based on numerical evaluation at reduced resolution, the studied approaches were also subjected to visual assessment at full scale. The decision to exclude full-scale image quality measures was made to avoid introducing evaluation bias in the specific problem under analysis, which involves selecting a reference-free index as the cost function for optimizing network parameters.

The selection of datasets is also crucial. We utilized all datasets provided by Maxar [62], which offer a comprehensive testbed for evaluating algorithm performance across diverse image types. This benchmark ensures that the techniques in this study are appropriately positioned for future comparisons with newly developed algorithms.

### *2.3. Related Studies*

The literature provides various contributions to training pansharpening networks using reference-free cost functions, ranging from simple networks to more advanced architectures.

The early works introducing learning techniques based on practically computable measures focused on the spatial and spectral distortions used in the Quality with No Reference (QNR) index [63]. In these works, modifications were often made to the functional form of the measures. For instance, in [64], the maximum value between spatial and spectral distortions was employed, while [65] used an additive combination rule with optimized weights. While the former study utilized a simple convolutional neural network (CNN), the latter developed a more complex network that incorporated a pyramid Transformer encoder for global feature extraction.

Several studies have proposed alternative distortion measures to evaluate the spectral and spatial quality of fused images. For example, ref. [54] used a combination of the mean square error (MSE) and the structural similarity (SSIM) index [66] between the original MS image and a Gaussian-filtered version of the fused MS image. In this method, the spatial index was calculated similarly by comparing the original PAN image with a linear combination of the MS bands, optimized to minimize MSE. A comparable approach, based on minimizing the mean absolute error (MAE) instead of MSE, was applied in [67], differing primarily in network architecture.

Building upon the A-PNN network [44], various loss functions combining spectral and spatial distortion measures have been introduced in [68–71]. In [68], the spectral cost function was defined as the absolute error between the original image and its degraded version for the fused image. The spatial cost function, on the other hand, was calculated using the local correlation coefficient between the fused image and the PAN image. Improvements in computing the spectral index were later introduced in [69], where a combination of the ERGAS index [72] and the spectral distortion index proposed by Khan et al. [73] was used. This approach also featured a more efficient implementation by selecting specific image tiles for processing, significantly reducing the number of iterations without compromising performance [70]. The critical choice of combination coefficients for spatial and spectral distortions was further refined in [71].

Recent techniques have also explored the use of attention mechanisms to improve focus on critical details in the PAN image without requiring ground truth. In [74], a spatial attention module based on high-frequency components of the PAN image was used to enhance spatial details during the fusion process, alongside a cost function that computes quality terms at both reduced and full resolutions. Similarly, in [75], attention mechanisms were applied to both the MS image (to improve encoding) and the PAN image (to estimate spatially varying detail extraction and injection coefficients).

In more sophisticated directions, unsupervised approaches have also been proposed. For instance, ref. [76] introduced a hybrid reference-free loss where relationships between the high-resolution MS image, low-resolution MS image, and available PAN image were learned through neural networks. This allowed for the evaluation of the fused image's consistency via a reference-free loss. A more advanced approach was presented in Meta-Pan [77], which began with supervised training at reduced resolution before progressing to a meta-learning phase [78] for parameter optimization, and concluded with an unsupervised phase tailored to more specific images. In another study [79], reference-free networks

were trained using metrics based on perceptual losses that leverage high-level features instead of relying solely on pixel values [80]. Likewise, ref. [81] proposed an unsupervised image fusion method using an encoder–decoder network, optimized without training data, to extract spatial details and semantic features from a guidance image.

Generative approaches using reference-free distortion measures have also been investigated for pansharpening. For instance, generative adversarial networks (GANs) have been applied in [49,82], while diffusion models were explored in [83].

### *2.4. Contributions*

In this study, three widely used pansharpening networks were trained using cost functions based on the most popular image quality measures, including many that had not previously been applied to this task (Hybrid Quality with No Reference (HQNR) [84], Filter-based Quality with No Reference (FQNR), and Regression-based (RQNR) [85]). Among these, RQNR emerged as a particularly effective alternative to traditional measures. It demonstrated superior performance not only in terms of the quality of the fused images but also in its ease of implementation. RQNR requires far fewer computational resources, making it a more efficient choice without compromising on performance.

Below, we summarize the major contributions:

- We conducted a comprehensive evaluation of three well-established convolutional networks across various datasets, assessing their performance with effectively implementable learning methods and benchmarking their full potential using ground-truth-based quality measures in a reduced resolution setting.
- By adopting widely accepted reference-free quality measures as cost functions, we optimized network performance and assessed their impact on pansharpened image quality. This process included the introduction of the RQNR criterion, efficiently implemented through a suitable approximation that simplifies optimization.
- Our findings show that networks trained with reference-free cost functions, such as RQNR, can compete with, and even surpass, traditional state-of-the-art pansharpening methods. This demonstrates their viability for real-world applications, even in the absence of high-resolution reference data.

## 3. Pansharpening Algorithms

In this section, we briefly formalize the pansharpening problem and introduce the networks selected for this study, along with the cost functions employed for optimizing the pansharpening process.

Throughout this paper, the following specific notations and conventions are adopted. Two-dimensional and three-dimensional arrays are denoted by bold uppercase letters, such as $\mathbf{X}$. An MS image is represented as a three-dimensional array, $\mathbf{X} = \{\mathbf{X}_b\}_{b=1,\ldots,B}$, where $B$ is the total number of spectral bands. Each band is indexed by $b = 1, \ldots, B$, with $\mathbf{X}_b$ referring to the $b$-th band of the image. A PAN image is expressed as a two-dimensional matrix, represented by $\mathbf{Y}$. The symbol $\odot$ is used to denote the element-wise multiplication of matrices.

### *3.1. Formalization of the Pansharpening Problem*

Given an MS image with $B$ bands, denoted as $\mathbf{M} \in \mathbb{R}^{R \times C \times B}$, acquired by a specific sensor, and a companion PAN image $\mathbf{P} \in \mathbb{R}^{rR \times rC}$, the objective of a pansharpening algorithm $\mathcal{F}$ is to produce an estimate $\widehat{\mathbf{M}} \in \mathbb{R}^{rR \times rC \times B}$ of the hypothetical high-resolution image $\overline{\mathbf{M}} \in \mathbb{R}^{rR \times rC \times B}$. This high-resolution image represents what the sensor would capture at a

resolution that is $r$ times higher, where $r$ denotes the resizing factor [10]. The relationship can be expressed as:

$$\widehat{\mathbf{M}} = \mathcal{F}[\mathbf{M}, \mathbf{P}]. \tag{1}$$

The operator $\mathcal{F}$ is of a very general nature, capable of encompassing linear and nonlinear operations. Classical approaches to pansharpening typically involve two main steps: the extraction of details from the PAN image and their injection into the MS image. These classical techniques can be summarized by the following relation [5]:

$$\widehat{\mathbf{M}}_b = \widetilde{\mathbf{M}}_b + \mathbf{G}_b \odot \left(\mathbf{P}_b - \mathbf{P}_b^L\right), \; b = 1, \ldots, B, \tag{2}$$

where $\widetilde{\mathbf{M}} \in \mathbb{R}^{rR \times rC \times B}$ is the upsampled MS image, scaled to the dimensions of the PAN image, preferably using a 23-tap polynomial filter [13], $\mathbf{P}^L$ is a low-pass-filtered version of $\mathbf{P}$, and $\mathbf{G}$ is the injection coefficients matrix. The two groups of classical algorithms differ based on the origin of the $\mathbf{P}^L$ image: in Component Substitution (CS) methods, $\mathbf{P}^L$ is derived from $\widetilde{\mathbf{M}}$, whereas in Multi-Resolution Analysis (MRA) methods, it is derived from $\mathbf{P}$. The relationship (2) is applied band by band, and the dependence of $\mathbf{G}$, $\mathbf{P}$, and $\mathbf{P}^L$ on the channel $b$ indicates that these quantities can vary across bands [5].

In variational optimization (VO) approaches, it is assumed that the relationships linking the ideal high-resolution multispectral (HRMS) image $\overline{\mathbf{M}}$ to the two available images $\mathbf{M}$ and $\mathbf{P}$ are known, or can be estimated. Let $\mathcal{H}_M$ denote the operator that maps $\overline{\mathbf{M}}$ to $\mathbf{M}$, and $\mathcal{H}_P$ denote the operator that maps $\overline{\mathbf{M}}$ to $\mathbf{P}$, so that $\mathbf{M} = \mathcal{H}_M[\overline{\mathbf{M}}]$ and $\mathbf{P} = \mathcal{H}_P[\overline{\mathbf{M}}]$. Typically, the variational approach can be formalized as follows:

$$\widehat{\mathbf{M}} = \arg\min_{\check{\mathbf{M}}} \left[ \left\| \mathbf{M} - \mathcal{H}_M[\check{\mathbf{M}}] \right\|_{S_1} + \left\| \mathbf{P} - \mathcal{H}_P[\check{\mathbf{M}}] \right\|_{S_2} + \mathcal{R}(\check{\mathbf{M}}) \right], \tag{3}$$

where $\|\cdot\|_{S_1}$ and $\|\cdot\|_{S_2}$ represent appropriate norms in the functional spaces $S_1$ and $S_2$, respectively, to which $\mathbf{M}$ and $\mathbf{P}$ belong. The term $\mathcal{R}(\cdot)$ is a regularization functional designed to stabilize the minimization problem, which might otherwise be ill-posed.

In ML-based approaches, the optimal form of the relation (1) is approximated by an operator, often implemented via a neural network, which depends on a vector of parameters $\boldsymbol{\theta}$. This can be expressed as:

$$\widehat{\mathbf{M}} = \mathcal{F}_{\boldsymbol{\theta}}[\mathbf{M}, \mathbf{P}]. \tag{4}$$

The optimal parameter set $\hat{\boldsymbol{\theta}}$ is obtained by minimizing a cost function $\mathcal{C}(\cdot)$, which measures the quality of the fused output using the image pairs $\left\{\mathbf{M}^{(i)}, \mathbf{P}^{(i)}\right\}_{i \in \mathcal{TS}}$ from the training set $\mathcal{TS}$. In unsupervised learning, the cost function takes into account at most the images to fuse $\mathbf{M}^{(i)}$, $\mathbf{P}^{(i)}$ and the fused product $\mathcal{F}_{\boldsymbol{\theta}}\left[\mathbf{M}^{(i)}, \mathbf{P}^{(i)}\right]$:

$$\widehat{\theta} = \arg\min_{\theta} \mathcal{C}\left( \left\{ \mathbf{M}^{(i)}, \mathbf{P}^{(i)}, \mathcal{F}_{\boldsymbol{\theta}}\left[\mathbf{M}^{(i)}, \mathbf{P}^{(i)}\right] \right\}_{i \in \mathcal{TS}} \right), \tag{5}$$

whereas in supervised learning, the desired output $\overline{\mathbf{M}}^{(i)}$ is also available, leading to the following optimization:

$$\widehat{\theta} = \arg\min_{\theta} \mathcal{C}\left( \left\{ \mathbf{M}^{(i)}, \mathbf{P}^{(i)}, \mathcal{F}_{\boldsymbol{\theta}}\left[\mathbf{M}^{(i)}, \mathbf{P}^{(i)}\right], \overline{\mathbf{M}}^{(i)} \right\}_{i \in \mathcal{TS}} \right). \tag{6}$$

### *3.2. CNN Networks for Pansharpening*

The aim of this study is to investigate the possibility of efficiently optimizing the performance of a pansharpening network using a single-image fine-tuning phase. For this purpose, three well-known and widely appreciated convolutional networks were

selected: A-PNN [44], PanNet [43], and Fusion Net [86]. These networks were chosen for their shallow architectures, which allow for adequate generalization [45], and their use of a residual learning scheme that requires a smaller training dataset. Various cost functions, corresponding to the main nonreferenced quality measures used today, have been implemented and tested to assess the achievable performance.

#### 3.2.1. Algorithms

The Advanced PNN (A-PNN) was introduced in [44] as a residual-based extension of the original Pansharpening Neural Network (PNN) [42]. PNN was one of the first CNNs designed for pansharpening. It is based on the Super-Resolution Convolutional Neural Network (SRCNN) [87] and consists of three layers. These layers are inspired by the sparse representation of signals and implement the three fundamental steps: representation of low-resolution image patches, nonlinear transformation into high-resolution patches, and reconstruction of the high-resolution image by combining high-resolution features. The network input is formed by stacking the PAN image and the upsampled MS image. The A-PNN, illustrated on the left in Figure 1, modifies this process by applying it to the residuals, which are the details that need to be added to the low-resolution MS image to produce a high-resolution version. Additionally, A-PNN is designed for fine-tuning, beginning with a pre-trained PNN model. The A-PNN optimizes its performance using an $L_1$ cost function, which differs from the $L_2$ cost function used in PNN. The $L_1$ cost function is more sensitive to small errors, providing a finer level of detail in the output.

PanNet [43], depicted in the center of Figure 1, is a deeper network than PNN and is based on the ResNet architecture [88]. According to the original implementation [89], we performed the upsampling of the MS image using a transposed convolution layer, which was found in preliminary tests to slightly outperform traditional 23-tap interpolators. The primary goal of the PanNet architecture is to optimize both spectral and spatial consistencies. Spectral consistency is achieved through the residual architecture, while spatial consistency is enhanced by using only the high-frequency components of the inputs. To accomplish this, the input is constructed by stacking high-pass versions of both the PAN and MS images, which are generated using a box filter.

FusionNet (here abbreviated as FusNet), one of the architectures proposed in [86], implements a machine-learning-based method that is derived from classical pansharpening approaches. The referenced study introduces three residual-type architectures, all based on the methodology represented by Equation (2). These architectures take as input the difference between the panchromatic (PAN) image and a low-resolution version of it. For CS-Net, the low-resolution version is generated by combining the channels of the upsampled MS image, while MRA-Net degrades the PAN image using a modulation transfer function (MTF)-shaped filter. FusNet, illustrated on the right in Figure 1, directly uses the difference between the PAN image (appropriately duplicated across channels) and the upsampled MS image as input. Unlike CS-Net and MRA-Net, which operate on monochromatic inputs, FusNet uses multichannel inputs, enabling the model to more effectively capture spectral relationships across different channels.

