

Article

# PCDRN: Progressive Cascade Deep Residual Network for Pansharpening

Yong Yang<sup>1</sup>, Wei Tu<sup>1</sup>, Shuying Huang<sup>2,\*</sup> and Hangyuan Lu<sup>1</sup>

<sup>1</sup> School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China; greatyangy@jxufe.edu.cn (Y.Y.); weitu@stu.jxufe.edu.cn (W.T.); 2201810059@stu.jxufe.edu.cn (H.L.)

<sup>2</sup> School of Software and Communication Engineering, Jiangxi University of Finance and Economics, Nanchang 330032, China

\* Correspondence: huangshuying@jxufe.edu.cn; Tel.: +86-791-8398-3830

Received: 14 January 2020; Accepted: 17 February 2020; Published: 19 February 2020



**Abstract:** Pansharpening is the process of fusing a low-resolution multispectral (LRMS) image with a high-resolution panchromatic (PAN) image. In the process of pansharpening, the LRMS image is often directly upsampled by a scale of 4, which may result in the loss of high-frequency details in the fused high-resolution multispectral (HRMS) image. To solve this problem, we put forward a novel progressive cascade deep residual network (PCDRN) with two residual subnetworks for pansharpening. The network adjusts the size of an MS image to the size of a PAN image twice and gradually fuses the LRMS image with the PAN image in a coarse-to-fine manner. To prevent an overly-smooth phenomenon and achieve high-quality fusion results, a multitask loss function is defined to train our network. Furthermore, to eliminate checkerboard artifacts in the fusion results, we employ a resize-convolution approach instead of transposed convolution for upsampling LRMS images. Experimental results on the Pléiades and WorldView-3 datasets prove that PCDRN exhibits superior performance compared to other popular pansharpening methods in terms of quantitative and visual assessments.

**Keywords:** pansharpening; deep residual network; loss function

## 1. Introduction

Remote sensing satellites such as Pléiades, WorldView, and GeoEye provide low spatial resolution multispectral (LRMS) and high spatial resolution panchromatic (PAN) images. To obtain a fused high-resolution multispectral (HRMS) image by fusing an LRMS image with a PAN image of the same scene, pansharpening is considered a powerful image fusion technique. The HRMS image can effectively integrate the spectral characteristics of the LRMS image with the spatial information of the PAN image [1,2].

In recent decades, numerous approaches have been put forward for pansharpening. The conventional methods for pansharpening can be classified into three major categories: component substitution (CS)-based methods, multiresolution analysis (MRA)-based methods, and model-based methods. The CS-based methods primarily include the intensity-hue-saturation (IHS) method [3,4], the principal component analysis method (PCA) [5], and the Gram–Schmidt (GS) transform-based method [6]. Although CS-based methods can usually be quickly and easily implemented [7], obvious spectral distortions may be produced in the spectral domain of the fused image [1]. The MRA-based methods primarily include the Laplacian Pyramid method [8], à trous wavelet transform (ATWT) method [9], discrete wavelet transform (DWT) [10], and non-subsampled Contourlet transform (NSCT) [11]. MRA-based methods generally outperform CS-based methods in spectral preservation. However, they often lead to the problem of spatial distortion [12]. There are several

model-based methods that have been recently proposed [13–16]. Although model-based methods generally exhibit better spectral information preservation, they are characterized by high computational complexity and poor real-time performance.

