remote sensing

MDPI

Article

A Three Stages Detail Injection Network for Remote Sensing Images Pansharpening

Yuanyuan Wu [1], Siling Feng [1], Cong Lin [1], Haijie Zhou [1] and Mengxing Huang [1,2,*]

[1] School of Information and Communication Engineering, Hainan University, Haikou 570228, China; wuyuanyuan992@hainanu.edu.cn (Y.W.); fengsiling@hainanu.edu.cn (S.F.); lincong@hainanu.edu.cn (C.L.); 20081000110034@hainanu.edu.cn (H.Z.)

[2] State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou 570228, China

* Correspondence: huangmx09@hainanu.edu.cn

Abstract: Multispectral (MS) pansharpening is crucial to improve the spatial resolution of MS images. MS pansharpening has the potential to provide images with high spatial and spectral resolutions. Pansharpening technique based on deep learning is a topical issue to deal with the distortion of spatio-spectral information. To improve the preservation of spatio-spectral information, we propose a novel three-stage detail injection pansharpening network (TDPNet) for remote sensing images. First, we put forward a dual-branch multiscale feature extraction block, which extracts four scale details of panchromatic (PAN) images and the difference between duplicated PAN and MS images. Next, cascade cross-scale fusion (CCSF) employs fine-scale fusion information as prior knowledge for the coarse-scale fusion to compensate for the lost information during downsampling and retain high-frequency details. CCSF combines the fine-scale and coarse-scale fusion based on residual learning and prior information of four scales. Last, we design a multiscale detail compensation mechanism and a multiscale skip connection block to reconstruct injecting details, which strengthen spatial details and reduce parameters. Abundant experiments implemented on three satellite data sets at degraded and full resolutions confirm that TDPNet trades off the spectral information and spatial details and improves the fidelity of sharper MS images. Both the quantitative and subjective evaluation results indicate that TDPNet outperforms the compared state-of-the-art approaches in generating MS images with high spatial resolution.

Keywords: multispectral images; pansharpening; convolutional neural network; cascade cross-scale; detail compensation mechanism

Citation: Wu, Y.; Feng, S.; Lin, C.; Zhou, H.; Huang, M. A Three Stages Detail Injection Network for Remote Sensing Images Pansharpening. *Remote Sens.* **2022**, *14*, 1077. https://doi.org/10.3390/rs14051077

Academic Editor: Giuseppe Scarpa

Received: 29 January 2022
Accepted: 18 February 2022
Published: 22 February 2022

1. Introduction

Remote sensing images (RSIs) are broadly employed in different aspects, for instance, obtaining geographic data, obtaining earth resource information, hazard prediction and analysis, urban investigation, yield estimation and others [1]. However, in these applications, RSIs with high spatial, spectral or time resolution are usually required [2–4]. The spatial and spectral resolutions of RSIs constrain each other limited by sensor technology, i.e., panchromatic (PAN) images are with high spatial and low spectral resolutions, multispectral (MS) images are with low spatial and multispectral resolutions (LRMS) and hyperspectral (HS) images are with low spatial and high spectral resolutions [1,5]. The PAN images and MS or HS images need to be fused to produce high spatial resolution MS (HRMS) images or high spatial resolution HS (HRHS) images. This technique is also called panchromatic sharpening (pansharpening). Pansharpening for MS and PAN images is studied. Pansharpening methods can be approximately comprised of traditional approaches and deep learning (DL) methods [1,2,5]. Traditional approaches comprise the component substitution (CS) approach, multiresolution analysis (MRA) method and variational optimization (VO) technique [2,5].

CS methods first transform the MS images to other coordinate systems, and extract the spatial detail information. Then, the spatial detail information is substituted for the PAN image. Eventually, the substituted image is projected to the original coordinate system by inverse transformation to generate the HRMS image. Various CS methods have been developed, mainly including intensity-hue-saturation (IHS) [6], adaptive IHS (AIHS) [7], generalized IHS (GIHS) [8], Gram-Schmidt (GS) [9], GS adaptive (GSA) [1], Brovey [5] and partial replacement adaptive component substitution (PRACS) [10]. CS methods are simple and easy to implement, which greatly improve the spatial resolution of MS images. CS methods have the disadvantage of severe spectrum distortion and oversharpening.

MRA methods first decompose MS images and PAN images into images of different scales. Then, the corresponding scale images are fused by a fusion technique. Finally, an HRMS image is generated by an inverse transformation. The decomposition methods used generally include the discrete wavelet transform (DWT) [11,12], fusion for MS and PAN images employing the Induction scaling approach (Indusion) [13], generalized Laplacian pyramid (GLP) transform [14], modulation transfer function-GLP (MTF_GLP) transform [15], à trous wavelet transform (ATWT) [16], and other spectral and wavelet decomposition techniques [17–19]. MRA methods retain more spectral information and lessen the spectrum distortion. However, the spatial information is not rich and the resolution is lower.

The key to the VO method is to establish an energy function and optimization method [20–23]. Bayesian-based methods [24,25] and sparse representation-based methods [26–29] can also be classified into this category. Although the VO method can reduce spectral distortion, the optimization calculation is more complicated.

With the rapid development of DL, various types of convolutional neural networks (CNN)-based models are increasingly used in pansharpening and closely related tasks [30–32] of RSIs. Giuseppe et al. [33] proposed a CNN-based pansharpening method (PNN). PNN consists of only three layers and uses the nonlinear mapping of the CNN to reconstruct the LRMS image generating the HRMS image. The advantage of the PNN is that it has few layers and is easy to implement, but it also has the disadvantage of overfitting and limited expression ability. Wei et al. [34] put forward a deep residual network-based pansharpening technique (DRPNN). The DRPNN employs residual blocks to improve the fusion ability and reduce overfitting. Yang et al. [35] put forward a PanNet method based on residual modules and trained it in the frequency domain. To retain more spectrum information, add the upsampled MS image to the residual information. The model is trained in the frequency field to retain more spatial structure information, and the generalization ability of the network is better. Scarpa et al. [36] proposed a target-adaptive pansharpening means based on a CNN (TA-PNN). TA-PNN proposes a target-adaptive usage mode to deal with problems of data mismatch, multisensor data, and insufficient data. Liu et al. [37] designed a PsGan technique, which contains a generator and discriminator. The generator is a two-stream fusion structure that generates an HRMS image with MS and PAN images as inputs. The discriminator is composed of a full CNN to discriminate between the reference image and the produced HRMS image. Ma et al. [38] proposed a generative adversarial network-based model for pansharpening (Pan-Gan). Pan-Gan employs a spectral discriminator to discriminate the spectral information between the fused HRMS image and LRMS image. The spatial discriminator is employed to discriminate the spatial structure information between the HRMS and PAN images. Zhao et al. [39] designed an FGF-GAN method. The FGF-GAN generator uses a fast guided filter to retain details and a spatial attention module for fusion. FGF-GAN reduces network parameters and training time. Deng et al. [40] designed a CS/MRA model-based detail injection network (FusionNet). The injected details are acquired through the deep CNN based on residual learning, and then the upsampled MS image is added to the output of the detail extraction network. For FusionNet, the difference between the MS and PAN images (i.e., the duplicated PAN image of N channels, N is the channels of the MS image) is taken as the input. The multispectral information is introduced into the detail extraction network to lessen the spectral distortion.

Wu et al. [41] designed a residual module-based distributed fusion network (RDFNet). RDFNet extracts multilevel features of MS images and PAN images, respectively. Then the corresponding level features and the fusion result of the previous step are fused to obtain the HRMS image. Although the network uses multilevel MS and PAN features as much as possible, it is affected by the depth of the network and cannot obtain more details and spectral information. Obviously, although various networks are used for pansharpening of RSIs and have acquired good results, there is rising space in terms of model complexity, implementation time, generalization ability, spectrum fidelity, retention of spatial details and so on.

In this article, we propose a novel three-stage detail injection network for pansharpening of RSIs by preserving spectral information to reduce spectral distortion and preserving details to strengthen spatial resolution. The main contributions of the work are as follows.

- A dual-branch multiscale feature extraction block is established to extract four-scale details of PAN images and the difference between duplicated PAN and MS images. The details are retained, and the MS image is introduced to preserve the spectrum information.
- Cascade cross-scale fusion (CCSF) employs fine-scale fusion information as prior knowledge for the coarse-scale fusion based on residual learning. CCSF combines the fine-scale and coarse-scale fusion, which compensates for the loss of information and retains details.
- A multiscale high-frequency detail compensation mechanism and a multiscale skip connection block are designed to reconstruct the fused details, which strengthen spatial details and reduce parameters.
- The quantitative evaluation and subjective evaluation of three satellite data sets are implemented at reduced and full resolutions.

Section 2 represents the data sets, evaluation indicators, implementation settings and proposed method at full length. Section 3 introduces comparative experiments on three data sets. Section 4 discusses the experimental results. Section 5 draws conclusions.

2. Materials and Methods

2.1. Data Sets

To prove the performance of the designed pansharpening approach, three data sets are used for evaluation. The specific information of these data sets is as follows.

Gaofen-1 (GF-1) data set: These data are collected from Guangdong and Yunnan, China. The MS images have 4 bands. The resolutions of the PAN and MS images are 2 m and 8 m, respectively, and the radiometric resolution is 10 bits. The reduced resolution data are produced by Wald's protocol [42], and the training samples, validation data and testing data are randomly selected. The number of data pairs in the data set is 21,560, 6160 and 3080, respectively. The number of full resolution testing data is 147.

QuickBird data set: These data are collected from Chengdu, Beijing, Shenyang and Zhengzhou, China. The MS images have 4 bands. The resolutions of the PAN and MS images are 0.6 m and 2.4 m, respectively, and the radiometric resolution is 11 bits. The number of training samples, validation data and testing data pairs is 20,440, 5840 and 2920, respectively. The number of full resolution testing data is 158.

Gaofen-2 (GF-2) data set: These data are collected from Beijing and Shenyang, China. The MS images have 4 bands. The resolutions of the PAN and MS images are 1 m and 4 m, respectively, and the radiometric resolution is 10 bits. The number of training samples, validation data and testing data pairs is 21,000, 6000 and 3000, respectively. The number of full resolution testing data is 286.

The sizes of the LMS (the reduced resolution form of the MS image), MS and P_L (the reduced resolution form of the PAN image) images of the training data are $16 \times 16 \times 4$, $64 \times 64 \times 4$ and $64 \times 64 \times 1$, respectively. The sizes of the MS and PAN images of the testing data at full resolution are $100 \times 100 \times 4$ and $400 \times 400 \times 1$, respectively.

2.2. Evaluation Indicators

The evaluation of pansharpening performance is performed at reduced and full resolutions. Subjective visual evaluation and objective evaluation are implemented on the experimental results. The objective evaluation indicators used in the reduced resolution experiment include the universal image quality index (UIQI) [43] extended to 4-band (Q4) [44], spectral angle mapping (SAM) [44], structural similarity (SSIM) [45], spatial correlation coefficient (SCC) [44] and erreur relative global adimensionnelle de Synthése (ERGAS) [44].

UIQI [43] assesses image quality from three sides: correlation, luminance and contrast. The representation of *UIQI* is Equation (1).

$$UIQI(g,f) = \frac{4\sigma_{gf} \cdot \bar{g} \cdot \bar{f}}{\left(\sigma_g^2 + \sigma_f^2\right)\left[(\bar{g})^2 + (\bar{f})^2\right]} \tag{1}$$

where g and f indicate the ground truth (GT) and pansharpened images, respectively. σ_{gf} means the covariance between g and f images. $\bar{g}$ and σ_g are the means and variance of g. $\bar{f}$ and σ_f are the means and variance of f. The optimal value for *UIQI* is 1, and the pansharpening result is optimum.

The expression for $Q4$ [44] is Equation (2).

$$Q4 = \frac{4\left|\sigma_{g_z f_z}\right| \cdot \left|\overline{\mathbf{g}_z}\right| \cdot \left|\overline{\mathbf{f}_z}\right|}{\left(\sigma_{g_z}^2 + \sigma_{f_z}^2\right)\left(\left|\overline{\mathbf{g}_z}\right|^2 + \left|\overline{\mathbf{f}_z}\right|^2\right)} \tag{2}$$

where $\mathbf{g}_z$ and $\mathbf{f}_z$ indicate the 4-band GT and pansharpened images, respectively. $\sigma_{g_z f_z}$ is the covariance between $\mathbf{g}_z$ and $\mathbf{f}_z$. $\overline{\mathbf{g}_z}$ and σ_{g_z} are the means and variance of $\mathbf{g}_z$. $\overline{\mathbf{f}_z}$ and σ_{f_z} are the means and variance of $\mathbf{f}_z$.