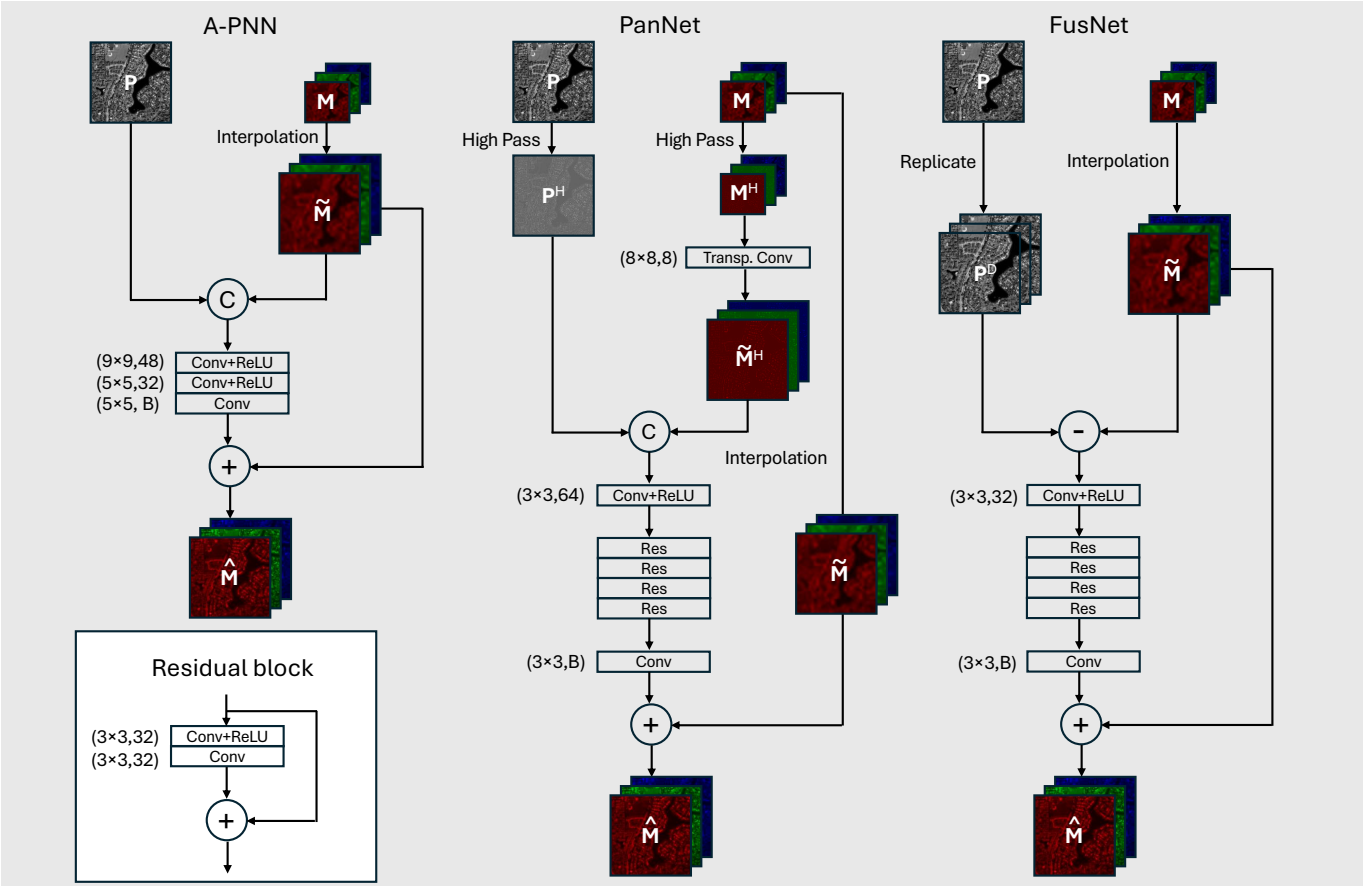



**Figure 1.** The three architectures analyzed in this work [89–91]. In the figure, the Conv and the Transp. Conv boxes represent a convolutional layer and a transposed convolutional layer, respectively, with their specifications provided alongside. The notation $(W \times H, C)$ denotes a kernel of dimensions $W \times H$ with $C$ filters. The activation function used is ReLU (rectified linear unit). The symbols Ⓒ, ⊕, and ⊖ indicate image concatenation, summation, and difference, respectively. For images, additional definitions are introduced: $\mathbf{P}^H$ and $\mathbf{M}^H$ represent the high-pass-filtered versions of the **P** and **M** images, while $\mathbf{P}^D$ is obtained by replicating the **P** image to match the number of channels in **M**. An exploded view of the residual block, labeled as Res in the architectures, is provided in the bottom left corner of the figure.

### 3.2.2. Loss Functions

In this study, we employed several widely used reference-free quality metrics, as reviewed in [60], to serve as cost functions for fine-tuning the networks. Before introducing these reference-free metrics, it is beneficial to describe two reference-based indices that will also be used in the performance evaluation phase, and which form the basis for many of the tested cost functions.

The Universal Image Quality Index (UIQI), or simply Q-index, was proposed to replace the Mean Square Error (MSE) as a quantitative measure of image quality, aiming to align more closely with human visual perception [92]. It is composed of three terms corresponding to the correlation coefficient, the luminance error, and the contrast error, with the following analytical expression:

$$Q(\mathbf{X}, \mathbf{Y}) = \frac{\sigma_{\mathbf{X},\mathbf{Y}}}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}} \frac{2\mu_{\mathbf{X}}\mu_{\mathbf{Y}}}{\mu_{\mathbf{X}}^2 + \mu_{\mathbf{Y}}^2} \frac{2\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}}{\sigma_{\mathbf{X}}^2 + \sigma_{\mathbf{Y}}^2} \tag{7}$$

where **X** and **Y** are two monochromatic images, with means $\mu_{\mathbf{X}}$ and $\mu_{\mathbf{Y}}$, variance $\sigma_{\mathbf{X}}^2$ and $\sigma_{\mathbf{Y}}^2$, and covariance $\sigma_{\mathbf{X},\mathbf{Y}}$.

If **X** and **Y** are multispectral images, the extension is typically achieved by averaging the Q-index across the corresponding bands:

$$\mathrm{Q}_{\mathrm{avg}}(\mathbf{X},\mathbf{Y}) = \frac{1}{B}\sum_{b=1}^{B} Q(\mathbf{X}_b,\mathbf{Y}_b). \tag{8}$$

A more comprehensive evaluation of image quality is given by the $\mathrm{Q2}^n$ index [93,94], which extends the Q-index using hypercomplex operators to replace the standard statistical operators. In this case, each pixel's values across multiple channels are treated as the real and complex parts of a hypercomplex number.

All reference-free quality indices consist of two components: one quantifying spectral distortion, $D_\lambda$, and another measuring spatial distortion, $D_S$. The PAN image and the MS image to be fused are used as references for these distortions, respectively. The underlying principle is to ensure that the spectral coherence of the fused image's channels matches that of the original MS image, while the spatial correlation between the PAN image and the fused image is preserved [10].

The first and most common reference-free index in the literature is the Quality with No Reference (QNR) metric [63]. The spectral distortion index for QNR is defined as:

$$D_\lambda^Q = \frac{1}{B(B-1)}\sum_{b=1}^{B}\sum_{\substack{k=1\\k\neq b}}^{B}\left|Q\left(\widehat{\mathbf{M}}_b,\widehat{\mathbf{M}}_k\right) - Q\left(\widetilde{\mathbf{M}}_b,\widetilde{\mathbf{M}}_k\right)\right|, \tag{9}$$

and the spatial distortion index is given by:

$$D_S^Q = \frac{1}{B}\sum_{b=1}^{B}\left|Q\left(\widehat{\mathbf{M}}_b,\mathbf{P}\right) - Q\left(\widetilde{\mathbf{M}}_b,\mathbf{P}^L\right)\right|, \tag{10}$$

where $\mathbf{P}^L$ is a low-pass-filtered version of the **P** image. The overall QNR is computed by combining these two indices multiplicatively:

$$\mathrm{QNR} = (1-D_\lambda^Q)^\alpha(1-D_S^Q)^\beta, \tag{11}$$

where $\alpha$ and $\beta$ are weighting parameters.

However, QNR evaluations often do not align well with visual analysis, and as a result, QNR is no longer widely used as a reference metric [9]. Theoretical analyses have shown that QNR tends to favor minimal detail injection into the fused image to avoid spectral distortions, with a strong coupling between the spatial and spectral indices [60].

The Filter-based QNR (FQNR) index, also known as the Khan index [73], separately analyzes the low-pass and high-pass components of the images for evaluating spectral and spatial distortions. The spectral distortion index implements the consistency property of the Wald protocol [10] and is defined as:

$$D_\lambda^F = 1 - Q2^n\left(\widehat{\mathbf{M}}^{L\downarrow},\mathbf{M}\right) \tag{12}$$

where $\widehat{\mathbf{M}}^{L\downarrow}$ is obtained by downsampling the low-pass-filtered version $\widehat{\mathbf{M}}^L$ of the fused image $\widehat{\mathbf{M}}$. The spatial distortion index compares the high-pass-filtered versions of the fused image and the original PAN, as also suggested by the Zhou protocol [95]:

$$D_S^F = \frac{1}{B}\sum_{b=1}^{B}\left|Q\left(\widehat{\mathbf{M}}_b^H,\mathbf{P}^H\right) - Q\left(\mathbf{M}_b^H,\mathbf{P}^{L\downarrow H}\right)\right|. \tag{13}$$

The high-pass-filtered versions are obtained by subtracting the low pass versions from the original image, namely:

$$\widehat{\mathbf{M}}_b^H = \widehat{\mathbf{M}}_b - \widehat{\mathbf{M}}_b^L; \qquad \mathbf{M}_b^H = \mathbf{M}_b - \mathbf{M}_b^L; \tag{14}$$

$$\mathbf{P}^H = \mathbf{P} - \mathbf{P}^L; \qquad \mathbf{P}^{L\downarrow H} = \mathbf{P}^{L\downarrow} - \mathbf{P}^{L\downarrow L}, \tag{15}$$

with $\mathbf{P}^{L\downarrow L}$ low-pass-filtered version of the $\mathbf{P}^{L\downarrow}$. The FQNR is then computed using the same multiplicative combination rule as QNR:

$$\text{FQNR} = (1 - D_\lambda^F)^\alpha (1 - D_S^F)^\beta. \tag{16}$$

The FQNR index has the merit of coming closer to the Wald protocol in the evaluation of the spectral quality. The definition is computationally less expensive than the $D_\lambda^Q$, since it requires the calculation of $B$ comparisons instead of $B^2$, so it can also be applied to images with a greater number of channels. The spatial distortion index, on the other hand, has the advantage of being more decoupled from the spatial distortion index, since it operates on a different part of the signal. The main problem is the dependence on the accuracy of the pass-through filter used.

The Hybrid QNR (HQNR) [84] combines the spectral index from FQNR with the spatial index from QNR, defined as:

$$\text{HQNR} = (1 - D_\lambda^F)^\alpha (1 - D_S^Q)^\beta, \tag{17}$$

inheriting advantages and disadvantages of the two measures.

Lastly, the Regression-based QNR (RQNR) [85] builds on FQNR by introducing a new spatial distortion index based on a linear regression model between the PAN image and the fused MS image:

$$\mathbf{P} = \sum_{b=1}^{B} w_i \widehat{\mathbf{M}}_b. \tag{18}$$

The optimal weights $\{\hat{w}_i\}_{i=1,\dots,B}$ are estimated by minimizing the mean square error (MMSE):

$$\hat{w}_i = \arg\min_{w_i} \left\| \mathbf{P} - \sum_{b=1}^{B} w_i \widehat{\mathbf{M}}_b \right\|_F^2 = \arg\min_{w_i} \|\boldsymbol{\epsilon}\|_F^2, \tag{19}$$

where the estimation error $\boldsymbol{\epsilon}$ is defined as:

$$\boldsymbol{\epsilon} = \mathbf{P} - \sum_{b=1}^{B} w_i \widehat{\mathbf{M}}_b \tag{20}$$

and $\|\cdot\|_F$ indicates the Frobenius norm. The spatial distortion index is then computed as the complement of the statistic $R^2$, which measures the goodness of the regression fit (18) [96]:

$$R^2 = 1 - \frac{\|\boldsymbol{\epsilon}\|_F^2}{\|\mathbf{P}\|_F^2} \tag{21}$$

and therefore, it is given by:

$$D_S^R = 1 - R^2 = \frac{\|\hat{\boldsymbol{\epsilon}}\|_F^2}{\|\mathbf{P}\|_F^2}, \tag{22}$$

where $\hat{\boldsymbol{\epsilon}}$ is the estimation error corresponding to the MMSE estimates $\{\hat{w}_i\}_{i=1,\dots,B}$. Finally, the RQNR is given by:

$$\text{RQNR} = (1 - D_\lambda^F)^\alpha (1 - D_S^R)^\beta. \tag{23}$$

While RQNR has the advantage of incorporating a more accurate spatial quality assessment, its primary challenge lies in estimating the linear regression coefficients and setting the weighting parameters $\alpha$ and $\beta$ due to the differing nature of the distortions [60].

## 4. Experimental Tests

Few papers have assessed the generalization capability of ML-based pansharpening algorithms, revealing that it tends to be limited [44,45], even for shallow networks. Consequently, this study focuses on fine-tuning pre-trained networks using the same images to be fused, specifically employing a reference-free cost function. This approach removes the need for a downsampling step, making it suitable for smaller images. It is especially effective for shallow networks with fewer parameters, allowing for the acquisition of statistically meaningful results.

In this section, the experimental phase is described below, detailing the dataset used, the image quality indexes employed for the assessment, the main operating conditions of the tests performed, and finally the performances obtained by the algorithms.

### *4.1. Testbed*

In this experimental study, the whole PairMax dataset, provided by Maxar Technologies for benchmarking purposes [97], was utilized (Figure 2).