Because deep learning can automatically learn additional features from various types of data, it has gained considerable attention in recent years [17–21]. Different from the conventional pansharpening methods, deep learning-based methods present more ideal solutions for improving the performance of pansharpening. Masi et al. [17] developed a pansharpening approach that is based on a three-layer convolutional neural network. PanNet was presented by Yang et al. [18], which achieved remote sensing image fusion by spectra-mapping and network parameters training. To obtain the high quality fused image, Wei et al. [19] presented a deep residual network (ResNet) for pansharpening. Shao et al. [20] developed a two-branch network that can separately obtain salient features from MS and PAN images. A multiscale and multidepth convolutional neural network (CNN) for pansharpening was introduced by Yuan et al. [12]. He et al. [21] proposed two convolutional neural network frameworks for pansharpening, i.e., DiCNN1 and DiCNN2, which can obtain a high quality fused image and converge quickly. In the above-mentioned pansharpening methods that are based on deep learning, the LRMS image is directly upsampled by a factor of 4 during the fusion process. However, this may result in loss of high-frequency details owing to the difficulty in learning nonlinear feature mapping. In addition, the mean squared error (MSE) loss function is often employed by many pansharpening methods through deep learning. However, it is difficult for MSE to capture the differences in high-frequency details between the fused image and the reference image. Consequently, this leads to the phenomenon of excessive smoothing and loss of high-quality details.

To overcome the above-mentioned limitations, a new progressive cascade deep residual network (PCDRN) for pansharpening was presented, which includes two residual subnetworks. Different from direct upsampling by a factor of 4 during the fusion process in other pansharpening methods, we first adopt two upsampling operations and then employ the two residual subnetworks to learn the nonlinear feature mapping from the source images to the ground truth in two scales. We finally realize the fusion of LRMS and PAN images in a step by step manner. To better train our network, a multitask loss function is designed to enable PCDRN to extract more precise features. Furthermore, to eliminate checkerboard artifacts in the fused image, the transposed convolution is replaced by the resize-convolution to upsample the LRMS image. Compared with several existing pansharpening methods, the experimental results demonstrate that PCDRN achieves better performance.

## 2. Related Work

### 2.1. Residual Network

He et al. [22] proposed a residual learning network architecture, which is substantially deeper than the plain network. Figure 1 illustrates the structure of a residual block. It not only makes the network deeper but also overcomes the vanishing gradient problem of the plain network.

Formally, the residual block is represented as:

$$y = F(x, \{W_i\}) + x \quad (1)$$

where  $x$  and  $y$  are the input and output vectors of the layers considered, respectively. The function  $F(x, \{W_i\})$  denotes the residual mapping to be learned. As shown in Figure 1,  $F(x)$  is defined as:

$$F(x) = W_2 \otimes R(W_1 \otimes x) \quad (2)$$

where  $W_1$  denotes the weight of the first layer,  $W_2$  denotes the weight of the second layer,  $\otimes$  denotes convolution,  $R$  denotes the activation function Relu, and  $R(x) = \max(0, x)$ .

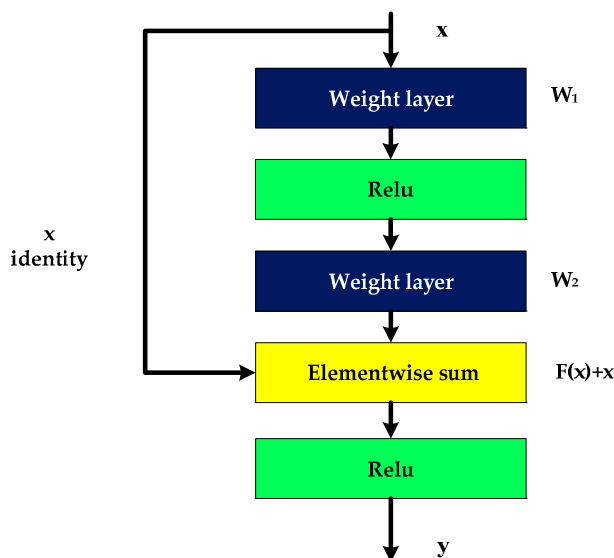


Figure 1. A residual block.

### 2.2. Universal Image Quality Index

Wang et al. [23] presented a universal image quality index (UIQI), which is used to measure the structure distortion degree. It is composed of three factors: loss of correlation, luminance distortion, and contrast distortion.