SAM [44] is an error indicator, which represents the angle difference of spectral vector between the GT and pansharpened images. The expression of *SAM* is Equation (3).

$$SAM(g,f) = \arccos\left(\frac{\langle g_v, f_v \rangle}{\|g_v\|_2 \|f_v\|_2}\right) \tag{3}$$

where g_v and f_v are the spectrum vector of the GT and pansharpened images. *SAM* expresses the spectral distortion. When *SAM* = 0, the pansharpening performance is the best.

SSIM [45] represents the proximity of structural information between the GT and pansharpened images. *SSIM* is shown as Equation (4).

$$SSIM(g,f) = \frac{\left(2\bar{g}\bar{f} + q_1\right)\left(2\sigma_{gf} + q_2\right)}{\left(\bar{g}^2 + \bar{f}^2 + q_1\right)\left(\sigma_g^2 + \sigma_f^2 + q_2\right)} \tag{4}$$

where q_1 and q_2 are constants. When *SSIM* = 1, the pansharpening result is the best.

SCC [44] represents the correlation of spatial details between the GT and pansharpened images. Spatial details are acquired through the high-pass filter. When *SCC* = 1, the pansharpened and GT images are most relevant.

ERGAS [44] is an error indicator, which shows the global effect of the pansharpened image. The representation of *ERGAS* is Equation (5).

$$ERGAS = 100\frac{R_P}{R_M}\sqrt{\frac{1}{N}\sum_{b=1}^{N}\left(\frac{\mathrm{RMSE}(b)}{\mu(b)}\right)^2} \tag{5}$$

where R_P and R_M indicate the spatial resolution of PAN and MS images, respectively. N is the number of bands. $RMSE(b)$ expresses the root mean square error of the bth band between the GT image and pansharpened image. $\mu(b)$ indicates the mean of the bth band. The smaller the *ERGAS*, the better the pansharpening result. The optimum value for *ERGAS* is 0.

The objective evaluation indicators used in the full resolution experiment include the quality with no-reference (QNR), D_λ and D_S [46].

D_λ denotes the spectrum distortion, the representation is Equation (6).

$$D_\lambda = \sqrt[s]{\frac{1}{N(N-1)} \sum_{b=1}^{N} \sum_{\substack{c=1 \\ c \neq b}}^{N} \left| UIQI(\tilde{m}_b, \tilde{m}_c) - UIQI\left(\hat{f}_b, \hat{f}_c\right) \right|^s} \tag{6}$$

where $\tilde{m}_b$ and $\tilde{m}_c$ are the b-band and c-band low spatial resolution MS images, $\hat{f}_b$ and $\hat{f}_c$ are the b-band and c-band pansharpened images, and s is a positive integer to amplify the difference.

D_S denotes the spatial distortion, the expression is Equation (7).

$$D_S = \sqrt[t]{\frac{1}{N} \sum_{b=1}^{N} \left| UIQI\left(\hat{f}_b, p\right) - UIQI(\tilde{m}_b, \tilde{p}) \right|^t} \tag{7}$$

where p and $\tilde{p}$ mean the PAN image and degraded version of PAN image, and t is a positive integer to amplify the difference.

The representation for *QNR* [46] is shown as Equation (8).

$$QNR = (1 - D_\lambda)^\eta (1 - D_S)^\rho \tag{8}$$

where η and ρ are constants. When $D_\lambda = 0$ and $D_S = 0$, *QNR* is the largest and the pansharpening effect is the best. The optimum value for *QNR* is 1, and the ideal values for D_λ and D_S are 0.

2.3. Implementation Settings

The implementation of the pansharpening network is the TensorFlow framework and a workstation containing an NVIDIA Tesla V100 PCIE GPU with 16 GB RAM and Intel Xeon CPU. The batch size is 32, and the number of iterations is 2.2×10^5. We employ the Adam optimizer [47] to optimize the pansharpening model, with the learning rate $\alpha = 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 10^{-8}$. The channels of input and output of the network can be set according to the image channels used. The number of output channels of the network is 4 since the bands of MS and PAN are 4 and 1. To compare the pansharpening effect fairly, the CNN-based experiments are completed on the GPU, and the CS/MRA-based experiments are conducted in MATLAB on the CPU.

2.4. Network Structure

We propose a novel three-stage detail injection pansharpening network (TDPNet) for RSIs. Figure 1 represents the structure of TDPNet. Because of the lack of reference images, TDPNet is trained on the reduced resolution image. We employ the reduced resolution images LMS and P_L of the MS and PAN images. The ↑LMS image is an upsampled LMS image, the same size as P_L. MS represents the reference image, and H_MS indicates the generated pansharpening result. Here, m is the ratio of the resolution of the MS and PAN images. TDPNet includes three stages: a dual-branch multiscale detail extraction stage, cascade cross-scale detail fusion stage, and reconstruction stage of injecting details. The dual-branch multiscale detail extraction stage extracts multiscale details from the PAN image and ↑LMS image generating details of four scales. The first stage consists of two branches. One branch extracts multiscale details from the PAN image. The other

branch extracts multiscale features from the difference image between the P_L^N image (the P_L image is duplicated N channels, where N is the channels of the LMS image.) and ↑LMS image. The second is a cascade cross-scale detail fusion stage. Cascade cross-scale fusion is achieved by combining fine-scale and coarse-scale fusion based on residual learning and prior information of four scales. Cascade cross-scale fusion employs the fine-scale fusion information as prior knowledge for the coarse-scale fusion to compensate for the loss of information caused by downsampling and retain details. The third is a reconstruction stage of injecting details. In this stage, the key information generated in the second stage is reconstructed. To compensate for the loss of information, we design a multiscale detail compensation mechanism. In addition, the fusion results generated in the cascade cross-scale fusion stage are used as the prior knowledge to reconstruct details (i.e., a multiscale skip connection block). This can strengthen spatial details and reduce parameters. Finally, the pansharpening result is produced by adding the reconstructed details to the ↑LMS image.

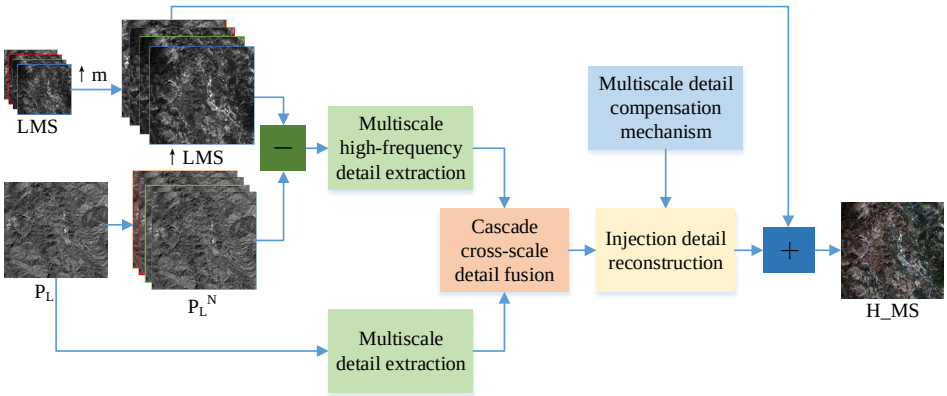

Figure 1. The structure of the three-stage detail injection network for pansharpening.

2.4.1. Dual-Branch Multiscale Detail Extraction Stage

FusionNet [40] introduces the details of the difference between MS and PAN images, which has better performance than directly obtaining the details from the PAN image. Inspired by this approach, we propose a dual-branch multiscale detail extraction network. The composition of the network is shown in Figure 2. The proposed network is trained using the reduced resolution P_L image and LMS image. One branch takes the difference between the ↑LMS and P_L^N images as the input to obtain the high-frequency details of four scales. The introduction of MS details reduces the spectrum distortion and details distortion caused by the lack of relevant spectrum information in the PAN image. The details extracted by this branch are named composite high-frequency details (CHFDs). The other branch is to acquire the details of four scales of the P_L image. In Figure 2, D_1–D_4 represent the extracted CHFDs of four scales. P_1–P_4 represent the extracted PAN details of the four scales.

The conv module in Figure 2 adopts a residual learning block [48], and Figure 3 shows the structure. t represents the dimension of the input, and n represents the dimension of the output. In Figure 2, the sizes of the convolution kernels of the four conv modules are $3 \times 3 \times 32$, $3 \times 3 \times 64$, $3 \times 3 \times 128$ and $3 \times 3 \times 256$. The downsampling adopts the maximum pooling operation, and the sizes of the convolution kernels are $2 \times 2 \times 64$, $2 \times 2 \times 128$ and $2 \times 2 \times 256$. The sizes of the D_4 and P_4 images are 1/8 of those of the D_1 and P_1 images. The expressions for extracting the CHFDs are Equations (9)–(12).

$$D_1 = H\left(P_L^N - \uparrow LMS\right) + \mathrm{F}\left(P_L^N - \uparrow LMS, W_{d1}\right) \tag{9}$$

$$H\left(P_L^N - \uparrow LMS\right) = W'_{d1} * \left(P_L^N - \uparrow LMS\right) \tag{10}$$

$$D_j = \varphi\left(H(D_{j-1}) + \mathrm{F}\left(D_{j-1}, W_{dj}\right)\right) \quad j = 2,3,4 \tag{11}$$

$$H(D_{j-1}) = W'_{dj} * D_{j-1} \quad j = 2,3,4 \tag{12}$$

where D_j ($j = 1,2,3,4$) shows the CHFDs of the jth scale, LMS and P_L are the reduced resolution images of the MS and PAN images, $\uparrow LMS$ is an upsampled LMS image, and $P_L^N - \uparrow LMS$ presents the difference between MS and PAN images. $H()$ represents the function of the direct connection part of the residual module. $*$ is a convolution operation. W'_{dj} ($j = 1,2,3,4$) represents the parameter of the direct connection part, and the convolution kernel size is 1×1, and the numbers are 32, 64, 128 and 256, respectively. $F()$ expresses the residual part function of the residual module, W_{dj} ($j = 1,2,3,4$) represents the parameter of the residual part, the convolution kernel size is 3×3, and the numbers are 32, 64, 128 and 256, respectively. $\varphi()$ indicates the function of the maximum pooling operation.

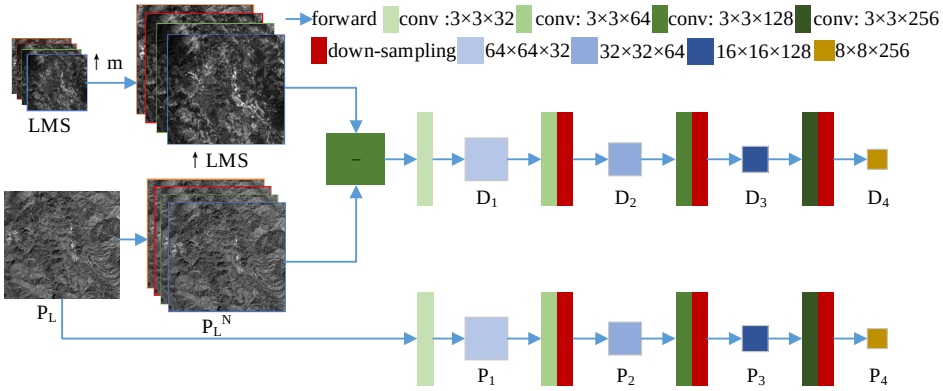

Figure 2. The dual-branch multiscale detail extraction network.

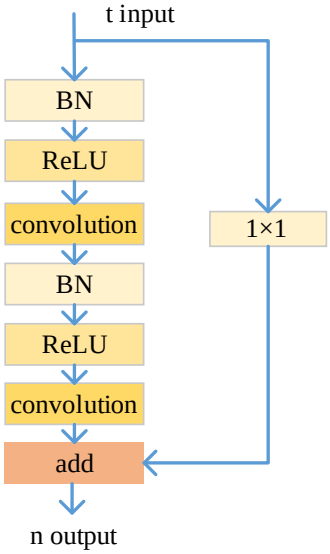

Figure 3. Residual learning block.