The dataset consists of nine scenes acquired from four different satellites: Geoeye-1 (GE-1), Worldview-4 (WV-4), Worldview-2 (WV-2), and Worldview-3 (WV-3). The MS sensors mounted on board the first two satellites (GE-1 and WV-4) capture images with four channels, while WV-2 and WV-3 collect MS images with eight channels. This distinction is critical, as the number of bands has a significant impact on performance, particularly regarding spectral coherence. In the case of eight-band sensors, the overlap between the Relative Spectral Response (RSR) of the MS and PAN sensors is typically reduced, influencing the analysis results.

All sensors provide very high spatial resolution, as shown in Table 1, with the Ground Sample Distance (GSD) of the PAN images being less than 0.5 m, and the GSD of the MS images around 2 m. However, these sensors belong to two different generations: the earlier generation mounted on GE-1 and WV-2 satellites, and the more recent generation on WV-4 and WV-3. Despite WV-4 no longer being operational, its data were included due to the superior sensor technology it housed compared to GE-1's four-channel sensor. All sensors provide images with an 11-bit radiometric resolution, which is preserved in the dataset. The original MS and PAN images have size of $512 \times 512$ and $2048 \times 2048$, corresponding to a quite common resize factor $r = 4$.

In this study, the RR protocol was used to assess algorithm performance. The RR protocol, detailed in [62], retains the original MS image as the Ground Truth (GT) while the algorithms are applied to the degraded versions of the original MS and PAN images [9]. In line with the Wald protocol [10], the MS image is degraded using a filter whose frequency response matches the sensor's MTF, and the PAN image is processed using an ideal filter [13]. Both images are then downsampled by a factor equal to the resize factor $r = 4$ employed in this study.

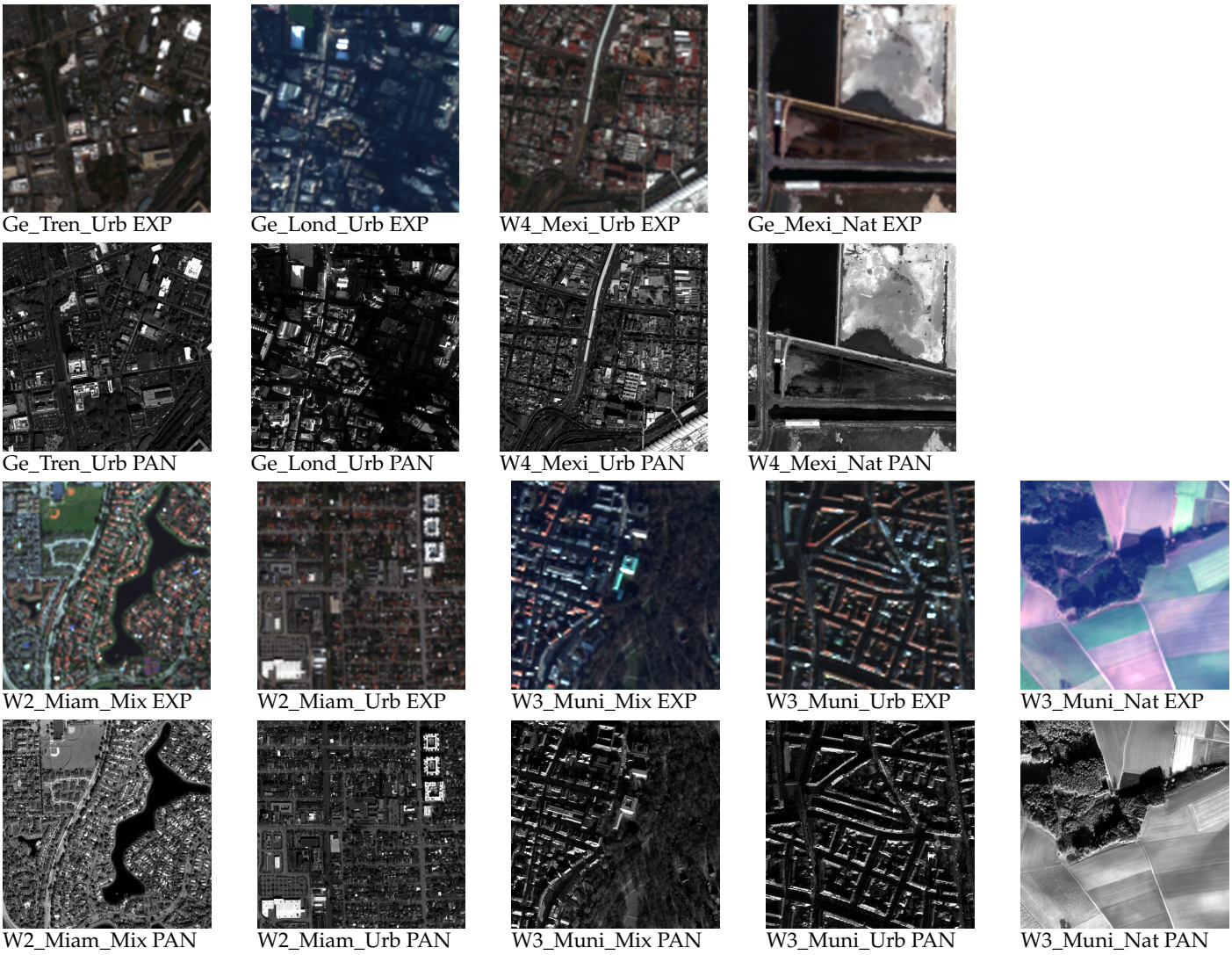



**Figure 2.** The PairMax dataset. The first two lines refer to images with four-channel MS data, the second two lines to images with eight-channel MS data.

**Table 1.** Main characteristics of the images composing the PairMax dataset.

| Satellite | Scene | GSD | Bands | Land Cover Type |
|---|---|---|---|---|
| GE-1 | `GE_Tren_Urb` | 0.46 m PAN, 1.84 m MS | 4 | Heterogeneous urban |
| GE-1 | `GE_Lond_Urb` | 0.46 m PAN, 1.84 m MS | 4 | Urban with long shadows |
| WV-4 | `W4_Mexi_Urb` | 0.31 m PAN, 1.24 m MS | 4 | Dense urban |
| WV-4 | `W4_Mexi_Nat` | 0.31 m PAN, 1.24 m MS | 4 | Vegetation and water |
| WV-2 | `W2_Miam_Mix` | 0.46 m PAN, 1.84 m MS | 8 | Urban with water |
| WV-2 | `W2_Miam_Urb` | 0.46 m PAN, 1.84 m MS | 8 | Urban |
| WV-3 | `W3_Muni_Mix` | 0.31 m PAN, 1.24 m MS | 8 | Urban with water |
| WV-3 | `W3_Muni_Urb` | 0.31 m PAN, 1.24 m MS | 8 | Urban and vegetated areas |
| WV-3 | `W3_Muni_Nat` | 0.31 m PAN, 1.24 m MS | 8 | Urban with water |

### *4.2. Image Quality Indexes*

To assess the quality of the pansharpened images, three widely used reference-based indices were employed: the Spectral Angle Mapper [98], the Erreur Relative Globale Adimensionnelle de Synthèse [72], and the $Q2^n$ index [93,94].

The Spectral Angle Mapper (SAM) [98] evaluates the spectral coherence between two images. It measures the angle between two spectral vectors, $\mathbf{v}_1 \in \mathbb{R}^B$ and $\mathbf{v}_2 \in \mathbb{R}^B$,

corresponding to the pixel values across the spectral bands. The SAM between two vectors is defined as:

$$\mathrm{SAM}(\mathbf{v}_1, \mathbf{v}_2) = \arccos\left(\frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_1\|_F \|\mathbf{v}_2\|_F}\right), \tag{24}$$

where $\langle \cdot, \cdot \rangle$ represents the scalar product of the vectors. The overall SAM for two images $\mathbf{X} \in \mathbb{R}^{rR \times rC \times B}$ and $\mathbf{Y} \in \mathbb{R}^{rR \times rC \times B}$ is the average of the SAM values computed for each pixel.

The ERGAS (Erreur Relative Globale Adimensionnelle de Synthèse) [72] is a radiometric index that normalizes the Mean Square Errors (MSEs) of each spectral channel relative to the average value of the channels. It is calculated as:

$$\mathrm{ERGAS}(\mathbf{X}, \mathbf{Y}) = \frac{100}{r} \sqrt{\frac{1}{B} \sum_{b=1}^{B} \frac{\mathrm{MSE}(\mathbf{X}_b, \mathbf{Y}_b)}{\mu_{\mathbf{X}}^2}}, \tag{25}$$

where $\mathrm{MSE}(\mathbf{X}_b, \mathbf{Y}_b)$ is the MSE between the corresponding bands $\mathbf{X}_b \in \mathbb{R}^{rR \times rC}$ and $\mathbf{Y}_b \in \mathbb{R}^{rR \times rC}$ of $\mathbf{X} \in \mathbb{R}^{rR \times rC \times B}$ and $\mathbf{Y} \in \mathbb{R}^{rR \times rC \times B}$, and $\mu_{\mathbf{X}}$ is the mean of $\mathbf{X}$.

Lastly, we employed the $Q2^n$ index, defined in Section 3.2.2, as a global measure that evaluates both the spatial and spectral quality of the images. This index is particularly useful for assessing the overall performance of pansharpening methods.

### *4.3. Compared Algorithms*

The algorithms selected for this evaluation include those under investigation, namely A-PNN, PanNet, and FusNet, each tested with various cost functions. Additionally, algorithms from key classes previously analyzed in earlier reviews [9,62] were incorporated for comparison. In particular, results for the same algorithms presented in the paper that introduced the PairMax dataset [62], have been reported here. It is also straightforward to obtain additional comparative values using the toolbox provided in [99], which accompanies the paper [9].

For CS approaches, the evaluation includes methods such as BDSD with physical constraints (BDSD-PC) [30], the Gram–Schmidt (GS) [20], and the Gram–Schmidt Adaptive (GSA) [27]. To represent the MRA class, three algorithms using the Generalized Laplacian Pyramid (GLP) with MTF-matched filters [13] were selected. These include the full-scale coefficient computation method (MTF-GLP-FS) [26], the HPM injection method with haze correction (MTF-GLP-HPM-H) [29], and a clustering-based implementation with projective injection (MTF-GLP-CBD) [31]. Additionally, the comparison considered a VO algorithm based on the sparse representation of details (SR-D) [38]. Three machine-learning-based algorithms were also evaluated. The A-PNN algorithm, fine-tuned for 50 epochs, represents the version available in the MATLAB toolbox (see Reference [99]) and is denoted as A-PNN-FT [44]. The bidirectional pansharpening network (BDPN) algorithm [100], one of the deep-learning-based methods included in the Python toolbox for recent deep learning algorithms [101], was fine-tuned at reduced resolution to prevent the use of images at the same resolution as those in the test set. The third method is the Z-PNN algorithm [68], which shares a methodological similarity with the approach explored in this study, as it employs a full-scale cost function without reference data. For completeness, values were calculated for MS image interpolation using a 23-tap polynomial kernel filter (EXP), providing a key baseline for comparison. The values associated with the reference method represent the ideal outcomes for the respective indices. These ideal values are achieved by using the reference image as the fused image in the case of reference-based indices. However, for indices without reference, the ideal values are rarely obtained.

The notation related to the CNN algorithms analyzed is intuitive and is composed first of the name of the network and then of the fine-tuning technique used. For example,

FusNet-HQNR indicates the FusNet network whose weights are optimized with the HQNR cost function. The performances related to the absence of fine-tuning (noFT) and to the use of the reference image during the training phase (GT) were also evaluated. The first technique corresponds to the value obtainable with the networks obtained by Transfer Learning, while the second technique represents a crucial value that allows evaluating the maximum performances obtainable with the method in question. These were obtained using 20,000 fine-tuning epochs, a value chosen to ensure an almost steady state value of the indices. In this case, it never happens that the performance decreases as the iterations increase since there is no risk of excessive specialization.

The other loss functions are obtained using the nonreference quality metrics described in Section 3.2.2 and are referred to by the same name in the acronyms. However, it is important to distinguish between metrics for evaluating image quality and those for optimizing network weights. The effectiveness of a network cost function is influenced by several factors that go beyond the accuracy of the quality metric. For example, the complexity of a metric affects both the calculation time—which can be a limiting factor, especially when fine-tuning is required for each image pair—and the difficulty of finding an optimal solution in parameter space.

Considering these factors, we opted to replace the $Q2^n$ index [93,94], which is used in the FQNR, HQNR, and RQNR quality metrics, with the average Q-index, defined in (8). The two indices show a very similar trend, as demonstrated by the scatterplot in Figure 3, which compares the Q and $Q2^n$ values calculated from the images produced by the classical algorithms used in this study, alongside the regression line between the two indices. Preliminary experiments confirmed that this substitution leads to comparable performance during network training, while also providing substantial computational savings. This modification is crucial for practical implementation and enhances the robustness of parameter estimation. The improved regularity of the error surface also enables more stable solutions, enhancing the network's ability to handle complex illuminated scenes.

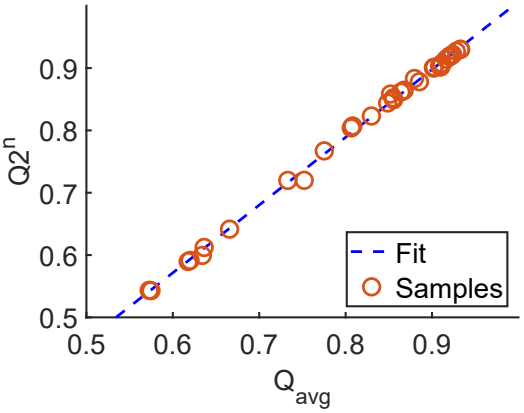



**Figure 3.** Scatterplot of $Q2^n$ vs. Q index. The regression coefficients are $-0.0809$ for the constant term and 1.0874 for the linear term, leading to a $R^2 = 0.9981$.

Regarding the QNR metric, the approach proposed by [64], which uses the combination rule $\text{QNR} = \min(D_\lambda^Q, D_S^Q)$, was evaluated in preliminary tests. However, it yielded worse results compared to the multiplicative rule described in (11).

All learning algorithms were fine-tuned for 2000 epochs, starting from nets trained from scratch on completely different datasets (those used in [9]).