The quality metrics is defined as:

$$UIQI(F, R) = \frac{\sigma_{FR}}{\sigma_F \sigma_R} \cdot \frac{2\mu_F \mu_R}{\mu_F^2 + \mu_R^2} \cdot \frac{2\sigma_F \sigma_R}{\sigma_F^2 + \sigma_R^2} \tag{3}$$

where  $\sigma_F$  and  $\sigma_R$  denote the standard deviation of the fused image  $F$  and referenced image  $R$ , respectively.  $\sigma_{FR}$  denotes the covariance of the fused image  $F$  and referenced image  $R$ .  $\mu_F$  and  $\mu_R$  denote the mean of the fused image  $F$  and referenced image  $R$ , respectively. The optimum value of UIQI is 1.

### 3. Proposed Method

In this section, we put forward a new PCDRN for pansharpening, which exploits the two residual subnetworks called ResNet to extract accurate features and progressively inject the details of a PAN image into an MS image in a coarse-to-fine manner. A multitask loss function is proposed to prevent the over-smooth phenomenon from preserving spatial information. Furthermore, to address the problem of checkerboard artifacts in the fusion results, resize-convolution is adopted rather than transposed convolution in the upsampling process of LRMS images.

#### 3.1. Flowchart of PCDRN

To inject additional spatial information of the PAN image into the LRMS image, we design the PCDRN for pansharpening. As shown in Figure 2, PCDRN consists of two residual subnetworks ResNet1 and ResNet2, which are progressively cascaded to learn nonlinear feature mapping from LRMS and PAN images to HRMS images.

In our experiments, the PCDRN was implemented through three stages.

Stage 1: The LRMS images are upsampled by a scale of 2 with nearest-neighbor interpolation, and the PAN images are downsampled by a scale of 2. The upsampled MS images are then concatenated with the downsampled PAN images to form the 5-band inputs.

Stage 2: The 5-band inputs are fed into ResNet1 to extract the coarse features. The ResNet includes 2 convolutional layers and 5 residual blocks. An element-wise sum is then performed on the feature

maps of ResNet1 and the upsampled LRMS image in a channel by channel manner. Subsequently,  $1 \times 1$  convolutional layers are utilized to reduce spectral dimensionality from 64 bands to 4 bands. The 4-band results are upsampled by a scale of 2 using nearest-neighbor interpolation and then concatenated with the PAN images to obtain new 5-band inputs.