The expressions for extracting details from the P_L image are Equations (13)–(16).

$$P_1 = H(P_L) + \mathrm{F}(P_L, W_{p1}) \tag{13}$$

$$H(P_L) = W'_{p1} * P_L \tag{14}$$

$$P_j = \varphi(H(P_{j-1}) + \mathrm{F}(P_{j-1}, W_{pj})) \quad j = 2,3,4 \tag{15}$$

$$H(P_{j-1}) = W'_{pj} * P_{j-1} \quad j = 2,3,4 \tag{16}$$

where P_j $(j = 1, 2, 3, 4)$ means the PAN detail of the jth scale. $W'_{pj}(j = 1, 2, 3, 4)$ represents the parameter of the direct connection part, and the convolution kernel size is 1×1, and the numbers are 32, 64, 128 and 256, respectively. W_{pj} $(j = 1, 2, 3, 4)$ represents the parameter of the residual part, the convolution kernel size is 3×3, and the numbers are 32, 64, 128 and 256, respectively.

2.4.2. Cascade Cross-Scale Detail Fusion Stage

This section describes the second stage of the TDPNet method, i.e., the cascade cross-scale detail fusion stage. The structure of this stage is shown in Figure 4. In this stage, the CHFDs and PAN details of four corresponding scales are concatenated, and then they are fused at the same scale. CCSF employs fine-scale fusion information as prior knowledge for coarse-scale fusion. CCSF is achieved by combining fine-scale and coarse-scale fusion based on residual learning and prior information of four scales. The representation of each content in Figure 4 is the same as the representation in Figure 2. First, the CHFDs $D_j(j = 1, 2, 3, 4)$ and the PAN details $P_j(j = 1, 2, 3, 4)$ are concatenated and then fused at the same scale to generate the prior fusion result $P\_D_j(j = 1, 2, 3, 4)$. Then, the fine-scale fusion result P_D_1 provides the prior information for coarse-scale fusion. Cross-scale fusion (P_D_1 and P_D_2) requires a scale transfer module to convert the fine-scale information into coarse-scale information. The scale transfer module used is a maxpooling operation, as shown in the red module of Figure 4. Then P_D_1 is downsampled and P_D_5 is generated. The P_D_5 and P_D_2 are fused, and the fusion result P_D_6 provides a priori information for the fusion of the next scale (i.e., P_D_7). In this way, the CCSF of four scales is carried out by combining the fine-scale and coarse-scale fusion, and finally, the key information P_D_8 is generated.

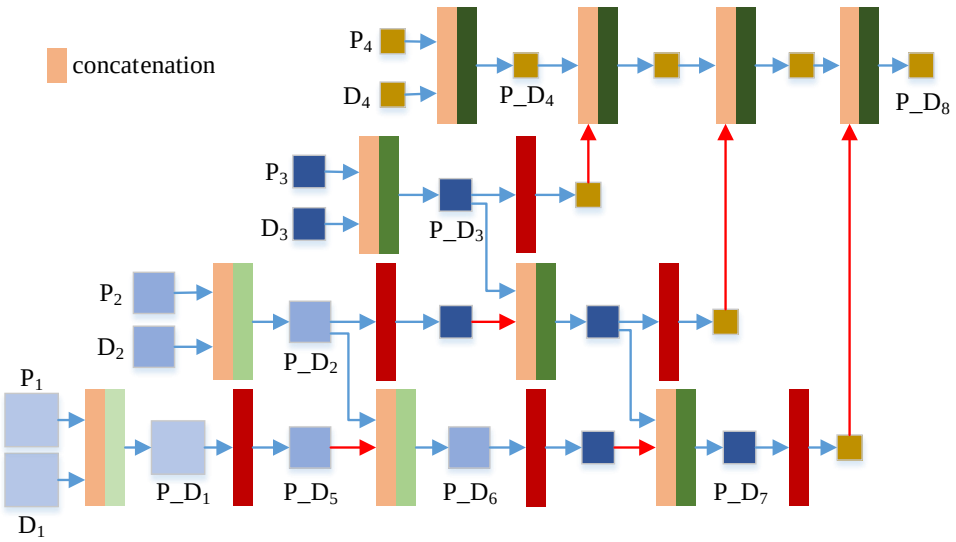

Figure 4. Structure of the cascade cross-scale detail fusion network.

The expression of the cascade cross-scale detail fusion stage is Equation (17).

$$P\_D_8 = C\_F\left(P_1, \cdots, P_n, D_1, \cdots, D_n, W_{cf}\right) \quad n = 4 \tag{17}$$

where P_D_8 shows the fused information of the cascade cross-scale detail fusion stage, C_F indicates the function of the cascade cross-scale detail fusion network, and W_{cf} means the parameter.

2.4.3. Injection Detail Reconstruction Stage

This section is the third stage of the proposed TDPNet approach generating injection detail, i.e., the reconstruction stage of injection detail. The structure is shown in Figure 5. To compensate for the lost information, we design a multiscale high-frequency detail

compensation mechanism. In addition, we take the fusion results generated in the cascade cross-scale fusion stage as a multiscale prior compensation module by multiscale skip connections. Finally, the pansharpening result is produced by adding the reconstructed injection detail to the ↑LMS image. This stage consists of three upsampling operations (i.e., deconvolution operation) and three convolutions after concatenating operations.

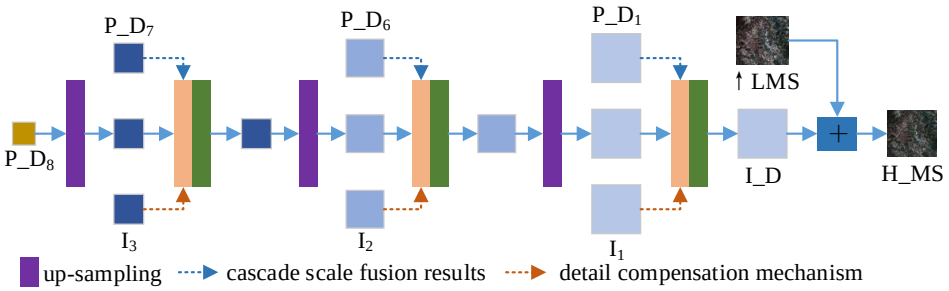

Figure 5. The structure of the high-frequency details reconstruction network.

As shown in Figure 5, considering that some details will be lost in the downsampling process, to enhance the injection details generated by reconstruction, the fusion results of the first scale P_D_1, the second scale P_D_6 and the third scale P_D_7 are introduced, forming multiscale skip connections. The multiscale skip connections not only reduce the network parameters but also compensate for the details. To further compensate for the information lost in the downsampling operations, we also design a multiscale high-frequency detail compensation mechanism, i.e., I_1–I_3 in Figure 5. To match the scale of the reconstructed information, we design a scale transfer block for the compensation details. Figure 6 describes the structure diagram of the compensation details I_1–I_3. P_L_D means the reduced form of the P_L image. I_1 is obtained from the difference between P_L and P_L_D, and then two-scale details I_2 and I_3 are obtained by the downsampling operation. In this way, the detail compensation mechanism can further compensate for the information lost in the downsampling of the fusion stage and enhance the reconstruction details.

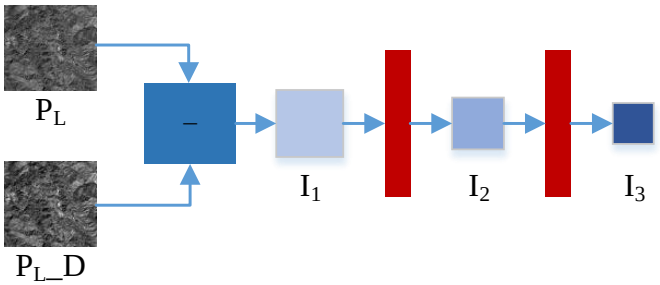

Figure 6. Structure diagram for obtaining compensated details.

As shown in Figure 5, the result P_D_8 of CCSF undergoes a deconvolution operation, and convolution is performed after concatenating P_D_7 and I_3, generating a finer scale image. After three such operations in turn, the final injection detail I_D is generated. Finally, I_D and ↑LMS images are added to obtain the pansharpening image.

The expression of obtaining the injection detail is Equation (18).

$$I\_D = d(LMS, P_L, \Theta) \tag{18}$$

where I_D presents the injection detail, d is the function of the three-stage injection detail extraction network, and Θ indicates the parameter of the network.

The expression for pansharpening model TDPNet is Equation (19).

$$H\_MS = d(LMS, P_L, \Theta) + \uparrow LMS \tag{19}$$

where H_MS indicates the generated pansharpening result.

The optimization objective of the pansharpening network TDPNet is the loss function, expressed as Equation (20).

$$L(\Theta) = \frac{1}{K}\sum_{i=1}^{K} \| \uparrow LMS + d(LMS, P_L, \Theta) - MS\|^2 \tag{20}$$

where $L(\Theta)$ is the loss function, K indicates the number of training data in each iteration, and MS represents the reference image.

3. Results

To validate the pansharpening ability of the proposed TDPNet approach , the state-of-the-art CS/MRA-based techniques PRACS [10], Indusion [13] and MTF_GLP [15] and the CNN-based methods PNN [33], DRPNN [34], PanNet [35], FusionNet [40] and RDFNet [41] are employed for comparative experiments. The experimental results of reduced resolution, full resolution and evaluation indices on three data sets are as follows.

3.1. Experimental Results on GF-1 Data Set

This section describes the reduced and full resolution experiments on the GF-1 data set. Figure 7 shows the experimental results at reduced resolution of various comparison methods. Figure 7a shows a reduced resolution PAN (LPAN) image, and Figure 7b shows an upsampled reduced resolution MS image obtained by a polynomial kernel with 23 coefficients (EXP) [49]. The pansharpened results of the PRACS, Indusion, MTF_GLP, PNN, DRPNN, PanNet, FusionNet, RDFNet and TDPNet approaches are shown in Figure 7c–k, respectively. Figure 7l shows the reference MS image, i.e., the GT image. To conveniently observe the differences between various methods, we enlarge the red box area. To observe the pansharpening performance more conveniently, we also show the average intensity difference map and the average spectral difference map (calculated according to SAM) between the pansharpened result and the reference image, as shown in Figures 8 and 9. We employ color to represent the difference, and the value gradually increases from blue to yellow. To highlight the difference, the values of the color bar of the average intensity difference map and the average spectral difference map are 0–0.5 and 0–1, respectively. The top row and the third row of Figures 8 and 9 represent the difference map of the whole pansharpened result. The second row and the bottom row of Figures 8 and 9 show an enlarged view of the corresponding region in the red box of Figure 7. Table 1 shows the quantitative evaluation indices of the experiment at reduced resolution.

From Figure 7, it is clearly found that the Indusion and MTF_GLP methods exhibit severe spectral distortion. The pansharpening results of the PRACS, Indusion, MTF_GLP and PanNet methods are relatively blurred. The pansharpening results of the PNN, DRPNN and FusionNet are also slightly vague compared to the GT image. RDFNet and the proposed method TDPNet have better pansharpening results. However, combined with Figures 8 and 9, we find that the difference between TDPNet and GT is smaller, and the pansharpening result is better. In Table 1, the objective evaluation indices Q4, UIQI, SCC, SSIM and SAM of the proposed method are better. Although the ERGAS of the proposed method ranks second, it is only 0.0006 from the optimal value. The TDPNet is better in preserving spectrum information and details.