The combination coefficients $\alpha = 1$ and $\beta = 0.1$ resulted in overall better performance, and thus, they were adopted for all algorithms. The overall pansharpening process is outlined in Algorithm 1.

**Algorithm 1:** Fine-tuning using reference-free metrics.

**Data:** The PAN and the MS images, **P** and **M**, respectively. Preliminary trained network $\mathcal{F}_{\boldsymbol{\theta}^{(0)}}$. Spatial distortion $D_S$. Spectral distortion $D_\lambda$. Weighting coeffcients $\alpha$ and $\beta$. Number of training epochs $K$.

**Result:** Pansharpened image, $\widehat{\mathbf{M}}$.

**begin**

  **for** $k = 1, \ldots, K$ **do**

    - Fuse **P** and **M** through $\mathcal{F}_{\boldsymbol{\theta}^{(k-1)}}$ to get a guess image $\widehat{\mathbf{M}}$

    - Update $\widehat{\theta}^{(k)} \leftarrow \widehat{\theta}^{(k-1)}$ backpropagating the loss $\mathcal{L}\left(\mathbf{M}, \mathbf{P}, \widehat{\mathbf{M}}\right) = 1 - (1 - D_\lambda)^\alpha (1 - D_S)^\beta$

  **end**

  - $\widehat{\mathbf{M}} = \mathcal{F}_{\boldsymbol{\theta}^{(K)}}[\mathbf{M}, \mathbf{P}]$

**end**

## 5. Experimental Results

The quality indices of the fused images obtained using the data collected by the various satellites are reported in Tables 2–5.

**Table 2.** Performance indexes computed for the two Geoeye-1 datasets. The best overall results are shown in boldface. The best results for each class of network algorithms are in blue. Results based on GT usage are in red.

| Method | Cost | Ge_Tren_Urb | | | Ge_Lond_Urb | | |
|---|---|---|---|---|---|---|---|
| | | $Q2^n$ | SAM [°] | ERGAS | $Q2^n$ | SAM [°] | ERGAS |
| Reference | | **1.0000** | **0.0000** | **0.0000** | **1.0000** | **0.0000** | **0.0000** |
| EXP | | 0.5828 | 7.3837 | 10.2034 | 0.5924 | 4.0029 | 10.8015 |
| BDSD-PC | | 0.9054 | 6.7506 | 5.1234 | 0.9191 | 2.8032 | 5.1443 |
| GS | | 0.8419 | 7.0974 | 6.6726 | 0.8014 | 3.9220 | 7.1635 |
| GSA | | 0.8988 | 6.8567 | 5.2663 | 0.9142 | 2.7727 | 5.2407 |
| MTF-GLP-FS | | 0.9033 | 6.8017 | 5.1500 | **0.9200** | 2.7727 | 5.2407 |
| MTF-GLP-HPM-H | | 0.8966 | 5.6987 | 5.0639 | 0.9189 | 3.0366 | 5.0510 |
| C-MTF-GLP-CBD | | 0.9026 | 6.5620 | 5.1373 | 0.9144 | 3.7263 | 7.6096 |
| SR-D | | 0.8915 | 6.0835 | 5.3648 | 0.8987 | 2.9751 | 5.5659 |
| A-PNN-FT | | 0.8859 | 4.8576 | 5.4296 | 0.8850 | **2.7801** | 5.9491 |
| Z-PNN | | 0.8961 | 6.1223 | **1.9377** | 0.9104 | 3.0097 | **2.0803** |
| BDPN | | 0.8655 | 6.0979 | 5.8951 | 0.8785 | 3.0806 | 6.2586 |
| A-PNN | GT | **0.9535** | **3.8445** | **3.4378** | **0.9548** | **2.1113** | **3.4200** |
| | noFT | 0.8886 | **4.8091** | 5.3980 | 0.8511 | 3.4854 | 6.8094 |
| | QNR | 0.8319 | 7.9953 | 6.7754 | 0.6855 | 16.9485 | 13.7052 |
| | FQNR | 0.8882 | 4.8213 | 5.4046 | 0.8511 | 3.4854 | 6.8148 |
| | HQNR | 0.8841 | 6.4077 | 5.6986 | 0.8959 | 3.9714 | 6.3242 |
| | RQNR | **0.9055** | 6.0904 | 5.0863 | 0.9160 | 3.0791 | 5.1363 |
| PanNet | GT | **0.9476** | **4.7969** | **3.5433** | **0.9530** | **2.3379** | **3.0896** |
| | noFT | 0.6843 | 9.2563 | 16.2686 | 0.7122 | 5.1803 | 17.6376 |
| | QNR | 0.8017 | 9.4059 | 7.5641 | 0.6508 | 22.2134 | 15.7088 |
| | FQNR | 0.7627 | 7.9346 | 9.3827 | 0.7636 | 4.6869 | 10.2782 |
| | HQNR | 0.8498 | 7.5531 | 6.5605 | 0.8720 | 4.3555 | 7.3408 |
| | RQNR | 0.8761 | 7.5013 | 6.0438 | 0.8857 | 3.7265 | 6.5994 |

**Table 2.** *Cont.*

| Method | Cost | Ge_Tren_Urb | | | Ge_Lond_Urb | | |
|---|---|---|---|---|---|---|---|
| | | $Q2^n$ | SAM [°] | ERGAS | $Q2^n$ | SAM [°] | ERGAS |
| FusNet | GT | 0.9583 | 3.8833 | 3.0637 | 0.9600 | 2.0936 | 2.4835 |
| | noFT | 0.7354 | 7.3138 | 8.4148 | 0.7758 | 5.2339 | 8.0505 |
| | QNR | 0.7514 | 7.1086 | 8.0709 | 0.8095 | 3.9659 | 7.5705 |
| | FQNR | 0.7553 | 6.7547 | 8.0589 | 0.7853 | 3.7334 | 7.7321 |
| | HQNR | 0.8127 | 7.1086 | 7.2014 | 0.8387 | 4.6421 | 7.5271 |
| | RQNR | 0.8451 | 6.4997 | 6.4915 | 0.8689 | 3.6976 | 6.3910 |

**Table 3.** Performance indexes computed for the two WV-4 datasets. The best overall results are shown in boldface. The best results for each class of network algorithms are in blue. Results based on GT usage are in red.

| Method | Cost | W4_Mexi_Urb | | | W4_Mexi_Nat | | |
|---|---|---|---|---|---|---|---|
| | | $Q2^n$ | SAM [°] | ERGAS | $Q2^n$ | SAM [°] | ERGAS |
| Reference | | 1.0000 | 0.0000 | 0.0000 | 1.0000 | 0.0000 | 0.0000 |
| EXP | | 0.5501 | 6.8951 | 8.2120 | 0.7274 | 2.1751 | 2.8821 |
| BDSD-PC | | **0.9237** | 5.6125 | 3.8415 | 0.8964 | 2.0565 | 1.4891 |
| GS | | 0.8547 | 6.4392 | 5.1652 | 0.8346 | 2.3298 | 1.7905 |
| GSA | | 0.9192 | 5.9993 | 3.9933 | 0.8977 | 2.0034 | 1.4845 |
| MTF-GLP-FS | | 0.9232 | 5.9122 | 3.9070 | **0.9039** | 1.9916 | 1.4669 |
| MTF-GLP-HPM-H | | 0.9091 | 5.2514 | 3.9181 | 0.8928 | 1.9729 | 1.5576 |
| C-MTF-GLP-CBD | | 0.9200 | 5.5694 | 3.8782 | 0.9006 | 1.9707 | 1.5010 |
| SR-D | | 0.9047 | 5.4267 | 4.1513 | 0.8895 | 1.9127 | 1.5768 |
| A-PNN-FT | | 0.8838 | **5.0580** | 4.5679 | 0.8820 | **1.8233** | 1.6688 |
| Z-PNN | | 0.9213 | 5.2778 | **1.4016** | 0.8921 | 1.9305 | **0.5685** |
| BDPN | | 0.8763 | 5.5259 | 4.7245 | 0.8572 | 2.2024 | 1.8446 |
| A-PNN | GT | 0.9578 | 3.8320 | 2.8365 | 0.9229 | 1.4784 | 1.0875 |
| | noFT | 0.8620 | 5.5830 | 5.0462 | 0.8772 | 1.9305 | 1.7649 |
| | QNR | 0.8880 | 7.2556 | 4.8764 | 0.7960 | 3.0015 | 2.5659 |
| | FQNR | 0.8620 | 5.5852 | 5.0466 | 0.8754 | 1.9527 | 1.7648 |
| | HQNR | 0.9071 | 6.5711 | 4.4800 | 0.8915 | 2.2020 | 1.6582 |
| | RQNR | 0.9187 | 5.9896 | 4.0801 | 0.8877 | 1.9744 | 1.4676 |
| PanNet | GT | 0.9532 | 4.4072 | 2.9277 | 0.9198 | 1.5861 | 1.0908 |
| | noFT | 0.8801 | 6.4540 | 6.6713 | 0.8510 | 2.3553 | 2.0538 |
| | QNR | 0.8431 | 8.3765 | 7.5864 | 0.8170 | 2.8177 | 2.2989 |
| | FQNR | 0.8739 | 7.1998 | 6.8432 | 0.8537 | 2.3298 | 2.0138 |
| | HQNR | 0.8914 | 6.0499 | 6.5202 | 0.8861 | 2.4871 | 1.7446 |
| | RQNR | 0.9113 | 5.3557 | 5.1472 | 0.8730 | 2.3397 | 1.7157 |
| FusNet | GT | 0.9627 | 3.7815 | 2.5317 | 0.9257 | 1.2977 | 0.8683 |
| | noFT | 0.8152 | 6.5717 | 5.7308 | 0.8207 | 2.4736 | 2.0346 |
| | QNR | 0.8226 | 6.4546 | 5.6650 | 0.8382 | 2.3326 | 1.9801 |
| | FQNR | 0.8173 | 6.5130 | 5.6931 | 0.8352 | 2.2264 | 1.9715 |
| | HQNR | 0.8301 | 6.5467 | 5.5673 | 0.8448 | 2.2693 | 1.9891 |
| | RQNR | 0.8369 | 6.4443 | 5.4567 | 0.8700 | 2.1502 | 1.7537 |

**Table 4.** Performance indexes computed for the two WV-2 datasets. The best overall results are shown in boldface. The best results for each class of network algorithms are in blue. Results based on GT usage are in red.

| Method | Cost | W2_Miam_Mix | | | W2_Miam_Urb | | |
|---|---|---|---|---|---|---|---|
| | | $Q2^n$ | SAM [°] | ERGAS | $Q2^n$ | SAM [°] | ERGAS |
| Reference | | **1.0000** | **0.0000** | **0.0000** | **1.0000** | **0.0000** | **0.0000** |
| EXP | | 0.5372 | 10.1668 | 9.2066 | 0.5457 | 9.5731 | 9.6275 |
| BDSD-PC | | **0.8625** | 9.1420 | 5.4843 | **0.9041** | 8.9339 | 4.8079 |
| GS | | 0.7368 | 10.8601 | 7.0942 | 0.8089 | 9.3340 | 6.5674 |
| GSA | | 0.8425 | 8.9703 | 5.7451 | 0.8862 | 9.4191 | 5.1924 |
| MTF-GLP-FS | | 0.8417 | 8.9726 | 5.7485 | 0.8856 | 9.3134 | 5.1692 |
| MTF-GLP-HPM-H | | 0.8584 | 8.4659 | 5.4797 | 0.8946 | 8.5635 | 5.0204 |
| C-MTF-GLP-CBD | | 0.8390 | 10.3918 | 6.2157 | 0.8814 | 9.4683 | 5.3149 |
| SR-D | | 0.8229 | 9.1678 | 5.9910 | 0.8676 | 8.8079 | 5.5048 |
| A-PNN-FT | | 0.8509 | **7.5701** | 5.5486 | 0.8849 | **7.3383** | 5.1656 |
| Z-PNN | | 0.8490 | 7.9199 | **1.9725** | 0.8894 | 7.6715 | **1.7885** |
| BDPN | | 0.8273 | 8.7961 | 5.8042 | 0.8962 | 8.2598 | 5.4184 |
| A-PNN | GT | **0.9286** | **5.8308** | **3.6884** | **0.9537** | **6.0006** | **3.3662** |
| | noFT | 0.8385 | 7.6758 | 5.5320 | 0.8806 | 7.4818 | 5.0525 |
| | QNR | 0.7732 | 9.0526 | 6.7335 | 0.8473 | 11.0628 | 6.6230 |
| | FQNR | 0.8412 | 7.6690 | 5.5060 | 0.8823 | 7.4658 | 5.0299 |
| | HQNR | 0.8486 | 8.7778 | 5.7198 | 0.8911 | 8.9517 | 5.2766 |
| | RQNR | 0.8548 | 8.7239 | 5.4556 | 0.8940 | 9.3177 | 5.2476 |
| PanNet | GT | **0.9181** | **6.5842** | **3.9874** | **0.9453** | **6.8017** | **3.6538** |
| | noFT | 0.8223 | 9.7975 | 6.3167 | 0.8656 | 8.9616 | 5.5937 |
| | QNR | 0.7668 | 11.0501 | 7.0909 | 0.8674 | 9.2218 | 5.6654 |
| | FQNR | 0.8254 | 10.0940 | 6.3647 | 0.8675 | 8.9518 | 5.5963 |
| | HQNR | 0.8277 | 9.3128 | 6.2284 | 0.8822 | 9.1715 | 5.4902 |
| | RQNR | 0.8410 | 9.4068 | 5.7506 | 0.8885 | 9.3821 | 5.2766 |
| FusNet | GT | **0.9400** | **5.3525** | **3.1342** | **0.9589** | **5.8183** | **3.1385** |
| | noFT | 0.8254 | 9.6081 | 6.3144 | 0.8660 | 8.8994 | 5.6480 |
| | QNR | 0.8006 | 9.7934 | 6.4368 | 0.8656 | 9.1602 | 5.7859 |
| | FQNR | 0.8229 | 9.7144 | 6.3266 | 0.8646 | 8.8694 | 5.6783 |
| | HQNR | 0.7953 | 9.3522 | 6.5664 | 0.8675 | 8.9951 | 5.7043 |
| | RQNR | 0.8541 | 9.3238 | 5.6702 | 0.8953 | 8.8366 | 5.0717 |