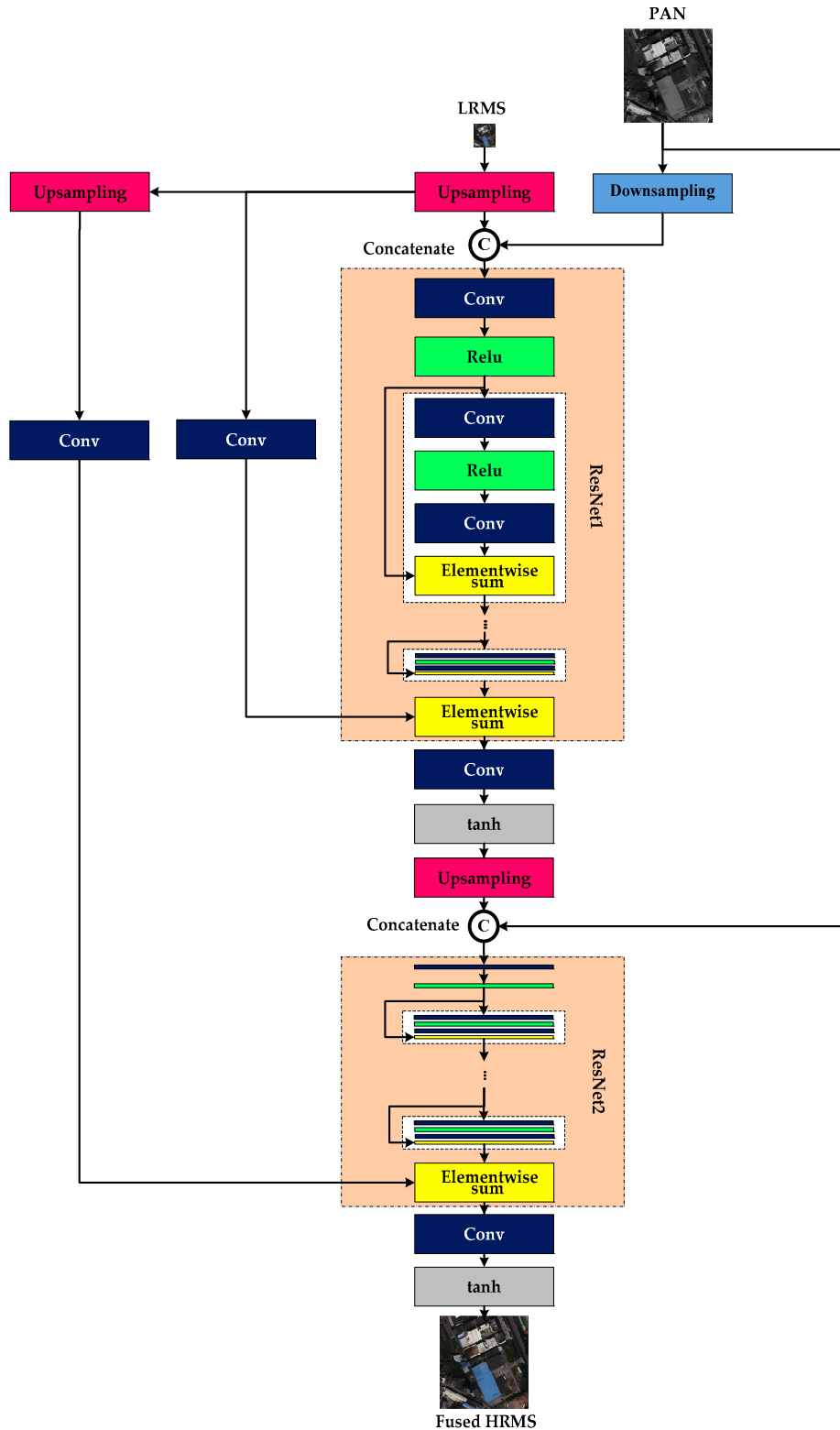
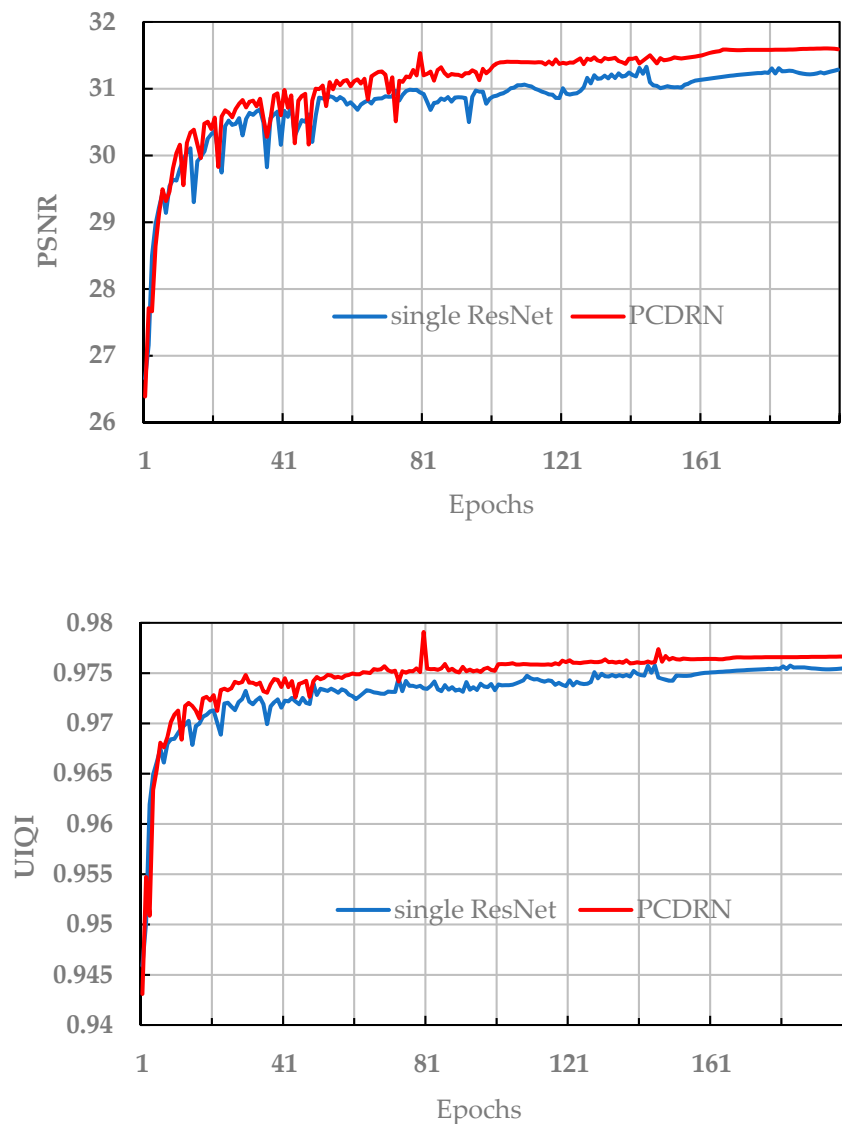