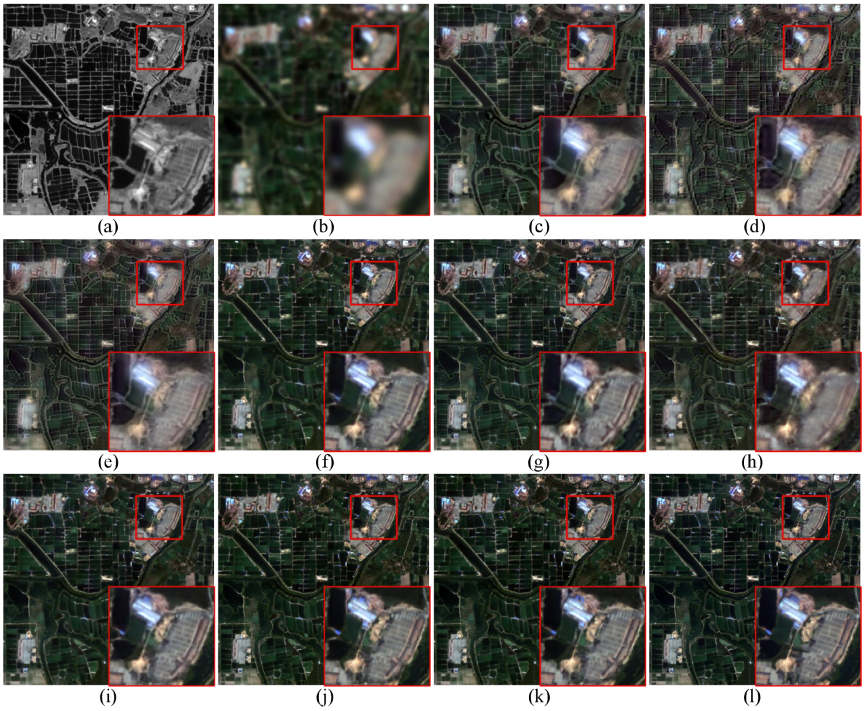

Figure 7. The experimental results of various compared methods on GF-1 testing data at reduced resolution. (**a**) LPAN. (**b**) EXP. (**c**) PRACS. (**d**) Indusion. (**e**) MTF_GLP. (**f**) PNN. (**g**) DRPNN. (**h**) PanNet. (**i**) FusionNet. (**j**) RDFNet. (**k**) TDPNet. (**l**) Ground Truth.

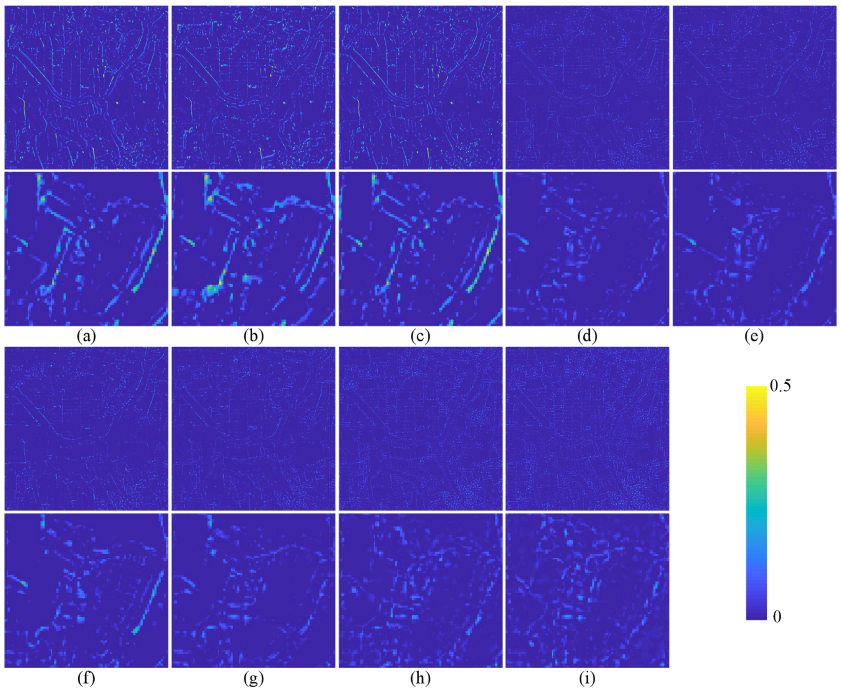

Figure 8. The average intensity difference map between the generated image and GT on GF-1 testing data. The top row and the third row represent the difference map of the whole pansharpened result. The second row and the bottom row show the enlarged view of the corresponding region in the red box of Figure 7. (**a**) PRACS. (**b**) Indusion. (**c**) MTF_GLP. (**d**) PNN. (**e**) DRPNN. (**f**) PanNet. (**g**) FusionNet. (**h**) RDFNet. (**i**) TDPNet.

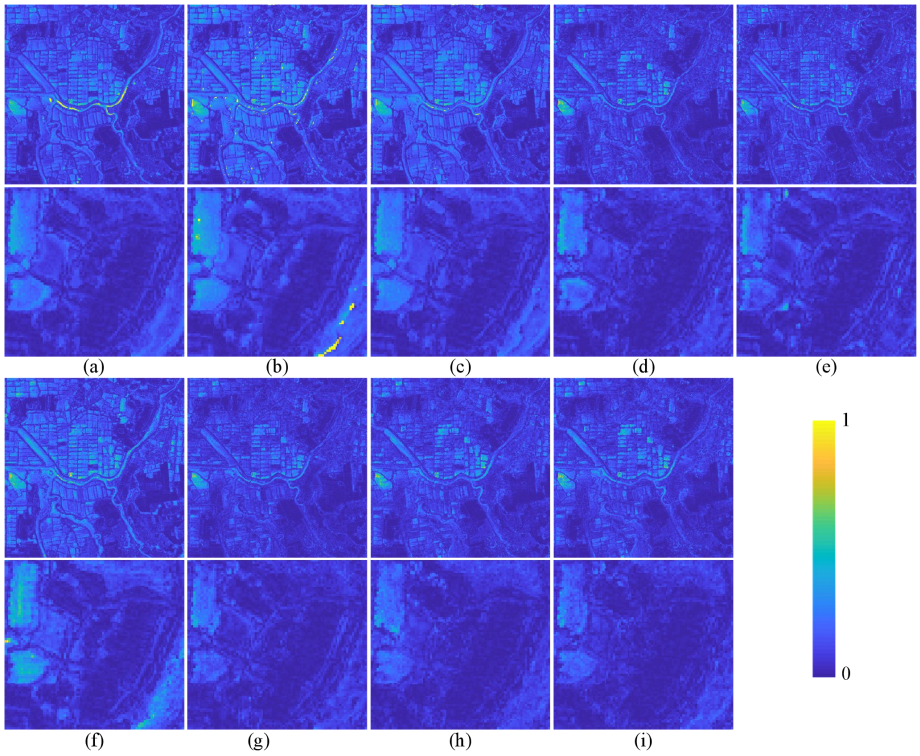

Figure 9. The average spectral difference map between the fused image and GT on GF-1 testing data. The top row and the third row represent the difference map of the whole pansharpened result. The second row and the bottom row show the enlarged view of the corresponding region in the red box of Figure 7. (**a**) PRACS. (**b**) Indusion. (**c**) MTF_GLP. (**d**) PNN. (**e**) DRPNN. (**f**) PanNet. (**g**) FusionNet. (**h**) RDFNet. (**i**) TDPNet.

Table 1. Quantitative evaluation of various compared methods on GF-1 testing data at reduced resolution.

	Q4	*UIQI*	*SCC*	*SSIM*	*SAM*	*ERGAS*
EXP	0.6220	0.6313	0.5894	0.4517	4.8390	3.9460
PRACS	0.6018	0.6175	0.6455	0.4191	4.9552	4.5335
Indusion	0.5938	0.6092	0.5872	0.3738	5.1425	4.4969
MTF_GLP	0.6122	0.6262	0.6433	0.3948	4.7667	4.2926
PNN	0.9396	0.9391	0.9388	0.9420	1.9413	1.5371
DRPNN	0.9160	0.9157	0.9061	0.9218	2.4796	1.8033
PanNet	0.8512	0.8846	0.8594	0.8873	3.6488	2.6688
FusionNet	0.9454	0.9463	0.9477	0.9495	1.7443	1.4278
RDFNet	0.9523	0.9545	0.9526	0.9533	1.5625	**1.3292**
TDPNet	**0.9530**	**0.9554**	**0.9536**	**0.9539**	**1.5492**	1.3298
Ideal value	1	1	1	1	0	0

Figure 10 presents the experimental results at full resolution of each method. Figure 10a indicates the PAN image at full resolution, Figure 10b means the corresponding upsampled MS image, and the pansharpened results of the PRACS, Indusion, MTF_GLP, PNN, DRPNN, PanNet, FusionNet, RDFNet and TDPNet approaches are shown in Figure 10c–k, respectively. Table 2 presents the quantitative evaluation metrics of experimental results at full resolution. In the quantitative evaluation indices of the tables, black bold formatting indicates the best result.

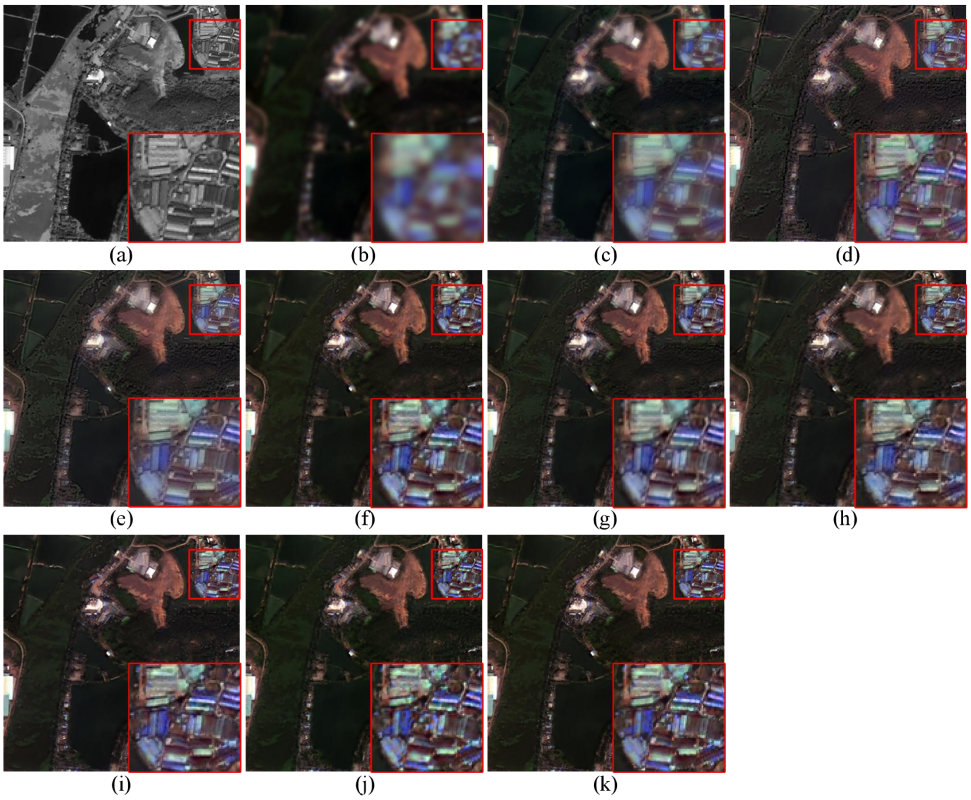

Figure 10. The experimental results of various compared approaches of GF-1 testing image at full resolution. (**a**) PAN. (**b**) EXP. (**c**) PRACS. (**d**) Indusion. (**e**) MTF_GLP. (**f**) PNN. (**g**) DRPNN. (**h**) PanNet. (**i**) FusionNet. (**j**) RDFNet. (**k**) TDPNet.

Table 2. Quantitative evaluation of various compared approaches on GF-1 testing data at full resolution.

	D_λ	D_S	QNR
EXP	0.0000	0.1775	0.8225
PRACS	0.0958	**0.1150**	0.8003
Indusion	0.1762	0.1233	0.7223
MTF_GLP	0.2226	0.2293	0.5991
PNN	0.0130	0.1450	0.8438
DRPNN	0.0536	0.1433	0.8109
PanNet	0.0296	0.1479	0.8268
FusionNet	0.0234	0.1570	0.8233
RDFNet	0.0212	0.1467	0.8352
TDPNet	**0.0108**	0.1432	**0.8475**
Ideal Value	0	0	1

From Figure 10, we can clearly find that the spectral distortions of PRACS, Indusion and MTF_GLP are relatively severe and that the MTF_GLP result is relatively fuzzy. Compared with the pansharpening results of other methods, the result of TDPNet is clearer and retains more spectral information. From Table 2, the objective evaluation indicators indicate that the D_S of TDPNet is not the best; although PRACS and Indusion retain more details, their spectral distortions are more severe. The proposed method TDPNet has the best

retention of spectral information, and the value of QNR is also optimal. Comprehensively, the pansharpening result of TDPNet is better than those of the other approaches.