**Table 5.** Performance indexes computed for the three WV-3 datasets. The best overall results are shown in boldface. The best results for each class of network algorithms are in blue. Results based on GT usage are in red.

| Method | Cost | W3_Muni_Mix | | | W3_Muni_Urb | | | W3_Muni_Nat | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $Q2^n$ | SAM [°] | ERGAS | $Q2^n$ | SAM [°] | ERGAS | $Q2^n$ | SAM [°] | ERGAS |
| Reference | | **1.0000** | **0.0000** | **0.0000** | **1.0000** | **0.0000** | **0.0000** | **1.0000** | **0.0000** | **0.0000** |
| EXP | | 0.6382 | 6.0261 | 10.8510 | 0.6164 | 7.7873 | 15.5875 | 0.5981 | 4.2291 | 3.5814 |
| BDSD-PC | | 0.9322 | 4.6638 | 4.1681 | 0.9241 | 5.7073 | 7.4363 | 0.8073 | 3.0237 | 1.9018 |
| GS | | 0.7889 | 6.1109 | 7.0181 | 0.8126 | 7.9825 | 10.6141 | 0.7124 | 4.0648 | 2.6983 |
| GSA | | 0.9283 | 4.5476 | 4.4244 | 0.9208 | 5.4690 | 7.4548 | 0.8044 | 3.2656 | 2.1663 |
| MTF-GLP-FS | | 0.9279 | 4.5296 | 4.4457 | 0.9200 | **5.4407** | 7.3512 | 0.8059 | 3.2340 | 2.1405 |
| MTF-GLP-HPM-H | | **0.9333** | **4.0962** | 4.1965 | **0.9251** | 5.6494 | 7.1778 | **0.8074** | **2.7901** | 1.7986 |

**Table 5.** *Cont.*

| Method | Cost | W3_Muni_Mix | | | W3_Muni_Urb | | | W3_Muni_Nat | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $Q2^n$ | SAM [°] | ERGAS | $Q2^n$ | SAM [°] | ERGAS | $Q2^n$ | SAM [°] | ERGAS |
| C-MTF-GLP-CBD | | 0.9192 | 5.2943 | 4.8486 | 0.9114 | 6.2052 | 8.1904 | 0.8045 | 3.1474 | 2.1637 |
| SR-D | | 0.9016 | 4.7611 | 5.3522 | 0.8913 | 6.6353 | 8.3888 | 0.7846 | 3.0034 | 1.9844 |
| A-PNN-FT | | 0.8784 | 5.0636 | 5.9190 | 0.8902 | 6.2602 | 8.4805 | 0.7113 | 3.7218 | 2.4970 |
| Z-PNN | | 0.8594 | 6.4484 | **2.4753** | 0.8750 | 8.5958 | **3.5083** | 0.6807 | 4.2008 | **1.0317** |
| BDPN | | 0.8897 | 4.7229 | 5.4619 | 0.8689 | 8.2834 | 5.4274 | 0.7499 | 3.2543 | 2.1845 |
| A-PNN | GT | **0.9513** | **3.4987** | **3.1907** | **0.9606** | **4.2449** | **4.9433** | **0.8313** | **2.5637** | **1.4837** |
| | noFT | 0.7598 | 7.1208 | 8.4470 | 0.7939 | 9.7334 | 11.7944 | 0.5432 | 4.7176 | 3.4971 |
| | QNR | 0.7863 | 13.9427 | 8.4460 | 0.8137 | 14.7175 | 11.9744 | 0.4166 | 4.6608 | 4.1198 |
| | FQNR | 0.7757 | 7.0468 | 8.2400 | 0.8080 | 9.7122 | 11.4733 | 0.5963 | 4.4609 | 3.2489 |
| | HQNR | 0.8890 | 6.3182 | 6.3089 | 0.8955 | 7.7941 | 8.7193 | 0.6468 | 4.3714 | 3.4458 |
| | RQNR | 0.9031 | 5.8645 | 5.5480 | 0.8949 | 7.6130 | 8.7823 | 0.7679 | 3.7693 | 2.4275 |
| PanNet | GT | **0.9473** | **3.8989** | **3.3052** | **0.9614** | **4.5326** | **4.4907** | **0.8287** | **2.5792** | **1.4726** |
| | noFT | 0.8801 | 6.4540 | 6.6713 | 0.8923 | 8.6096 | 9.6488 | 0.7088 | 4.7256 | 3.1256 |
| | QNR | 0.8431 | 8.3765 | 7.5864 | 0.8934 | 7.7413 | 8.9435 | 0.6514 | 4.6211 | 3.2579 |
| | FQNR | 0.8739 | 7.1998 | 6.8432 | 0.9002 | 8.8535 | 9.7061 | 0.2514 | 10.8686 | 6.6024 |
| | HQNR | 0.8914 | 6.0499 | 6.5202 | 0.8995 | 7.5776 | 9.0165 | 0.7215 | 4.4007 | 2.8563 |
| | RQNR | 0.9113 | 5.3557 | 5.1472 | 0.8995 | 6.8194 | 7.7547 | 0.7504 | 4.0763 | 2.6787 |
| FusNet | GT | **0.9572** | **3.3237** | **2.4885** | **0.9681** | **4.2445** | **3.7786** | **0.8485** | **2.1527** | **1.2122** |
| | noFT | 0.7811 | 7.6588 | 7.3704 | 0.8711 | 9.7675 | 10.3209 | 0.5963 | 5.6120 | 3.9887 |
| | QNR | 0.8295 | 8.9779 | 7.4941 | 0.8766 | 11.6005 | 10.2290 | 0.6054 | 6.0992 | 3.3482 |
| | FQNR | 0.8713 | 6.6717 | 6.7944 | 0.8880 | 8.6498 | 10.0403 | 0.4070 | 6.1437 | 3.9901 |
| | HQNR | 0.8563 | 6.6414 | 7.1423 | 0.8829 | 8.0526 | 9.8364 | 0.4162 | 6.2201 | 3.6997 |
| | RQNR | 0.9001 | 6.3063 | 5.9143 | 0.9063 | 7.9454 | 8.6807 | 0.4044 | 6.3404 | 3.7323 |

The first two tables refer to satellites acquiring four-channel multispectral imagery. Among traditional methods, BDSD-PC and MTF-GLP-FS consistently show strong performance across both datasets, providing a strong balance between spectral and spatial quality, as indicated by $Q2^n$, SAM, and ERGAS metrics. GSA and C-MTF-GLP-CBD also perform well, closely trailing the two leaders in terms of overall quality. Very low ERGAS values, indicating better radiometric accuracy, are reported for the MTF-GLP-HPM-H among the classical methods, and particularly for the Z-PNN among the ML-based methods.

More specific observations related to the problem examined in this paper highlight the need to fine-tune the networks. Only the A-PNN without tuning has decent performance, indicating a greater generalization capability. The results are reasonably competitive but clearly fall behind the other tuning methods, such as HQNR and RQNR. The PanNet and the FusNet show significant degradation in all metrics across datasets without fine-tuning. SAM and ERGAS values are considerably higher, demonstrating that they struggle to maintain spatial and spectral fidelity. Across A-PNN, PanNet, and FusNet, RQNR consistently provides the best results in terms of $Q2^n$, SAM, and ERGAS across multiple datasets. It shows the best balance between spatial and spectral quality and can outperform all other methods in the `Ge_Tren_Urb` dataset.

The datasets with eight-channel MS images give very similar results, with some small differences. As for the classical methods, which obtain better results than CNNs, the highest performance is also obtained in this case by BDSD-PC, whereas for MRA techniques, the highest ranking is obtained by MTF-GLP-HPM-H. Once again, the Z-PNN method achieves the best performance among ML-based algorithms; however, the results indicate that the algorithm is more inclined to optimize radiometric distortion. For both WV-2 and WV-3 datasets, the performance of the pansharpening algorithms shows a clear trend where the application of fine-tuning enhances the quality of fused images. The improvement is seen across all metrics, with significant gains in spatial and spectral quality for fine-tuned versions of the models. This indicates that tuning to the specific characteristics of each dataset is critical for improving pansharpening performance. A-PNN shows consistent improvements with tuning, especially in spectral accuracy, making it competitive across both WV2 and WV3 datasets. PanNet performs well, especially in urban and mixed regions, with fine-tuning leading to substantial gains in both SAM and ERGAS. FusNet, while benefiting from fine-tuning, remains less effective compared to A-PNN and PanNet, especially in the natural region, where its SAM and ERGAS values indicate higher errors.

The figures enable the visual inspection of products generated by various algorithms in two scenarios: an urban scene with eight-channel MS data (`W3_Muni_Mix`), and a natural scene with four-channel MS data (`W4_Mexi_Nat`). For the urban scene, close-ups of the final product, detailed views, and MS maps are presented in Figures 4–6. For the natural scene, close-ups of the final product and detailed views are shown in Figures 7 and 8. Results from all combinations of the analyzed networks and tested cost functions are reported, while only representative examples are provided for classical reference algorithms, particularly those significant for the quality of the obtained results. Specifically, two examples from the CS methods—GSA and BDSD-PC—and two from the MRA methods—MTF-GLP-FS and MTF-GLP-HPM-H—are presented alongside the SR-D technique, which is notably effective among the VO approaches.

It is evident that CS techniques often yield spatially appealing results, although they can sometimes lead to overinjection of details. MRA techniques, on the other hand, tend to provide more spectrally consistent outcomes.

In both scenes, the differences between the various cost functions are evident in the merged products displayed in Figures 4 and 7. Additionally, detailed comparisons in Figures 5 and 8 also show the details, i.e., the difference between the fused image and the original image upsampled to the scale of the fused image.

First, it is important to note the high quality of the merged images produced by the networks trained with the GT, indicating their ability to generate results close to the ideal. In contrast, the images produced by the networks before the tuning phase are of significantly lower quality.