Figure 2. Flowchart of the proposed progressive cascade deep residual network (PCDRN).

Stage 3: The new 5-band inputs are entered into the ResNet2, which exhibits the same structure as that of the previous ResNet1, to extract finer features. After the LRMS image has been upsampled twice, an element-wise sum is performed on the results of ResNet2 and the upsampled LRMS image, channel by channel. The fused HRMS image is finally obtained by a  $1 \times 1$  convolutional layer and an activation function tanh. In addition,  $1 \times 1$  convolution was employed several times in the network to achieve the reduction of network parameters and boost the performance of PCDRN.

To validate the advantages of PCDRN, a performance comparison between the single ResNet and PCDRN is presented in Figure 3. We observe that the two indices obtained by PCDRN are significantly better than those obtained by the single ResNet.



**Figure 3.** Average peak signal-to-noise ratio (PSNR) [24] and universal image quality index (UIQI) [23] of the single ResNet and PCDRN on 180 groups of the simulated dataset from Pléiades.

### 3.2. Multitask Loss Function

A mean squared error (MSE) loss function is usually applied in deep learning-based pansharpening methods. However, the MSE loss function often loses high-frequency details, such as texture during the fusion process, which may result in poor perceptual quality and the over-smooth phenomenon. To address this problem, we design a novel multitask loss function comprising MSE loss and universal image quality index (UIQI) loss.

The UIQI is often used to measure the structure distortion degree in image quality evaluation. Based on the characteristics of UIQI, the UIQI loss in our network has been designed to preserve structure information. Therefore, our multitask loss function  $L^{Fusion}$  can improve the performance of the fusion network in preserving spatial details.

The  $L^{Fusion}$  is represented as:

$$L^{Fusion} = \alpha \cdot L_{NMSE}^{Fusion} + (1 - \alpha) \cdot L_{NUIQI}^{Fusion} \quad (4)$$

where  $L_{NMSE}^{Fusion}$  is the normalized MSE loss,  $L_{NUIQI}^{Fusion}$  is the normalized UIQI loss, and  $\alpha$  denotes the weight coefficient.