3.2. Experimental Results on QuickBird Data Set

This section describes the experiment on the QuickBird data set. Figure 11 shows the experimental outcomes of the compared approaches at reduced resolution. Figures 12 and 13 show the average intensity difference map and the average spectral difference map, respectively. Figure 14 shows the full resolution experimental results of the compared methods. Table 3 describes the quantitative evaluation indices of the experimental results at reduced resolution of the compared approaches. Table 4 presents the quantitative evaluation indices of the full resolution experimental results.

The pansharpening result of the DRPNN method in Figure 11 contain the most severe spectral distortion. Artifacts appear in the pansharpening result of the Indusion method. Combined with Figures 12 and 13, the result shows that the proposed method retains more details while retaining spectral information. As shown in Table 3, the evaluation indices Q4 and SCC of TDPNet rank second. For Q4, the difference between TDPNet and RDFNet is very small. The proposed method TDPNet is the best for indicators UIQI, SCC, SSIM, SAM and ERGAS. The spectral distortion of the PNN and DRPNN in Figure 14 is severe. The result of the Indusion method exhibits artifacts and is relatively fuzzy. From Table 4, although the indices D_λ and D_S of the proposed method TDPNet are both ranked second, TDPNet is the best in terms of retaining both spectral information and details simultaneously.

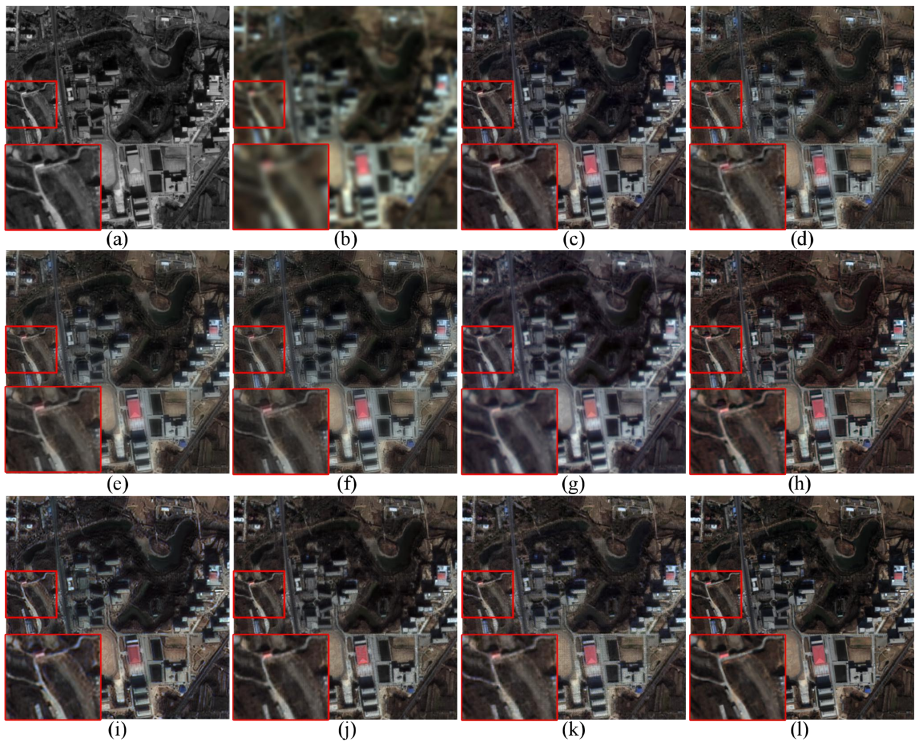

Figure 11. The experimental results of various compared approaches on QuickBird testing data at reduced resolution. (**a**) LPAN. (**b**) EXP. (**c**) PRACS. (**d**) Indusion. (**e**) MTF_GLP. (**f**) PNN. (**g**) DRPNN. (**h**) PanNet. (**i**) FusionNet. (**j**) RDFNet. (**k**) TDPNet. (**l**) Ground Truth.

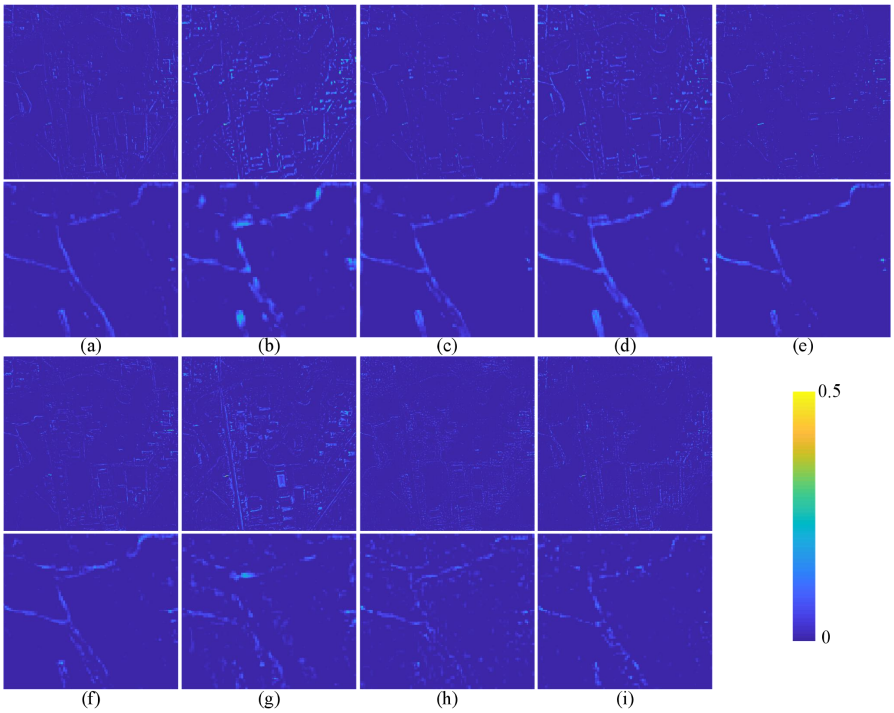

Figure 12. The average intensity difference map between the fused MS image and GT on QuickBird testing data. The top row and the third row represent the difference map of the whole pansharpened result. The second row and the bottom row show the enlarged view of the corresponding region in the red box of Figure 11. (**a**) PRACS. (**b**) Indusion. (**c**) MTF_GLP. (**d**) PNN. (**e**) DRPNN. (**f**) PanNet. (**g**) FusionNet. (**h**) RDFNet. (**i**) TDPNet.

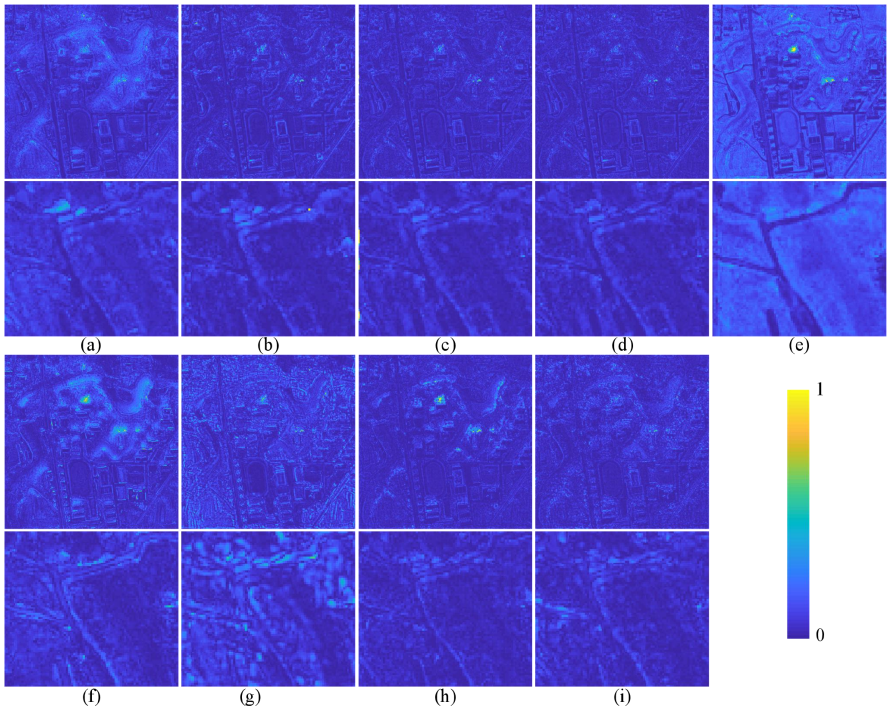

Figure 13. The average spectral difference map between the fused image and GT on QuickBird testing data. The top row and the third row represent the difference map of the whole pansharpened result. The second row and the bottom row show the enlarged view of the corresponding region in the red box of Figure 11. (**a**) PRACS. (**b**) Indusion. (**c**) MTF_GLP. (**d**) PNN. (**e**) DRPNN. (**f**) PanNet. (**g**) FusionNet. (**h**) RDFNet. (**i**) TDPNet.

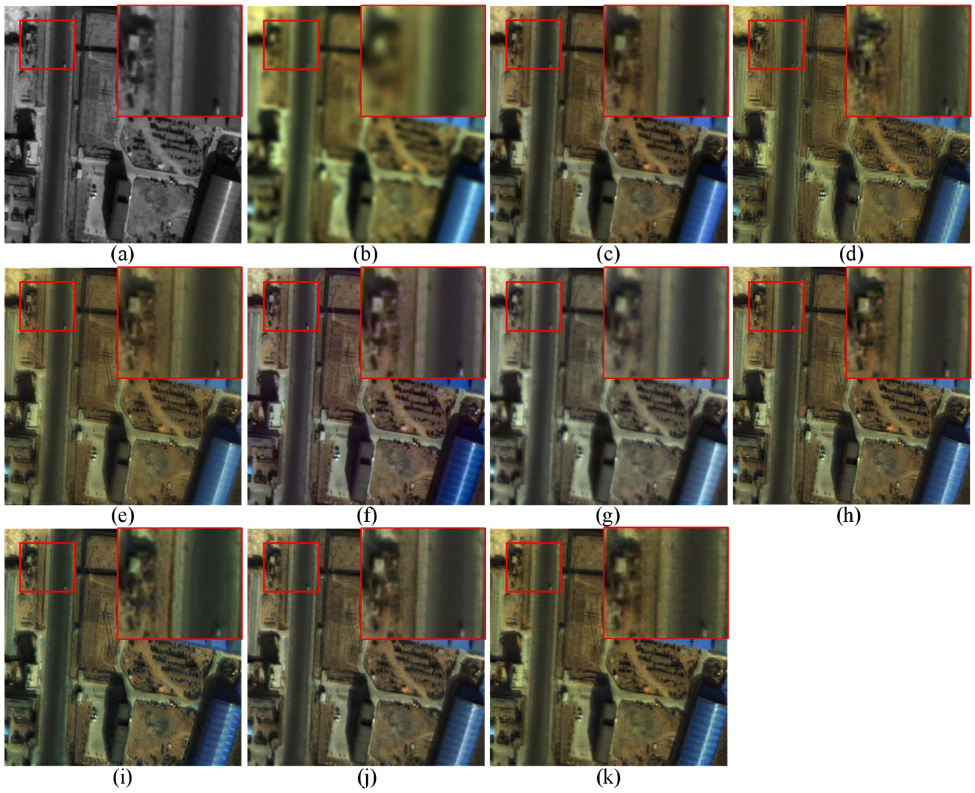

Figure 14. The experimental results of various compared approaches on the QuickBird testing data at full resolution. (**a**) PAN. (**b**) EXP. (**c**) PRACS. (**d**) Indusion. (**e**) MTF_GLP. (**f**) PNN. (**g**) DRPNN. (**h**) PanNet. (**i**) FusionNet. (**j**) RDFNet. (**k**) TDPNet.

Table 3. Quantitative evaluation of various compared approaches on QuickBird testing data at reduced resolution.