**Figure 4.** Close-ups of the fusion results for the `W3_Muni_Mix_RR` dataset.

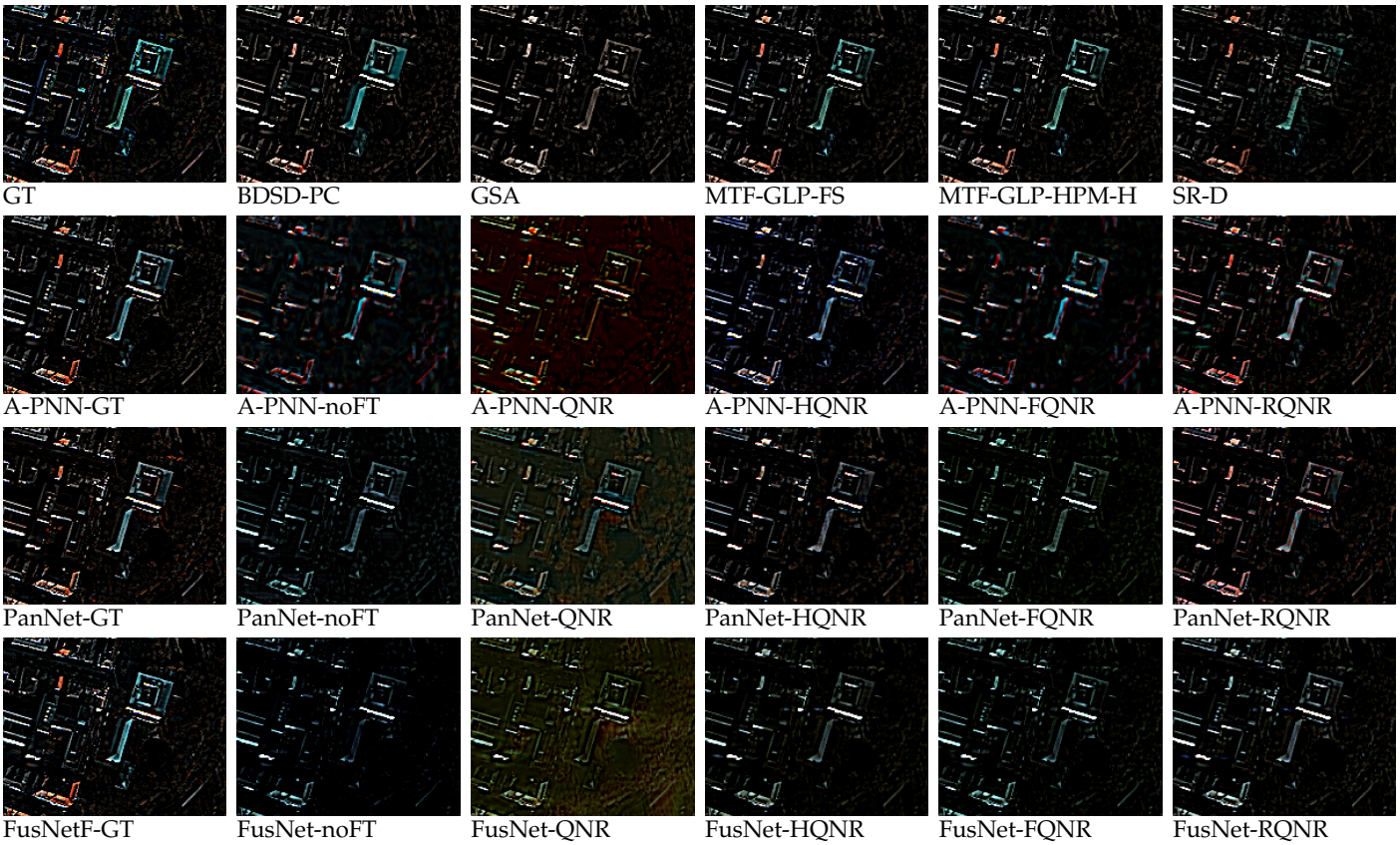



**Figure 5.** Close-ups of details of the fusion results for the `W3_Muni_Mix_RR` dataset.

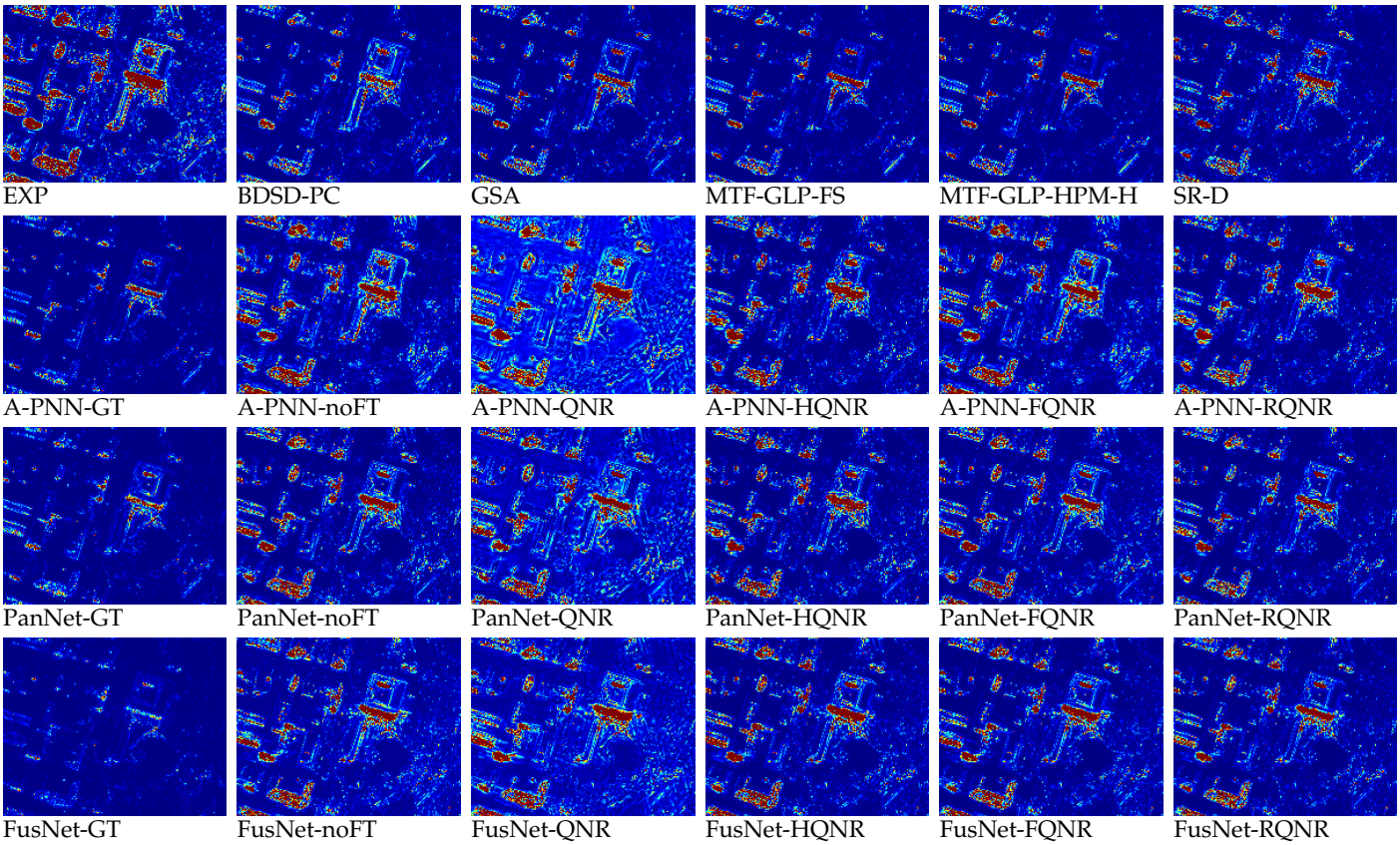



**Figure 6.** MSE maps related to the fusion results for the `W3_Muni_Mix_RR` dataset.

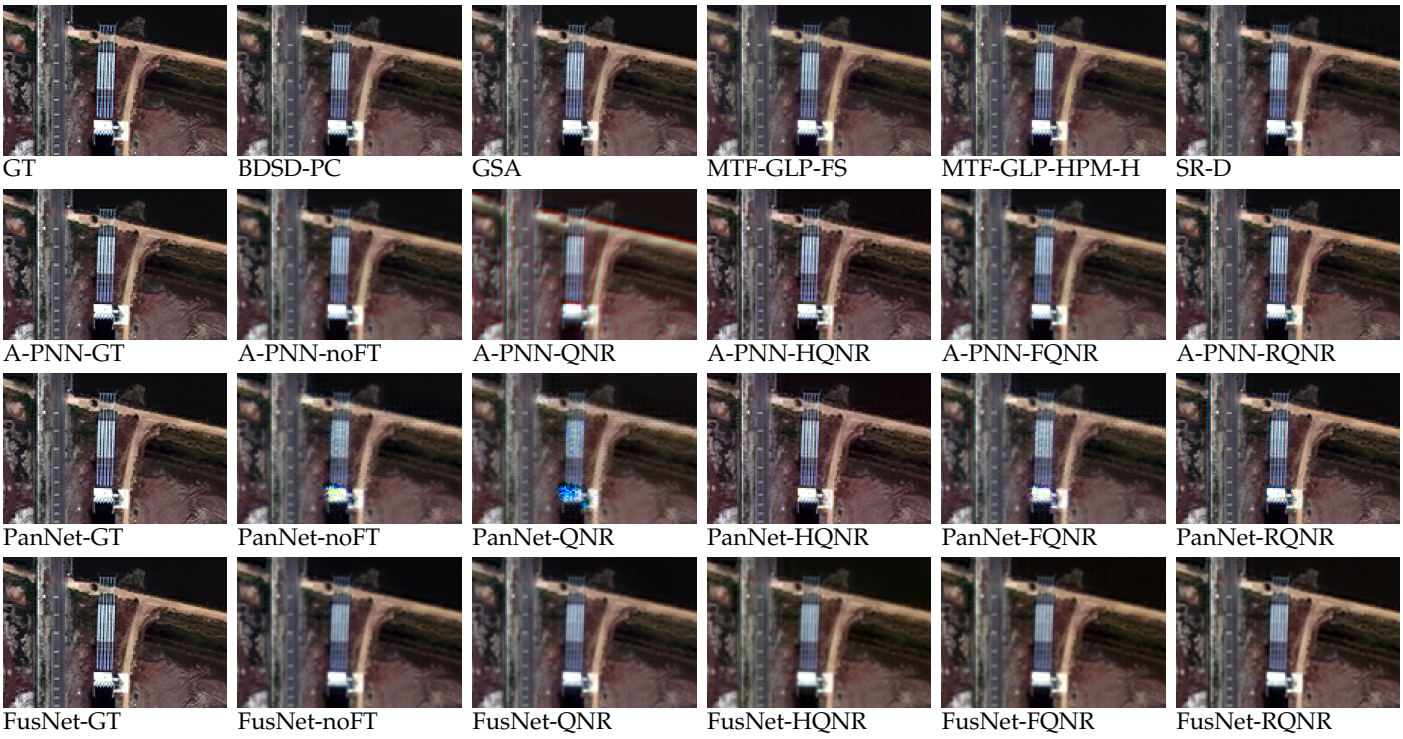



**Figure 7.** Close-ups of the fusion results for the `W4_Mexi_Nat_RR` dataset.

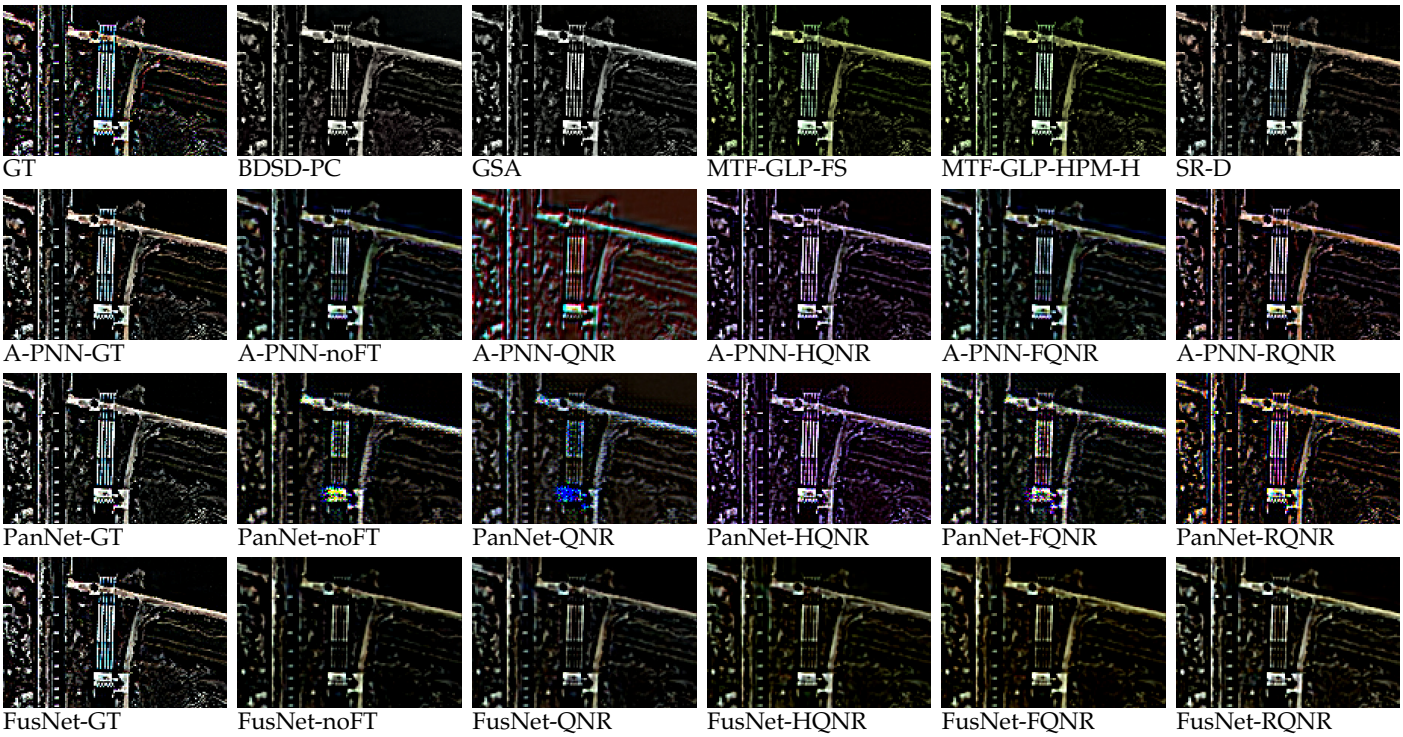



**Figure 8.** Close-ups of details of the fusion results for the `W4_Mexi_Nat_RR` dataset.

The cost function based on QNR does not improve results in the given case, and in fact, tends to introduce artifacts, as shown in in Figure 7 and produces spectrally wrong images. This behavior occurred consistently across the tests, as reflected in the results shown in Tables 2–5. In contrast, the other algorithms achieved significantly better outcomes, particularly after the tuning phase, enhancing performance compared to the unrefined cases. A more detailed examination reveals that the RQNR cost function makes a notable contribution in reproducing vegetation on the right side of Figure 4. This improvement is quantitatively confirmed by the reduction in MSE error in that area, as shown in Figure 6. Furthermore, Figure 6 demonstrates the improvement that RQNR achieves on building contours, which was previously only visible in the A-PNN-FQNR results in Figure 4, such as the edges of the central building.

Although the fused images in Figures 4 and 7 show several visible differences, the visual analysis is further aided by the images in Figures 5 and 8. These figures highlight the net contribution of pansharpening to the final product, particularly in terms of the quantity of injected details and their spectral properties. The RQNR cost function tends to promote higher levels of detail injection (which may not always be beneficial). Moreover, optimal details are observed to be spectrally accurate, with the PanNet network yielding the highest quality results, particularly in the `W3_Muni_Mix` scene.

Figure 9 provides an example of pansharpening algorithm application to images at the original scale. Numerical results are not presented here because selecting a specific cost function for full-scale evaluation would inherently favor one of the tested functions, introducing bias. Therefore, the assessment relies on visual analysis, which remains valuable for evaluating the robustness of the techniques across scales. However, relying on visual analysis of full-scale images, similar to [68], presents challenges due to difficulties in analyzing images with a high number of channels and with radiometric resolution exceeding 8 bits.