The  $L_{NMSE}^{Fusion}$  and  $L_{NUIQI}^{Fusion}$  are defined as:

$$L_{NMSE}^{Fusion} = \beta \cdot \left( \frac{1}{n} \sum_{i=1}^n \|F(x^{(i)}) - R^{(i)}\|^2 \right) \quad (5)$$

and

$$L_{NUIQI}^{Fusion} = \gamma \cdot \left( 1 - \frac{1}{n} \sum_{i=1}^n UIQI(F(x^{(i)}), R^{(i)}) \right) \quad (6)$$

where  $\beta$  and  $\gamma$  denote the normalized coefficients,  $n$  denotes the number of train image groups,  $x^{(i)}$  denotes a low-resolution MS image,  $F(x^{(i)})$  denotes a fused MS image, and  $R^{(i)}$  denotes a reference MS image.

In order to roughly balance the contribution of the MSE and UIQI losses, a normalized algorithm is introduced as following Algorithm 1.

---

**Algorithm 1. The Algorithm for Normalization method**

---

Input: LRMS image (*lrms*), PAN image (*pan*),

the number of training epochs (*max\_train\_epoch*)

Output: the normalized coefficients  $\beta$  and  $\gamma$

Initialize:

1)  $c \leftarrow 0.00001$

For  $i = 0$  to *max\_train\_epoch* do

2) Input *lrms*( $i$ ) and *pan*( $i$ ) to compute  $L_{MSE}(i)$  by  $L_{MSE}^{Fusion} = \frac{1}{n} \sum_{i=1}^n \|F(x^{(i)}) - R^{(i)}\|^2$  and  $L_{UIQI}(i)$  by

$L_{UIQI}^{Fusion} = 1 - \frac{1}{n} \sum_{i=1}^n UIQI(F(x^{(i)}), R^{(i)})$ , respectively

If (converge)

3) Compute  $\nabla L_{MSE}(i)$  by  $\nabla L_{MSE}(i) = |L_{MSE}(i) - L_{MSE}(i-1)|$

4) Compute  $\nabla L_{UIQI}(i)$  by  $\nabla L_{UIQI}(i) = |L_{UIQI}(i) - L_{UIQI}(i-1)|$

5) Compute  $\overline{\nabla L_{MSE}}$  by averaging the  $\nabla L_{MSE}(i)$

6) Compute  $\overline{\nabla L_{UIQI}}$  by averaging the  $\nabla L_{UIQI}(i)$

Endif

Endfor

7) Compute the coefficient  $\beta$  by  $\frac{c}{\overline{\nabla L_{MSE}}}$  and the coefficient  $\gamma$  by  $\frac{c}{\overline{\nabla L_{UIQI}}}$

---

Finally, the  $L_{NMSE}^{Fusion}$  and  $L_{NUIQI}^{Fusion}$  are obtained by formula (5) and formula (6), respectively.

To illustrate the validity of the proposed loss function, we compare the performances of the MSE loss function and the MSE + UIQI loss function, as shown in Table 1. The results demonstrate that the PSNR value obtained by using the proposed MSE + UIQI loss function is higher than that obtained by only using the MSE loss function. Similarly, our network comprising the MSE + UIQI loss function also obtains UIQI value, which is higher than that obtained by using the MSE loss function.

**Table 1.** Average PSNR and UIQI of PCDRN using MSE and MSE+UIQI loss on 180 groups of the simulated dataset from Pléiades.

Methods	PSNR (↑)	UIQI (↑)
PCDRN using MSE loss	31.3716	0.9754
PCDRN using MSE+UIQI loss	<b>31.5914</b>	<b>0.9767</b>

### 3.3. Resize-Convolution

Unlike the traditional convolution, the transposed convolution forms the connectivity in the backward direction [25]. The transposed convolution is usually employed for upsampling images. In addition, different from the weights in a traditional filter, the weights in the transposed convolution are learnable without being predefined. However, transposed convolution can easily produce uneven overlap, which may result in checkerboard artifacts in the image. Therefore, the resize-convolution approach is employed to avoid the problem of uneven overlap, which has been known to be robust against checkerboard artifacts [26]. The approach involves resizing an image using nearest-neighbor interpolation and then executing a convolutional operation. To better illustrate this phenomenon, we carried out a group of experiments to compare the performance of the fused image that employs transposed convolution versus resize-convolution, which can be seen in Figure 4. As shown in Figure 4, a small region is enlarged and shown on the left bottom for better visualization. From the enlarged box in Figure 4a, we can observe the obvious checkerboard artifacts. From the enlarged box in Figure 4b, we can observe that the region is smoother than that in Figure 4a.