	Q4	*UIQI*	*SCC*	*SSIM*	*SAM*	*ERGAS*
EXP	0.7327	0.7404	0.6621	0.4304	2.0954	3.1671
PRACS	0.9316	0.9469	0.9142	0.8740	1.7117	1.6558
Indusion	0.8457	0.8516	0.8790	0.7128	2.0692	2.5741
MTF_GLP	0.9280	0.9299	**0.9363**	0.8428	1.6645	1.7722
PNN	0.8912	0.8933	0.9108	0.7834	1.8165	2.1636
DRPNN	0.6226	0.6744	0.8714	0.7429	3.5481	6.4267
PanNet	0.9331	0.9448	0.8864	0.9482	2.3250	1.8568
FusionNet	0.8356	0.8977	0.8615	0.9166	2.8616	3.1899
RDFNet	**0.9352**	0.9492	0.8860	0.9468	1.8177	1.6423
TDPNet	0.9344	**0.9504**	0.9169	**0.9523**	**1.5098**	**1.6420**
Ideal value	1	1	1	1	0	0

Table 4. Quantitative evaluation of various compared approaches on QuickBird testing data at full resolution.

	D_λ	D_S	*QNR*
EXP	0.0000	0.1660	0.8340
PRACS	0.0349	0.0724	0.8952
Indusion	**0.0273**	0.1476	0.8291
MTF_GLP	0.0883	0.0657	0.8518
PNN	0.0825	0.1552	0.7751
DRPNN	0.1051	0.1287	0.7797
PanNet	0.0620	**0.0144**	0.9245
FusionNet	0.0335	0.0496	0.9185
RDFNet	0.0571	0.0212	0.9228
TDPNet	0.0322	0.0212	**0.9472**
Ideal Value	0	0	1

3.3. Experimental Results on GF-2 Data Set

This section describes the experiment on the GF-2 data set. Figure 15 shows the experimental results of the compared methods at reduced resolution. The average intensity difference map and the average spectral difference map are shown in Figures 16 and 17. Figure 18 shows the full resolution experimental results of the compared methods. The quantitative evaluation indices of the experimental results at reduced resolution for the compared methods are shown in Table 5. The quantitative evaluation indices of the full resolution experimental results of each method are shown in Table 6.

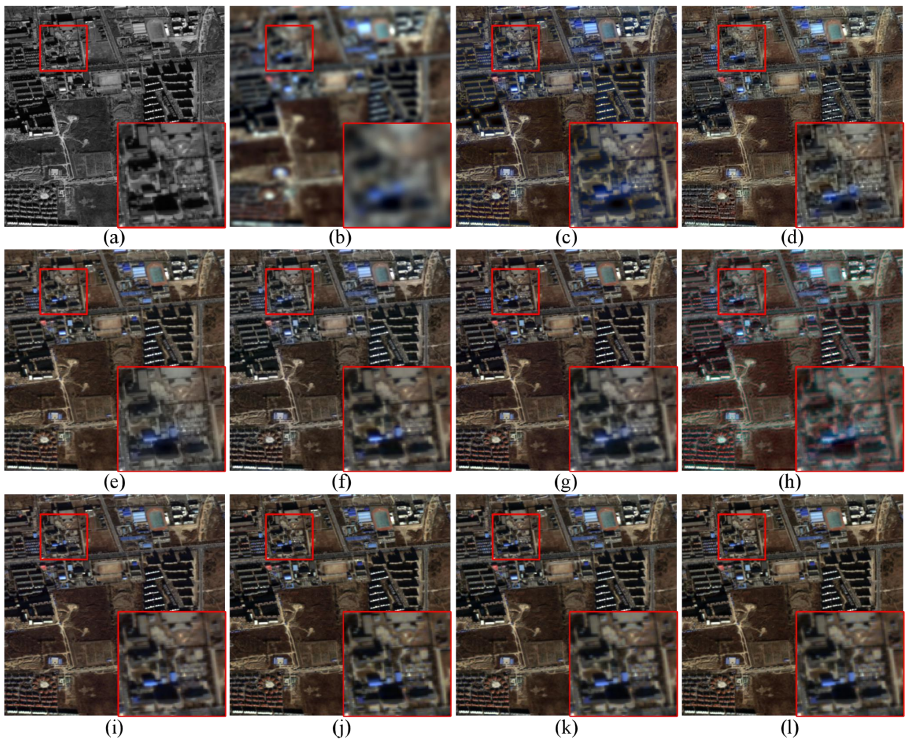

Figure 15. The experimental results of various compared approaches on GF-2 testing data at reduced resolution. (**a**) LPAN. (**b**) EXP. (**c**) PRACS. (**d**) Indusion. (**e**) MTF_GLP. (**f**) PNN. (**g**) DRPNN. (**h**) PanNet. (**i**) FusionNet. (**j**) RDFNet. (**k**) TDPNet. (**l**) Ground Truth.

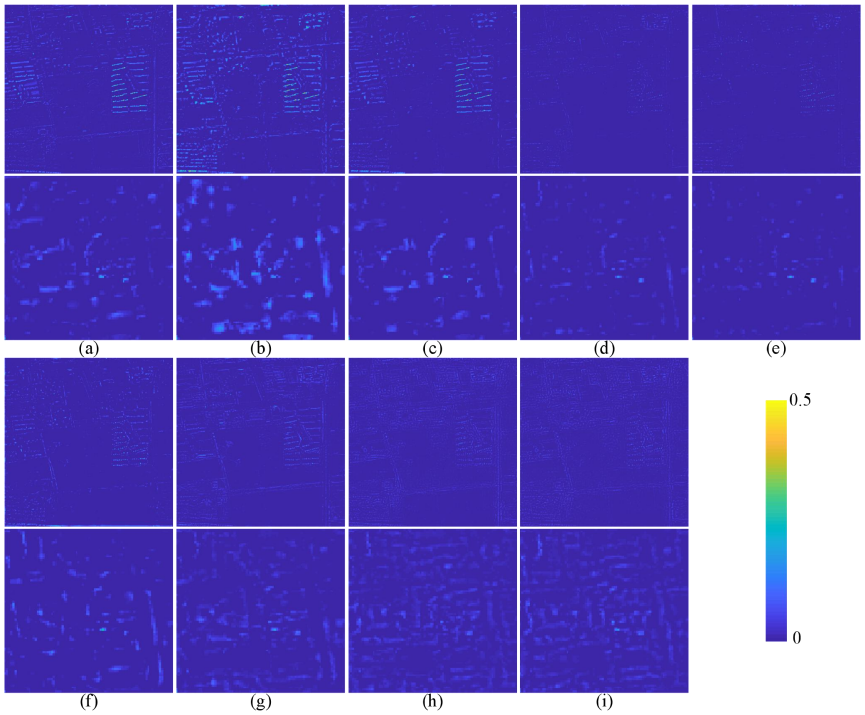

Figure 16. The average intensity difference map between the fused MS image and GT on GF-2 testing data. The top row and the third row represent the difference map of the whole pansharpened result. The second row and the bottom row show the enlarged view of the corresponding region in the red box of Figure 15. (**a**) PRACS. (**b**) Indusion. (**c**) MTF_GLP. (**d**) PNN. (**e**) DRPNN. (**f**) PanNet. (**g**) FusionNet. (**h**) RDFNet. (**i**) TDPNet.

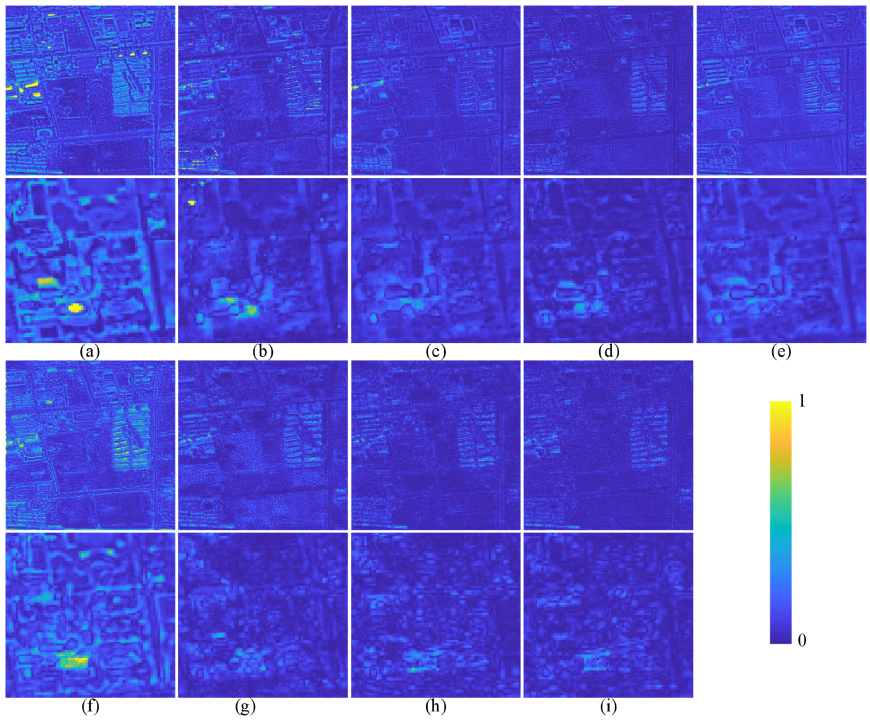

Figure 17. The average spectral difference map between the fused image and GT on GF-2 testing data. The top row and the third row represent the difference map of the whole pansharpened result. The second row and the bottom row show the enlarged view of the corresponding region in the red box of Figure 15. (**a**) PRACS. (**b**) Indusion. (**c**) MTF_GLP. (**d**) PNN. (**e**) DRPNN. (**f**) PanNet. (**g**) FusionNet. (**h**) RDFNet. (**i**) TDPNet.

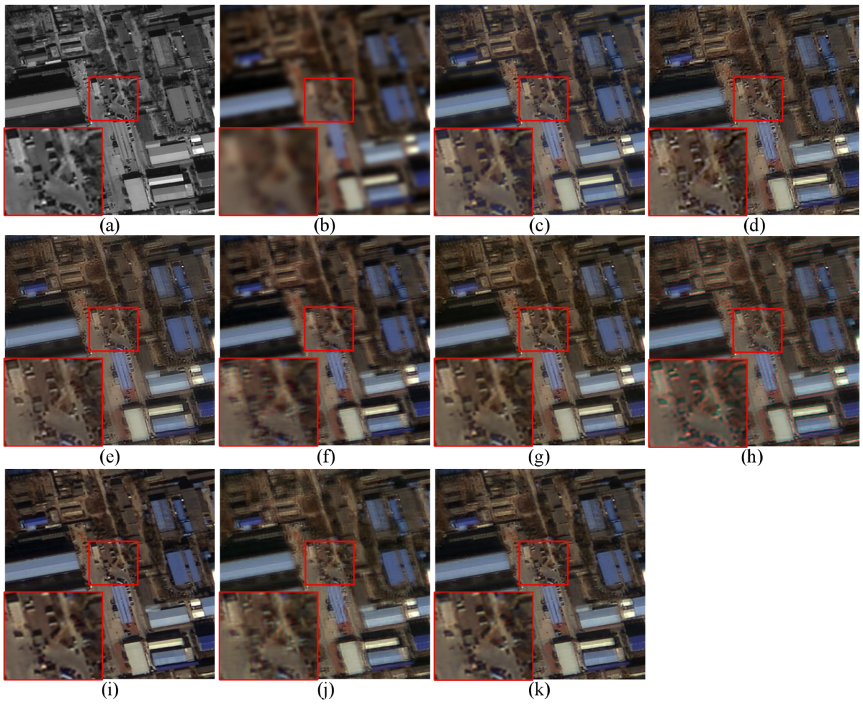

Figure 18. The experimental results of various compared approaches on the GF-2 testing data at full resolution. (**a**) PAN. (**b**) EXP. (**c**) PRACS. (**d**) Indusion. (**e**) MTF_GLP. (**f**) PNN. (**g**) DRPNN. (**h**) PanNet. (**i**) FusionNet. (**j**) RDFNet. (**k**) TDPNet.

Table 5. Quantitative evaluation of various compared approaches on GF-2 testing data at reduced resolution.