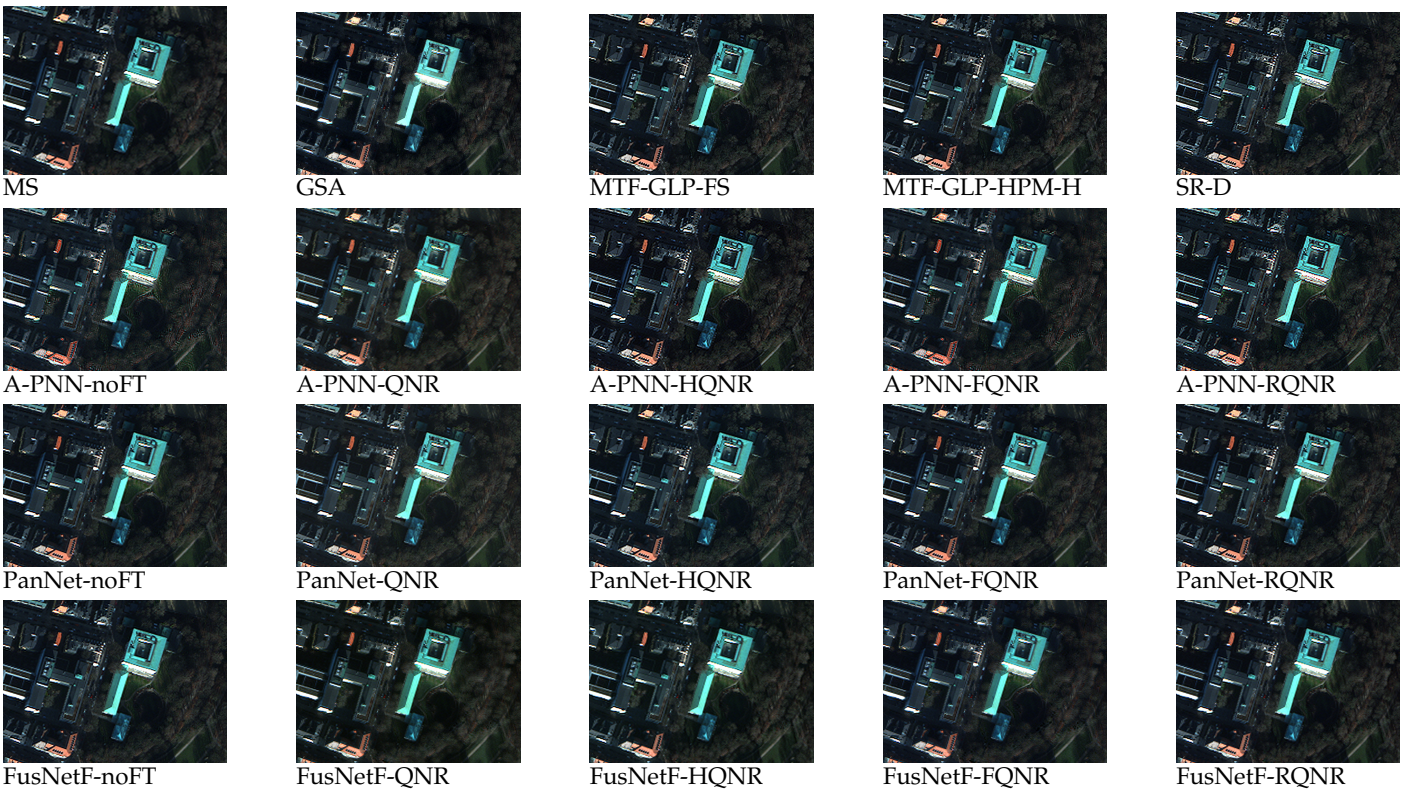



**Figure 9.** Close-ups of the fusion results for the `W3_Muni_Mix_FR` dataset.

A comparison between the images in Figure 9 and those in Figure 4 (depicting the same area) highlights the appearance of much smaller objects in the full-scale images. The improvements achieved by fine-tuning the networks over 1000 full-scale epochs are evident in the enhanced image quality, both spatially and spectrally, as observed through visual inspection. Among the tested cost functions, the one derived from the RQNR index consistently delivers the best performance across all three networks. In particular, the images produced by the PanNet network exhibit outstanding quality in both spatial and spectral dimensions.

The comparison between Figures 4 and 9 is particularly compelling, as the ranking of the tested algorithms remains virtually unchanged. This consistency supports the validity of the numerical results obtained at a reduced scale, even when the algorithms are applied at the actual scale.

An important evaluation metric in this study is the computational burden associated with different cost functions. Table 6 presents the training times for a single epoch, as measured using an Nvidia Titan XP GPU. The $L_1$ and $L_2$ cost functions, which are used for training with ground truth GT data, involve the least computational overhead. In contrast, reference-free cost functions require additional calculations, leading to increased computation times. Among these, the QNR function is the most computationally demanding, especially in the eight-channel case, due to the high complexity of calculating spectral distortion, which involves approximately $B^2$ comparisons between Q-indexes. The FQNR and HQNR cost functions exhibit similar computation times, while the RQNR index offers a small reduction in computational burden. This efficiency gain is primarily due to the simpler evaluation of spatial distortion, with the savings becoming more pronounced as the number of $B$ bands increases.

**Table 6.** Computational training times in seconds per epoch for the algorithms tested in this work. The best overall results are shown in boldface. The best results for each class of network algorithms are in blue. Results based on GT usage are in red.

| | W4_Mexi_Nat | | | W3_Muni_Mix | | |
|---|---|---|---|---|---|---|
| | **A-PNN** | **PanNet** | **FusNet** | **A-PNN** | **PanNet** | **FusNet** |
| GT | **0.1551** | **0.2362** | **1.3664** | **0.1285** | **0.2883** | **1.4061** |
| QNR | 0.4693 | 0.5001 | 1.6655 | 1.3259 | 1.3935 | 2.4722 |
| FQNR | 0.3317 | 0.4692 | 1.6503 | 0.5763 | 0.7093 | 1.9232 |
| HQNR | 0.3181 | 0.4532 | 1.6484 | 0.4963 | 0.6958 | 1.9099 |
| RQNR | **0.2761** | 0.4174 | 1.5803 | **0.3793** | 0.6045 | 1.8038 |

## 6. Discussion

The most significant evidence that emerged from this work is the great potential of ML-based algorithms, though effectively exploiting them remains challenging. The results obtained using GT far surpass those of traditional algorithms, highlighting the impressive ability of neural networks to address such problems and the clear advantage of nonlinear methods. However, parameter tuning to achieve these high performance levels is impossible, due to the lack of reference images.

Most contributions in the literature report results comparable to those obtained using the GT-based cost function in this study. This similarity arises because, although the training and test images are not identical, subimages from the same dataset are often used for training. As a result, the issue of different resolutions in the reference images is left unresolved. Another persistent challenge, which also remains unaddressed in this work, is the absence of a validation set that mirrors the critical aspects of the training set. This absence makes it difficult to select optimal network hyperparameters and assess potential overfitting, a particularly crucial issue when generalizing to different scales. A more specific analysis is needed to explore suitable tools for mitigating this problem. Traditional techniques such as cross-validation and regularization must be adapted to account for varying scales of observation.

This paper contributes to the evaluation of neural-network-based pansharpening algorithms by assessing their practical performance when using no-reference quality measures as cost functions. The best results were obtained using the RQNR cost function. This function combines a well-established spectral distortion index to assess the consistency between the fused image and the original MS image, along with a spatial distortion index based on a linear regression model between the fused image channels and the PAN image. However, this choice does not resolve the debate over how best to evaluate spatial information quality, as the regression-based approach implements a different rationale from the Wald protocol [60]. Moreover, the regression model could potentially be improved by using a nonlinear approach, possibly through the implementation of neural networks [102,103].

Using cost functions based on high-performing no-reference quality indices improves neural network performance. However, this study shows that the initial performance of these networks, without specific fine-tuning, is often poor, even for models with few parameters. This makes the fine-tuning process more challenging. Approaches focused on generalization, such as the general image fusion framework [53] or foundation models [104], may offer a better starting point.

## 7. Conclusions

In this study, we analyzed the performances achievable with various simple CNN architectures proposed for pansharpening, focusing on the impact of different cost functions derived from key no-reference image quality indices.

Using a reduced resolution assessment protocol, we highlighted the gap between the theoretically excellent performance of CNN architectures and the more limited results achievable with cost functions that can be practically implemented. Despite the potential of these architectures, this discrepancy underscores the challenges associated with selecting effective cost functions that align with real-world constraints.

Our findings underscore the superiority of a reference-free cost function that incorporates a highly accredited spectral quality index, combined with an innovative spatial quality index. This combination proved effective in linking the channels of the fused image to the available panchromatic image through a simple yet powerful relationship. This approach not only enhances the fusion process but also offers a promising pathway for achieving better pansharpening results without relying on ground-truth references.

This work provides a starting point for evaluating the performance of ML-based pansharpening techniques while highlighting key questions for future research. The most critical challenges involve optimizing the trade-off between algorithm generalization and image quality for specific datasets. On the one hand, this requires designing architectures that deliver robust performance across a broad range of problems and, in the context of pansharpening, ensure sufficient scale invariance. On the other hand, developing reliable techniques to monitor product quality during fine-tuning is essential. This task can be supported by using simplified cost functions, like those introduced in this study, which produce smoother error surfaces and facilitate optimization.

**Funding:** The work of R. Restaino was partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001—program "RESTART").

**Data Availability Statement:** The data presented in this study are available in PAirMax (Maxar Data) at https://resources.maxar.com/product-samples/pansharpening-benchmark-dataset (accessed on 20 December 2024).

**Acknowledgments:** We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan XP GPU used for this research.

**Conflicts of Interest:** The author declares no conflicts of interest.

## References

1. Fernandez-Beltran, R.; Latorre-Carmona, P.; Pla, F. Single-frame Super-resolution in Remote Sensing: A Practical Overview. *Int. J. Remote Sens.* **2017**, *38*, 314–354. [CrossRef]
2. Dalla Mura, M.; Prasad, S.; Pacifici, F.; Gamba, P.; Chanussot, J.; Benediktsson, J.A. Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing. *Proc. IEEE* **2015**, *103*, 1585–1601. [CrossRef]
3. Ghamisi, P.; Rasti, B.; Yokoya, N.; Wang, Q.; Hofle, B.; Bruzzone, L.; Bovolo, F.; Chi, M.; Anders, K.; Gloaguen, R.; et al. Multisource and Multitemporal Data Fusion in Remote Sensing: A Comprehensive Review of the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 6–39. [CrossRef]
4. Alparone, L.; Aiazzi, B.; Baronti, S.; Garzelli, A. *Remote Sensing Image Fusion*; CRC Press: Boca Raton, FL, USA, 2015.
5. Vivone, G.; Alparone, L.; Chanussot, J.; Dalla Mura, M.; Garzelli, A.; Licciardi, G.; Restaino, R.; Wald, L. A Critical Comparison Among Pansharpening Algorithms. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2565–2586. [CrossRef]
6. Meng, X.; Shen, H.; Li, H.; Zhang, L.; Fu, R. Review of the Pansharpening Methods for Remote Sensing Images Based on the Idea of Meta-analysis: Practical Discussion and Challenges. *Inf. Fusion* **2019**, *46*, 102–113. [CrossRef]
7. Sirguey, P.; Mathieu, R.; Arnaud, Y.; Khan, M.M.; Chanussot, J. Improving MODIS Spatial Resolution for Snow Mapping Using Wavelet Fusion and ARSIS Concept. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 78–82. [CrossRef]
8. Bovolo, F.; Bruzzone, L.; Capobianco, L.; Garzelli, A.; Marchesi, S.; Nencini, F. Analysis of the Effects of Pansharpening in Change Detection on VHR Images. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 53–57. [CrossRef]
9. Vivone, G.; Dalla Mura, M.; Garzelli, A.; Scarpa, G.; Ulfarsson, M.; Restaino, R.; Alparone, L.; Chanussot, J. A New Benchmark Based on Recent Advances in Multispectral Pansharpening: Revisiting Pansharpening with Classical and Emerging Pansharpening Method. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 53–81. [CrossRef]

10. Wald, L.; Ranchin, T.; Mangolini, M. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 691–699.
11. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A.; Selva, M. Twenty-Five Years of Pansharpening: A Critical Review and New Developments. In *Signal and Image Processing for Remote Sensing*, 2nd ed.; Chen, C.H., Ed.; CRC Press: Boca Raton, FL, USA, 2012; pp. 533–548.
12. Alparone, L.; Baronti, S.; Aiazzi, B.; Garzelli, A. Spatial Methods for Multispectral Pansharpening: Multiresolution Analysis Demystified. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2563–2576. [CrossRef]
13. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2300–2312. [CrossRef]
14. González-Audícana, M.; Otazu, X.; Fors, O.; Seco, A. Comparison between Mallat's and the 'à trous' discrete wavelet transform based algorithms for the fusion of multispectral and panchromatic images. *Int. J. Remote Sens.* **2005**, *26*, 595–614. [CrossRef]
15. Otazu, X.; González-Audícana, M.; Fors, O.; Núñez, J. Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2376–2385. [CrossRef]
16. Nencini, F.; Garzelli, A.; Baronti, S.; Alparone, L. Remote sensing image fusion using the curvelet transform. *Inf. Fusion* **2007**, *8*, 143–156. [CrossRef]
17. Shah, V.P.; Younan, N.H.; King, R.L. An efficient pan-sharpening method via a combined adaptive-PCA approach and contourlets. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1323–1335. [CrossRef]
18. Restaino, R.; Vivone, G.; Dalla Mura, M.; Chanussot, J. Fusion of Multispectral and Panchromatic Images Based on Morphological Operators. *IEEE Trans. Image Process.* **2016**, *25*, 2882–2895. [CrossRef] [PubMed]
19. Chavez, P.S., Jr.; Kwarteng, A.W. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens.* **1989**, *55*, 339–348.
20. Laben, C.A.; Brower, B.V. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening. U.S. Patent 6011875A, 4 January 2000.
21. Gillespie, A.R.; Kahle, A.B.; Walker, R.E. Color enhancement of highly correlated images-II. Channel ratio and "Chromaticity" Transform techniques. *Remote Sens. Environ.* **1987**, *22*, 343–365. [CrossRef]
22. Garzelli, A.; Nencini, F.; Capobianco, L. Optimal MMSE Pan sharpening of very high resolution multispectral images. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 228–236. [CrossRef]
23. Garzelli, A. Pansharpening of Multispectral Images Based on Nonlocal Parameter Optimization. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2096–2107. [CrossRef]
24. Schowengerdt, R.A. *Remote Sensing: Models and Methods for Image Processing*, 3rd ed.; Elsevier: Orlando, FL, USA, 2007.
25. Vivone, G.; Restaino, R.; Chanussot, J. A Regression-Based High-Pass Modulation Pansharpening Approach. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 984–996. [CrossRef]
26. Vivone, G.; Restaino, R.; Chanussot, J. Full Scale Regression-Based Injection Coefficients for Panchromatic Sharpening. *IEEE Trans. Image Process.* **2018**, *27*, 3418–3431. [CrossRef]
27. Aiazzi, B.; Baronti, S.; Selva, M. Improving component substitution pansharpening through multivariate regression of MS+Pan data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3230–3239. [CrossRef]
28. Alparone, L.; Wald, L.; Chanussot, J.; Thomas, C.; Gamba, P.; Bruce, L.M. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data fusion contest. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3012–3021. [CrossRef]
29. Lolli, S.; Alparone, L.; Garzelli, A.; Vivone, G. Haze Correction for Contrast-Based Multispectral Pansharpening. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2255–2259. [CrossRef]
30. Vivone, G. Robust Band-Dependent Spatial-Detail Approaches for Panchromatic Sharpening. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6421–6433. [CrossRef]
31. Restaino, R.; Dalla Mura, M.; Vivone, G.; Chanussot, J. Context-adaptive Pansharpening Based on Image Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 753–766. [CrossRef]
32. Vivone, G.; Simões, M.; Dalla Mura, M.; Restaino, R.; Bioucas-Dias, J.; Licciardi, G.; Chanussot, J. Pansharpening Based on Semiblind Deconvolution. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1997–2010. [CrossRef]
33. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O. A New Pansharpening Algorithm Based on Total Variation. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 318–322. [CrossRef]
34. Palsson, F.; Ulfarsson, M.O.; Sveinsson, J.R. Model-Based Reduced-Rank Pansharpening. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 656–660. [CrossRef]
35. Ulfarsson, M.O.; Palsson, F.; Dalla Mura, M.; Sveinsson, J.R. Sentinel-2 Sharpening Using a Reduced-Rank Method. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6408–6420. [CrossRef]
36. Li, S.; Yang, B. A New Pan-Sharpening Method Using a Compressed Sensing Technique. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 738–746. [CrossRef]