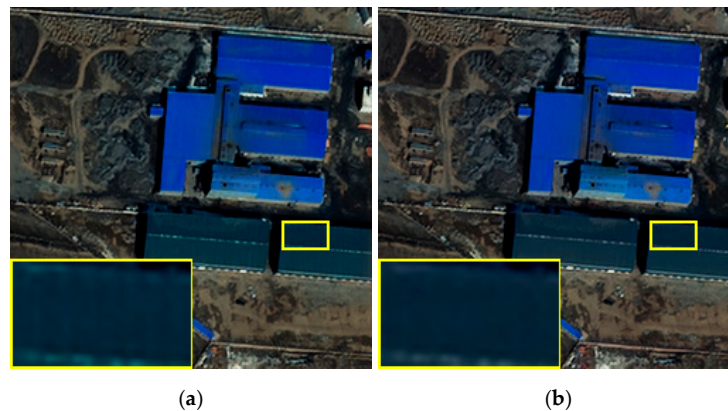
**Figure 4.** The comparison between transposed convolution and resize-convolution from the WorldView-3 dataset. (a) Fused image using transposed convolution; and (b) fused image using resize-convolution.

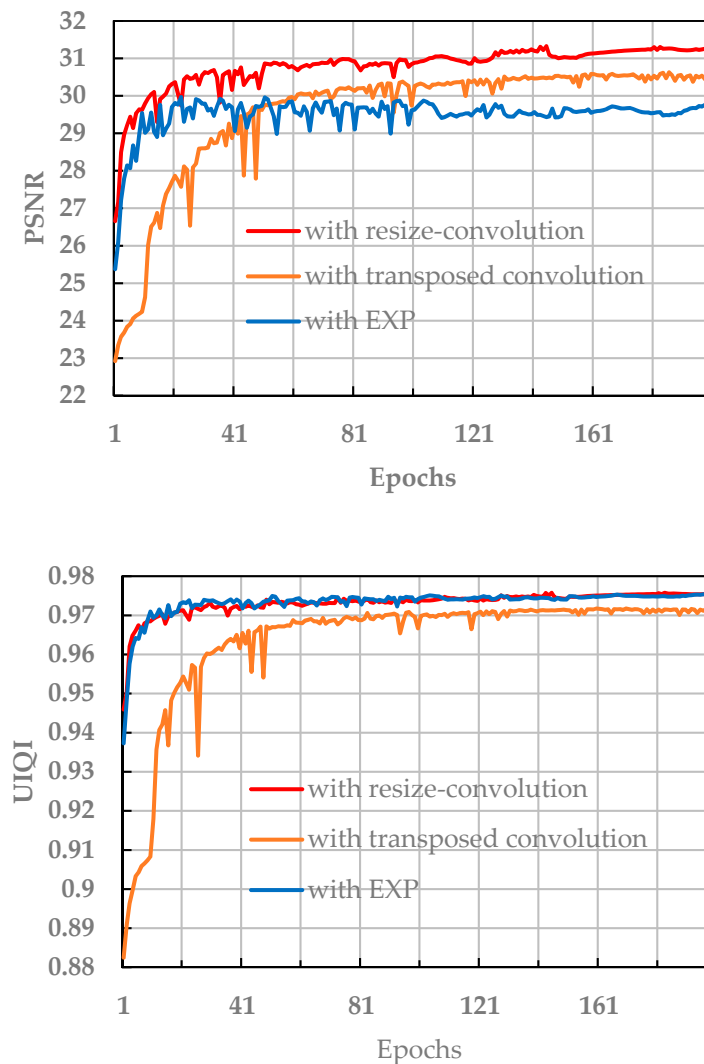
Table 2 shows the quantitative assessment of the results in Figure 4, in which the bold represents the best value. As we can observe from the table, the experimental results by using resize-convolution are better than those by using transposed convolution in most image quality indexes, including PSNR [24], the correlation coefficient (CC) [27], UIQI [23], the spectral angle mapper (SAM) [28], and the erreur relative globale adimensionnelle de synthèse (ERGAS) [29] except the  $Q2^n$  [30] index.