	Q4	*UIQI*	*SCC*	*SSIM*	*SAM*	*ERGAS*
EXP	0.6697	0.6756	0.5960	0.5150	2.0718	3.6576
PRACS	0.9254	0.9107	0.8843	0.8675	1.5629	1.8503
Indusion	0.8221	0.8293	0.8471	0.7677	1.9478	2.8731
MTF_GLP	0.9125	0.9122	0.9132	0.8710	1.5096	2.0406
PNN	0.9795	0.9820	0.9740	0.9833	**0.8900**	0.8749
DRPNN	0.9514	0.9758	0.9672	0.9801	1.1417	1.2794
PanNet	0.7047	0.8814	0.8586	0.9036	7.0690	5.8844
FusionNet	0.9696	0.9741	0.9710	0.9799	1.1054	1.0169
RDFNet	0.9807	0.9826	0.9773	0.9838	0.9451	0.8211
TDPNet	**0.9835**	**0.9850**	**0.9803**	**0.9856**	0.9157	**0.7855**
Ideal value	1	1	1	1	0	0

The pansharpening results of PRACS and PanNet in Figure 15 exhibit the most severe spectrum distortion. The pansharpening result of Indusion is blurry. Combined with Figures 16 and 17, the results show that the PNN, RDFNet and TDPNet methods retain more spectral information, whereas the TDPNet method retains more details. The objective indicators in Table 5, except SAM, indicate that the results of TDPNet are optimal. From the visual effect of Figure 18, the spectral distortion of PanNet is the most severe, and the results of the PNN and RDFNet methods are blurred. Although the index D_S of Indusion in Table 6 is the best, the spectral distortion is severe. Although the index D_λ of PNN is the best, more spatial details are lost. Combined with the overall spectral distortion and the

retention of spatial structure information, the pansharpening performance of the proposed method TDPNet is better than that of the other approaches.

Table 6. Quantitative evaluation of various compared methods on GF-2 testing data at full resolution.

	D_λ	D_S	QNR
EXP	0.0000	0.2558	0.7442
PRACS	0.0692	0.0950	0.8423
Indusion	0.0487	**0.0076**	0.9441
MTF_GLP	0.0700	0.0493	0.8841
PNN	**0.0193**	0.1059	0.8768
DRPNN	0.0472	0.0363	0.9182
PanNet	0.0996	0.0784	0.8298
FusionNet	0.0485	0.1037	0.8529
RDFNet	0.0205	0.1232	0.8589
TDPNet	0.0199	0.0123	**0.9680**
Ideal Value	0	0	1

3.4. Implementation Time

Table 7 shows the implementation time of pansharpening for the compared approaches at full resolution. The sizes of the MS and PAN images are $100 \times 100 \times 4$ and $400 \times 400 \times 1$, respectively. Note that the approaches based on CNN are realized on the GPU, and the approaches based on CS and MRA are realized on the CPU. Although the implementation time is not the shortest, it is acceptable for the CNN-based method.

Table 7. Implement time (seconds) of pansharpening for compared methods.

	GF-1	**QuikBird**	**GF-2**
PRACS	0.3208	0.3059	0.3100
Indusion	0.1536	0.1307	0.1351
MTF_GLP	0.2389	0.1908	0.2250
PNN	0.0195	0.0195	0.0196
DRPNN	0.1098	0.1028	0.1055
PanNet	0.0386	0.0406	0.0414
FusionNet	0.0483	0.0380	0.0466
RDFNet	0.3850	0.3747	0.4179
TDPNet	0.3484	0.3988	0.4069

4. Discussion

Based on the aforementioned experimental results, it is evidently found that the proposed TDPNet well trade off the contents of spectral and spatial. This section discusses the major compositions of TDPNet, the number of iterations and the consumed time in the training and testing processes.

4.1. Major Compositions of TDPNet

We discuss impacts of compositions of the three stages of the proposed TDPNet on the pansharpening performance from three main contents. The following discussion is based on the experimental results of the GF-2 data.

First, the dual-branch multiscale high-frequency detail extraction stage comprises two branches to extract the features of $P_L^N - \uparrow LMS$ and P_L images respectively. Verifying the effectiveness of features extracted from $P_L^N - \uparrow LMS$ image, we compare the performance of the dual-branch with $\uparrow LMS$ and P_L images (i.e., TDPNet-n-p) with that of the dual-branch with $P_L^N - \uparrow LMS$ and P_L images (TDPNet). TDPNet-n-p extracts the four-scale information from $\uparrow LMS$ and P_L respectively, and the other parts are the same as TDPNet. The compared results are shown in Figure 19. We clearly observe that the pansharpening performance of TDPNet is better than that of TDPNet-n-p in terms of preserving spectral and spatial information.

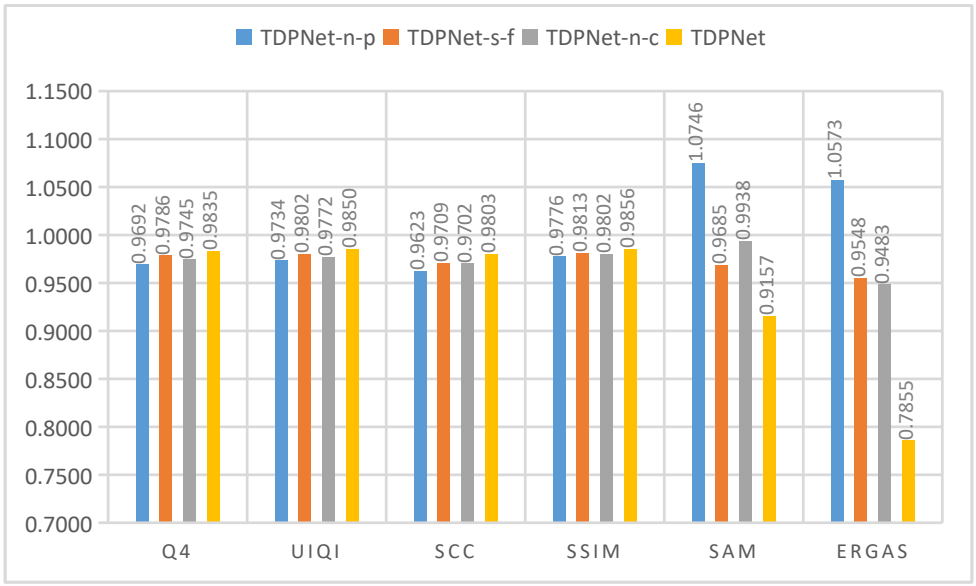

Figure 19. The effect of the components of TDPNet on the pansharpening results of the GF-2 data.

Second, we use a basic fusion block (i.e., Figure 20, the contents are the same as that in Figure 4.) to replace the CCSF in the second stage, and the other parts are the same as TDPNet, which is called TDPNet-s-f. In Figure 19, it is found that the performance of TDPNet is significantly better than that of TDPNet-s-f, and cross scale fusion in CCSF enhances spatial details and spectrum information.

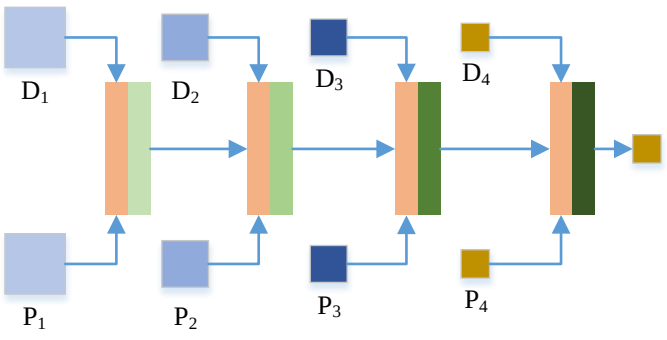

Figure 20. A basic fusion block.

Third, we remove the multiscale high-frequency detail compensation mechanism of the third stage, and the other parts of TDPNet remain the same, which is termed TDPNet-n-c. In Figure 19, it is evidently observed that the result of TDPNet is finer than TDPNet-n-c. From Figure 19, the main parts of the three stages promote the pansharpening performance of TDPNet, enhancing spatial details and reducing spectral distortion.

4.2. Iterations and Consumed Time

The number of iterations of all the CNN-based approaches is obtained on the same data set. The number of iterations and consumed time for each approach achieving the optimal performance are demonstrated in Figures 21 and 22. Note that the time spent on PanNet contains the time used to extract high-frequency details. In Figure 22, FusionNet [40] takes the least time reaching convergence because of the simple composition of it. Although the times of TDPNet reaching convergence is less, it spend longer time because of the relatively complex compositions of the network. In Figures 21 and 22, compared with other CNN-based approaches, the time spent on TDPNet is acceptable. In the next research, we can further simplify the network structure by employing group convolutions and separable convolutions and design a lightweight pansharpening network, which takes less time while achieving the same or higher performance.

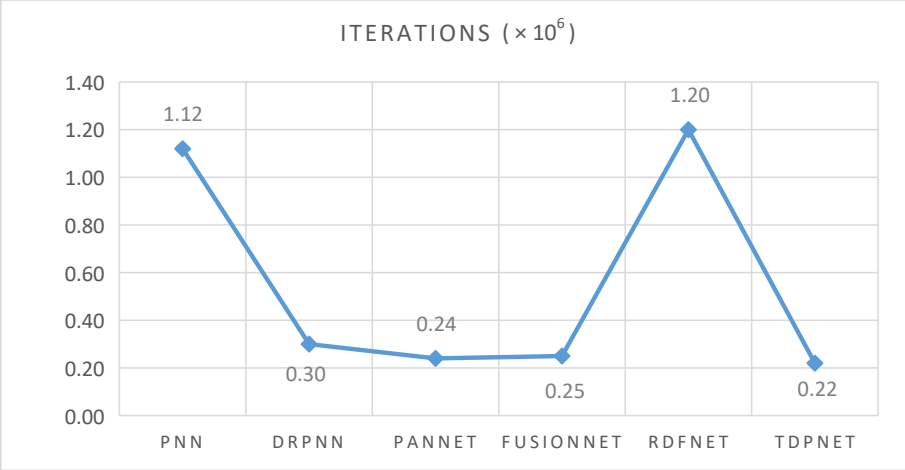

Figure 21. The number of iterations of comparative approaches based on CNN.

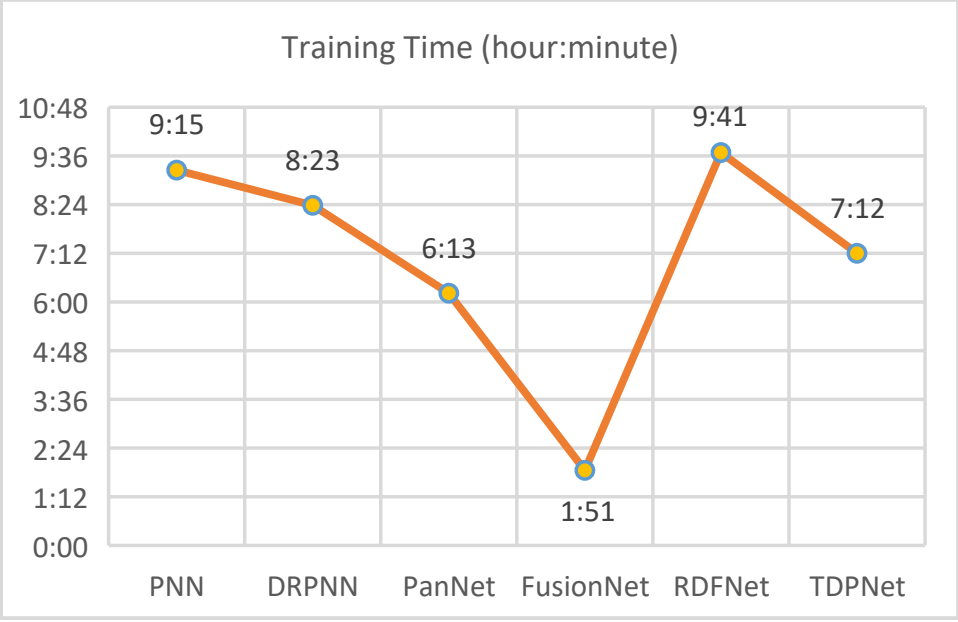

Figure 22. Training time of comparative approaches based on CNN.