37. Zhu, X.X.; Bamler, R. A Sparse Image Fusion Algorithm with Application to Pan-Sharpening. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2827–2836. [CrossRef]
38. Vicinanza, M.R.; Restaino, R.; Vivone, G.; Dalla Mura, M.; Licciardi, G.; Chanussot, J. A Pansharpening Method Based on the Sparse Representation of Injected Details. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 180–184. [CrossRef]
39. Liu, J.; Zhou, C.; Fei, R.; Zhang, C.; Zhang, J. Pansharpening Via Neighbor Embedding of Spatial Details. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4028–4042. [CrossRef]
40. Han, X.; Leng, W.; Xu, Q.; Li, W.; Tao, R.; Sun, W. A Joint Optimization Based Pansharpening via Subpixel-Shift Decomposition. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5407815. [CrossRef]
41. Huang, W.; Xiao, L.; Wei, Z.; Liu, H.; Tang, S. A New Pan-Sharpening Method with Deep Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1037–1041. [CrossRef]
42. Masi, G.; Cozzolino, D.; Verdoliva, L.; Scarpa, G. Pansharpening by Convolutional Neural Networks. *Remote Sens.* **2016**, *8*, 594. [CrossRef]
43. Yang, J.; Fu, X.; Hu, Y.; Huang, Y.; Ding, X.; Paisley, J. PanNet: A Deep Network Architecture for Pan-Sharpening. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 1753–1761.
44. Scarpa, G.; Vitale, S.; Cozzolino, D. Target-adaptive CNN-based pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5443–5457. [CrossRef]
45. Deng, L.J.; Vivone, G.; Paoletti, M.E.; Scarpa, G.; He, J.; Zhang, Y.; Chanussot, J.; Plaza, A. Machine Learning in Pansharpening: A Benchmark, from Shallow to Deep Networks. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 279–315. [CrossRef]
46. Lei, D.; Chen, H.; Zhang, L.; Li, W. NLRNet: An Efficient Nonlocal Attention ResNet for Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5401113. [CrossRef]
47. Meng, X.; Wang, N.; Shao, F.; Li, S. Vision Transformer for Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5409011. [CrossRef]
48. Liu, X.; Wang, Y.; Liu, Q. PSGAN: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpening. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 873–877.
49. Ma, J.; Yu, W.; Chen, C.; Liang, P.; Guo, X.; Jiang, J. Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. *Inf. Fusion* **2020**, *62*, 110–120. [CrossRef]
50. Ozcelik, F.; Alganci, U.; Sertel, E.; Unal, G. Rethinking CNN-Based Pansharpening: Guided Colorization of Panchromatic Images via GANs. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 3486–3501. [CrossRef]
51. Meng, Q.; Shi, W.; Li, S.; Zhang, L. PanDiff: A Novel Pansharpening Method Based on Denoising Diffusion Probabilistic Model. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5611317. [CrossRef]
52. Hosseiny, B.; Mahdianpari, M.; Hemati, M.; Radman, A.; Mohammadimanesh, F.; Chanussot, J. Beyond Supervised Learning in Remote Sensing: A Systematic Review of Deep Learning Approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 1035–1052. [CrossRef]
53. Wang, W.; Deng, L.J.; Vivone, G. A general image fusion framework using multi-task semi-supervised learning. *Inf. Fusion* **2024**, *108*, 102414. [CrossRef]
54. Luo, S.; Zhou, S.; Feng, Y.; Xie, J. Pansharpening via Unsupervised Convolutional Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4295–4310. [CrossRef]
55. Diao, W.; Zhang, F.; Sun, J.; Xing, Y.; Zhang, K.; Bruzzone, L. ZeRGAN: Zero-Reference GAN for Fusion of Multispectral and Panchromatic Images. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 8195–8209. [CrossRef] [PubMed]
56. Yang, G.; Zhang, K.; Zhang, F.; Wang, J.; Sun, J. Cross-Resolution Semi-Supervised Adversarial Learning for Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5402617. [CrossRef]
57. Cao, Q.; Deng, L.J.; Wang, W.; Hou, J.; Vivone, G. Zero-shot semi-supervised learning for pansharpening. *Inf. Fusion* **2024**, *101*, 102001. [CrossRef]
58. He, J.; Yuan, Q.; Li, J.; Xiao, Y.; Zhang, L. A self-supervised remote sensing image fusion framework with dual-stage self-learning and spectral super-resolution injection. *ISPRS J. Photogramm. Remote Sens.* **2023**, *204*, 131–144. [CrossRef]
59. Xing, Y.; Qu, L.; Zhang, S.; Zhang, K.; Zhang, Y.; Bruzzone, L. CrossDiff: Exploring Self-SupervisedRepresentation of Pansharpening via Cross-Predictive Diffusion Model. *IEEE Trans. Image Process.* **2024**, *33*, 5496–5509. [CrossRef] [PubMed]
60. Arienzo, A.; Vivone, G.; Garzelli, A.; Alparone, L.; Chanussot, J. Full-Resolution Quality Assessment of Pansharpening: Theoretical and hands-on approaches. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 168–201. [CrossRef]
61. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A.; Selva, M. MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 591–596. [CrossRef]
62. Vivone, G.; Dalla Mura, M.; Garzelli, A.; Pacifici, F. A Benchmarking Protocol for Pansharpening: Dataset, Preprocessing, and Quality Assessment. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6102–6118. [CrossRef]
63. Alparone, L.; Aiazzi, B.; Baronti, S.; Garzelli, A.; Nencini, F.; Selva, M. Multispectral and panchromatic data fusion assessment without reference. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 193–200. [CrossRef]

64. Xiong, Z.; Guo, Q.; Liu, M.; Li, A. Pan-Sharpening Based on Convolutional Neural Network by Using the Loss Function with No-Reference. *IEEE J. Sel. Top. Appl. Earth Observ.* **2021**, *14*, 897–906. [CrossRef]
65. Li, S.; Guo, Q.; Li, A. Pan-Sharpening Based on CNN+ Pyramid Transformer by Using No-Reference Loss. *Remote Sens.* **2022**, *14*, 624. [CrossRef]
66. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
67. Gao, J.; Li, J.; Su, X.; Jiang, M.; Yuan, Q. Deep Image Interpolation: A Unified Unsupervised Framework for Pansharpening. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 19–20 June 2022; pp. 609–618.
68. Ciotola, M.; Vitale, S.; Mazza, A.; Poggi, G.; Scarpa, G. Pansharpening by Convolutional Neural Networks in the Full Resolution Framework. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5408717. [CrossRef]
69. Ciotola, M.; Poggi, G.; Scarpa, G. Unsupervised Deep Learning-Based Pansharpening with Jointly Enhanced Spectral and Spatial Fidelity. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5405417. [CrossRef]
70. Ciotola, M.; Scarpa, G. Fast Full-Resolution Target-Adaptive CNN-Based Pansharpening Framework. *Remote Sens.* **2023**, *15*, 319. [CrossRef]
71. Ciotola, M.; Guarino, G.; Scarpa, G. An Unsupervised CNN-Based Pansharpening Framework with Spectral-Spatial Fidelity Balance. *Remote Sens.* **2024**, *16*, 3014. [CrossRef]
72. Wald, L. *Data Fusion: Definitions and Architectures—Fusion of Images of Different Spatial Resolutions*; Les Presses de l'École des Mines: Paris, France, 2002.
73. Khan, M.M.; Alparone, L.; Chanussot, J. Pansharpening quality assessment using the modulation transfer functions of instruments. *IEEE Trans. Geosci. Remote Sens.* **2009**, *11*, 3880–3891. [CrossRef]
74. Xiong, Z.; Liu, N.; Wang, N.; Sun, Z.; Li, W. Unsupervised Pansharpening Method Using Residual Network with Spatial Texture Attention. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5402112. [CrossRef]
75. Qu, Y.; Baghbaderani, R.; Qi, H.; Kwan, C. Unsupervised Pansharpening Based on Self-Attention Mechanism. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 3192–3208. [CrossRef]
76. Ni, J.; Shao, Z.; Zhang, Z.; Hou, M.; Zhou, J.; Fang, L.; Zhang, Y. LDP-Net: An Unsupervised Pansharpening Network Based on Learnable Degradation Processes. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5468–5479. [CrossRef]
77. Wang, D.; Zhang, P.; Bai, Y.; Li, Y. MetaPan: Unsupervised Adaptation with Meta-Learning for Multispectral Pansharpening. *IEEE Signal Process. Lett.* **2022**, *19*, 5513505. [CrossRef]
78. Zhang, P.; Bai, Y.; Wang, D.; Bai, B.; Li, Y. Few-Shot Classification of Aerial Scene Images via Meta-Learning. *Remote Sens.* **2021**, *13*, 108. [CrossRef]
79. Zhou, C.; Zhang, J.; Liu, J.; Zhang, C.; Fei, R.; Xu, S. PercepPan: Towards Unsupervised Pan-Sharpening Based on Perceptual Loss. *Remote Sens.* **2020**, *12*, 2318. [CrossRef]
80. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; pp. 694–711.
81. Uezato, T.; Hong, D.; Yokoya, N.; He, W. Guided Deep Decoder: Unsupervised Image Pair Fusion. In *Computer Vision—ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer: Cham, Switzerland, 2020; pp. 87–102.
82. Zhou, H.; Liu, Q.; Wang, Y. PGMAN: An Unsupervised Generative Multiadversarial Network for Pansharpening. *IEEE J. Sel. Top. Signal Process.* **2021**, *14*, 6316–6327. [CrossRef]
83. Rui, X.; Cao, X.; Pang, L.; Zhu, Z.; Yue, Z.; Meng, D. Unsupervised hyperspectral pansharpening via low-rank diffusion model. *Inf. Fusion* **2024**, *107*, 102325. [CrossRef]
84. Aiazzi, B.; Alparone, L.; Baronti, S.; Carlà, R.; Garzelli, A.; Santurri, L. Full scale assessment of pansharpening methods and data products. In Proceedings of the Volume 9244, Image and Signal Processing for Remote Sensing XX, Amsterdam, The Netherlands, 22–25 September 2014; p. 924402.
85. Alparone, L.; Garzelli, A.; Vivone, G. Spatial Consistency for Full-scale Assessment of Pansharpening. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1–3.
86. Deng, L.; Vivone, G.; Jin, C.; Chanussot, J. Detail Injection-based Deep Convolutional Neural Networks for Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 6995–7010. [CrossRef]
87. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [CrossRef]
88. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
89. Fu, X. PanNet: A Deep Network Architecture for Pan-Sharpening—Training Code. 2017. Available online: https://xueyangfu.github.io/projects/iccv2017.html (accessed on 5 September 2024).

90. Vitale, S. Pansharpening-cnn-Matlab-Version. 2019. Available online: https://github.com/sergiovitale/pansharpening-cnn-matlab-version (accessed on 5 September 2024).
91. Deng, L. FusionNet. 2020. Available online: https://github.com/liangjiandeng/FusionNet (accessed on 5 September 2024).
92. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]
93. Alparone, L.; Baronti, S.; Garzelli, A.; Nencini, F. A global quality measurement of pan-sharpened multispectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2004**, *1*, 313–317. [CrossRef]
94. Garzelli, A.; Nencini, F. Hypercomplex quality assessment of multi/hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2009**, *6*, 662–665. [CrossRef]
95. Zhou, J.; Civco, D.L.; Silander, J.A. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. *Int. J. Remote Sens.* **1998**, *19*, 743–757. [CrossRef]
96. Draper, N.; Smith, H. *Applied Regression Analysis*; Joh Wiley & Sons: Hoboken, NJ, USA, 1998.
97. Maxar Technologies. Pansharpening Benchmark Dataset ©IEEE. 2021. Available online: https://resources.maxar.com/product-samples/pansharpening-benchmark-dataset/ (accessed on 5 September 2024).
98. Yuhas, R.H.; Goetz, A.F.H.; Boardman, J.W. Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm. In Proceedings of the Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 2: TIMS Workshop, Pasadena, CA, USA, 1–5 June 1992; pp. 147–149.
99. Open Remote Sensing. Pansharpening Benchmark Dataset ©IEEE. 2020. Available online: https://openremotesensing.net/knowledgebase/a-new-benchmark-based-on-recent-advances-in-multispectral-pansharpening-revisiting-pansharpening-with-classical-and-emerging-pansharpening-methods/ (accessed on 5 September 2024).
100. Zhang, Y.; Liu, C.; Sun, M.; Ou, Y. Pan-Sharpening Using an Efficient Bidirectional Pyramid Network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5549–5563. [CrossRef]
101. Deng, L. FusionNet. 2022. Available online: https://github.com/liangjiandeng/DLPan-Toolbox (accessed on 12 December 2024).
102. Wen, R.; Deng, L.; Wu, Z.; Wu, X.; Vivone, G. A Novel Spatial Fidelity with Learnable Nonlinear Mapping for Panchromatic Sharpening. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5401915. [CrossRef]
103. Zhang, H.; Wang, H.; Tian, X.; Ma, J. P2Sharpen: A progressive pansharpening network with deep spectral transformation. *Inf. Fusion* **2023**, *91*, 103–122. [CrossRef]
104. Lu, S.; Guo, J.; Zimmer-Dauphinee, J.; Nieusma, J.; Wang, X.; VanValkenburgh, P.; Wernke, S.; Huo, Y. AI Foundation Models in Remote Sensing: A Survey. *arXiv* **2024**, arXiv:2408.03464.


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.