5. Conclusions

To improve the preservation of spatio-spectral information, we proposed a novel three stages detail injection network for remote sensing images pansharpening to reconstruct details. A dual-branch multiscale feature extraction block obtains four scale details of PAN image and the difference between duplicated PAN and MS images, respectively. This utilizes the abundant spatial details of the PAN images and retains the spectral information by introducing MS image. CCSF employs the fine-scale fusion information as prior knowledge for the coarse-scale fusion information compensating for the lost information during downsampling and retaining high-frequency details. The multiscale detail compensation mechanism and multiscale skip connection block compensate for the lost information and reduce parameters during reconstructing injection detail. Extensive experimental analysis on GF-1, QuickBird and GF-2 data sets both at degraded and full resolutions testify that TDPNet trades off the spectral information and spatial details and improves the fidelity of sharper MS images. The quantitative evaluation and subjective evaluation results indicate that TDPNet has better ability in retaining spectrum information and spatial details than the other compared approaches. Nevertheless, the implementation time of proposed TDPNet is not the least.

In the next research, we can further simplify the network structure and design a lightweight pansharpening network, which takes less time while achieving the same or higher performance. Besides, designing a lightweight pansharpening network with the same or better performance has the following advantages: (1) There is less requirement for the server during training. (2) Fewer parameters. (3) It is more suitable for implementing on devices with limited memory. Our future work also needs to optimize the network structure and design a lightweight pansharpening network by group convolutions and separable convolutions to decrease the implementation time and further boost the pansharpening ability.

Author Contributions: Conceptualization, Y.W.; methodology, Y.W.; software, Y.W.; validation, Y.W.; formal analysis, Y.W., C.L., H.Z. and M.H.; investigation, H.Z., C.L. and M.H.; resources, M.H. and S.F.; data curation, M.H. and S.F.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W., H.Z., C.L., M.H. and S.F.; visualization, Y.W.; supervision, M.H. and S.F.; project administration, M.H.; funding acquisition, M.H. and S.F. All authors have read and agreed to the published version of the manuscript.

Funding: The work was encouraged by Hainan Provincial Natural Science Foundation of China under Grant 2019CXTD400, and the National Key Research and Development Program of China under Grant 2018YFB1404400.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data provided in this study can be provided at the request of the corresponding author. The data has not been made public because it is still used for further research in the study field.

Acknowledgments: Thanks for PNN codes in [33] supplied by Giuseppe Scarpa. Thanks for Wei and Yuan providing DRPNN codes in [34]. Thanks for PanNet codes in [35] provided by Yang, J. Thanks to Deng L. J. in [40] for providing FusionNet codes.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MS	multispectral
TDPNet	three-stage detail injection pansharpening network
PAN	panchromatic
CCSF	cascade cross-scale fusion
RSIs	Remote sensing images
LRMS	low spatial and multispectral resolutions
HS	hyperspectral
HRMS	high spatial resolution MS
HRHS	high spatial resolution HS
pansharpening	panchromatic sharpening
DL	deep learning
CS	component substitution
MRA	multiresolution analysis
VO	variational optimization
IHS	intensity-hue-saturation
AIHS	adaptive IHS
GIHS	generalized IHS
GS	Gram-Schmidt
GSA	GS adaptive
PRACS	partial replacement adaptive component substitution
DWT	discrete wavelet transform
Indusion	fusion for MS and PAN images employing the Induction scaling approach
GLP	generalized Laplacian pyramid
MTF_GLP	modulation transfer function-GLP
ATWT	à trous wavelet transform
CNN	convolutional neural networks
GF-1	Gaofen-1
GF-2	Gaofen-2
LMS	the reduced resolution form of the MS image
P_L	the reduced resolution form of the PAN image
UIQI	universal image quality index
Q4	UIQI extended to 4-band
SCC	structural correlation coefficient
SSIM	structural similarity
SAM	spectral angle mapping
ERGAS	erreur relative global adimensionnelle de Synthése
QNR	quality with no-reference
GT	ground truth
$\uparrow LMS$	an upsampled LMS image
P_L^N	duplicated N channels of the P_L image
H_MS	the generated pansharpening image
CHFDs	composite high-frequency details

References

1. Yilmaz, C.S.; Yilmaz, V.; Gungor, O. A theoretical and practical survey of image fusion methods for multispectral pansharpening. *Inf. Fusion* **2022**, *79*, 1–43. [CrossRef]
2. Javan, F.D.; Samadzadegan, F.; Mehravar, S.; Toosi, A.; Khatami, R.; Stein, A. A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 101–117. [CrossRef]
3. Chen, Q.; Huang, M.; Wang, H. A Feature Discretization Method for Classification of High-Resolution Remote Sensing Images in Coastal Areas. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 8584–8598. [CrossRef]
4. Xiao, T.; Cai, Z.; Lin, C.; Chen, Q. A Shadow Capture Deep Neural Network for Underwater Forward-Looking Sonar Image Detection. *Mob. Inf. Sys.* **2021**, *2021*, 3168464. [CrossRef]
5. 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]

6. Yang, C.; Zhan, Q.; Liu, H.; Ma, R. An IHS-based pan-sharpening method for spectral fidelity improvement using ripplet transform and compressed sensing. *Sensors* **2018**, *18*, 3624. [CrossRef]
7. Zhang, L.; Zhang, J. A novel remote-sensing image fusion method based on hybrid visual saliency analysis. *Int. J. Remote Sens.* **2018**, *39*, 7942–7964. [CrossRef]
8. Hsu, C.B.; Lee, J.C.; Tu, T.M. Generalized IHS-BT framework for the pansharpening of high-resolution satellite imagery. *J. Appl. Remote Sens.* **2018**, *12*, 046008. [CrossRef]
9. Meng, X.; Xiong, Y.; Shao, F.; Shen, H.; Sun, W.; Yang, G.; Yuan, Q.; Fu, R.; Zhang, H. A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 18–52. [CrossRef]
10. Choi, J.; Yu, K.; Kim, Y. A new adaptive component-substitution-based satellite image fusion by using partial replacement. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 295–309. [CrossRef]
11. Liu, D.; Yang, F.; Wei, H.; Hu, P. Remote sensing image fusion method based on discrete wavelet and multiscale morphological transform in the IHS color space. *J. Appl. Remote Sens.* **2020**, *14*, 016518. [CrossRef]
12. Gharbia, R.; Hassanien, A.E.; El-Baz, A.H.; Elhoseny, M.; Gunasekaran, M. Multi-spectral and panchromatic image fusion approach using stationary wavelet transform and swarm flower pollination optimization for remote sensing applications. *Future Gener. Comput. Syst.* **2018**, *88*, 501–511. [CrossRef]
13. Khan, M.M.; Chanussot, J.; Condat, L.; Montanvert, A. Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 98–102. [CrossRef]
14. Vivone, G.; Marano, S.; Chanussot, J. Pansharpening: Context-Based Generalized Laplacian Pyramids by Robust Regression. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6152–6167. [CrossRef]
15. 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]
16. Shensa, M.J. The discrete wavelet transform: Wedding the à trous and Mallat algorithms. *IEEE Trans. Signal Process.* **1992**, *40*, 2464–2482. [CrossRef]
17. Dong, W.; Xiao, S.; Li, Y.; Qu, J. Hyperspectral Pansharpening Based on Intrinsic Image Decomposition and Weighted Least Squares Filter. *Remote Sens.* **2018**, *10*, 445. [CrossRef]
18. Constans, Y.; Fabre, S.; Seymour, M.; Crombez, V.; Deville, Y.; Briottet, X. Hyperspectral Pansharpening in the Reflective Domain with a Second Panchromatic Channel in the SWIR II Spectral Domain. *Remote Sens.* **2022**, *14*, 113. [CrossRef]
19. Ghaderpour, E.; Pagiatakis, S.D.; Hassan, Q.K. A Survey on Change Detection and Time Series Analysis with Applications. *Appl. Sci.* **2021**, *11*, 6141. [CrossRef]
20. Fu, X.; Lin, Z.; Huang, Y.; Ding, X. A variational pan-sharpening with local gradient constraints. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 10265–10274.
21. Tian, X.; Chen, Y.; Yang, C.; Gao, X.; Ma, J. A variational pansharpening method based on gradient sparse representation. *IEEE Signal Process. Lett.* **2020**, *27*, 1180–1184. [CrossRef]
22. Vivone, G.; Addesso, P.; Restaino, R.; Dalla Mura, M.; Chanussot, J. Pansharpening based on deconvolution for multiband filter estimation. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 540–553. [CrossRef]
23. Li, W.; Li, Y.; Hu, Q.; Zhang, L. Model-based variational pansharpening method with fast generalized intensity–hue–saturation. *J. Appl. Remote Sens.* **2019**, *13*, 036513. [CrossRef]
24. Wang, T.; Fang, F.; Li, F.; Zhang, G. High-quality Bayesian pansharpening. *IEEE Trans. Image Process.* **2019**, *28*, 227–239. [CrossRef] [PubMed]
25. Vivone, G.; Restaino, R.; Chanussot, J. A Bayesian procedure for full-resolution quality assessment of pansharpened products. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4820–4834. [CrossRef]
26. Ayas, S.; Gormus, E.T.; Ekinci, M. An efficient pan sharpening via texture based dictionary learning and sparse representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2448–2460. [CrossRef]
27. Fei, R.; Zhang, J.; Liu, J.; Du, F.; Chang, P.; Hu, J. Convolutional sparse representation of injected details for pansharpening. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1595–1599. [CrossRef]
28. Fei, R.; Zhang, J.; Liu, J.; Du, F.; Hu, J.; Chang, P.; Zhou, C.; Sun, K. Weighted manifold regularized sparse representation of featured injected details for pansharpening. *Int. J. Remote Sens.* **2021**, *42*, 4199–4223. [CrossRef]
29. Yin, H. PAN-Guided Cross-Resolution Projection for Local Adaptive Sparse Representation- Based Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4938–4950. [CrossRef]
30. Yi, C.; Zhao, Y.Q.; Chan, C.W. Spectral Super-Resolution for Multispectral Image Based on Spectral Improvement Strategy and Spatial Preservation Strategy. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9010–9024. [CrossRef]
31. Lanaras, C.; Bioucas-Dias, J.; Galliani, S.; Baltsavias, E.; Schindler, K. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 305–319. [CrossRef]
32. Gargiulo, M.; Mazza, A.; Gaetano, R.; Ruello, G.; Scarpa, G. Fast Super-Resolution of 20 m Sentinel-2 Bands Using Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 2635. [CrossRef]
33. Giuseppe, M.; Davide, C.; Luisa, V.; Giuseppe, S. Pansharpening by Convolutional Neural Networks. *Remote Sens.* **2016**, *8*, 594.
34. Wei, Y.; Yuan, Q.; Shen, H.; Zhang, L. Boosting the Accuracy of Multispectral Image Pansharpening by Learning a Deep Residual Network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1795–1799. [CrossRef]

35. 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, Venice, Italy, 22–29 October 2017; pp. 1753–1761.
36. Scarpa, G.; Vitale, S.; Cozzolino, D. Target-Adaptive CNN-Based Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5443–5457. [CrossRef]
37. Liu, X.; Wang, Y.; Liu, Q. Psgan: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpening. In Proceedings of the IEEE Transactions on Geoscience and Remote Sensing, (ICIP), Athens, Greece, 7–10 October 2018; pp. 873–877.
38. 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]
39. Zhao, Z.; Zhan, J.; Xu, S.; Sun, K.; Huang, L.; Liu, J.; Zhang, C. FGF-GAN: A Lightweight Generative Adversarial Network for Pansharpening via Fast Guided Filter. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; pp. 1–6.
40. Deng, L.J.; Vivone, G.; Jin, C.; Chanussot, J. Detail Injection-Based Deep Convolutional Neural Networks for Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 6995–7010. [CrossRef]
41. Wu, Y.; Huang, M.; Li, Y.; Feng, S.; Wu, D. A Distributed Fusion Framework of Multispectral and Panchromatic Images Based on Residual Network. *Remote Sens.* **2021**, *13*, 2556. [CrossRef]
42. 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.
43. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]
44. Vivone, G.; Mura, M.; Garzelli, A.; Pacifici, F. A Benchmarking Protocol for Pansharpening: Dataset, Pre-processing, and Quality Assessment. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6102–6118. [CrossRef]
45. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
46. 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]
47. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 1–15.
48. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 September 2016; pp. 630–645.
49. 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]