**Table 2.** Quantitative assessment of the results in Figure 4.

Methods	PSNR (↑)	CC (↑)	UIQI (↑)	$Q2^n$ (↑)	SAM (↓)	ERGAS (↓)
Transposed convolution	26.3580	0.9705	0.9576	<b>0.8460</b>	5.0280	5.3059
Resize-convolution	<b>26.6129</b>	<b>0.9760</b>	<b>0.9621</b>	0.8412	<b>4.4173</b>	<b>5.1850</b>



To further demonstrate the superiority of the resize-convolution method, the polynomial (EXP) [31], transposed convolution, and resize-convolution methods were evaluated by performing experiments on 180 groups of the simulated dataset from Pléiades through a single ResNet. Figure 5 shows the average PSNR and UIQI of these three methods. From Figure 5, we can find that single ResNet with resize-convolution outperforms the other two methods on PSNR. In addition, it not only achieves higher UIQI values than the method with transposed convolution but also achieves similar UIQI values to the method with EXP. Thus, on the whole, the single ResNet with resize-convolution is superior to the methods with transposed convolution and EXP.



**Figure 5.** Average PSNR and UIQI of 3 methods on 180 groups of the simulated dataset from Pléiades.

## 4. Experimental Results

### 4.1. Experimental Settings

#### 4.1.1. Datasets

In the experiments, the MS images and PAN images were captured by the Pléiades and WorldView-3 satellites, respectively. The MS images of the Pléiades satellite contain four bands: red, green, blue, and near-infrared. We selected red, green, blue, and near-infrared1 bands from the MS image of the WorldView-3 satellite to comprise the new 4-band MS image. The spatial resolutions of Pleiades and WorldView-3 datasets are described in Table 3.



**Table 3.** The spatial resolutions of Pleiades and WorldView-3 datasets.

	PAN	MS
Pleiades	0.5 m GSD (0.7 m GSD at nadir)	2 m GSD (2.8 m GSD at nadir)
WorldView-3	0.31 m GSD at nadir	1.24 m GSD at nadir

The source images were degraded to lower resolution by a factor of 4 using Wald’s protocol [32,33]. In the approach, the original MS images were degraded by using a low-pass filter matched with modulation transfer function (MTF) of the MS sensor, and the original PAN images were degraded by using the ‘bicubic’ method. The sizes of the original MS and PAN images are  $256 \times 256$  and  $1024 \times 1024$  pixels, respectively. In training, the sizes of the degraded MS and PAN images are  $64 \times 64$  and  $256 \times 256$  pixels, respectively. The original MS images are considered referenced images. To enhance the generalization of our network, data augmentation approaches in [34] were used in the training process. The 180 groups of simulated data from the Pléiades satellite and the 108 groups of simulated data from WorldView-3 satellite were used as test datasets, respectively. Each group of simulated data is composed of a  $64 \times 64$  degraded MS image, a  $256 \times 256$  degraded PAN image, and a  $256 \times 256$  reference MS image. In addition, the 165 groups of real data from Pléiades satellite and 100 groups of real data from WorldView-3 satellite were used to assess the performance of PCDRN. It should be noted that the test datasets are not used to train PCDRN.

#### 4.1.2. Training Details

The training of PCDRN is achieved in 200 epochs with a batch size of 9 for using adaptive moment estimation (Adam). The learning rate  $\epsilon$  is initialized to  $8 \times 10^{-5}$ , which is divided by a factor of 2 for every 50 epochs. The biases in the network are initialized to zero.

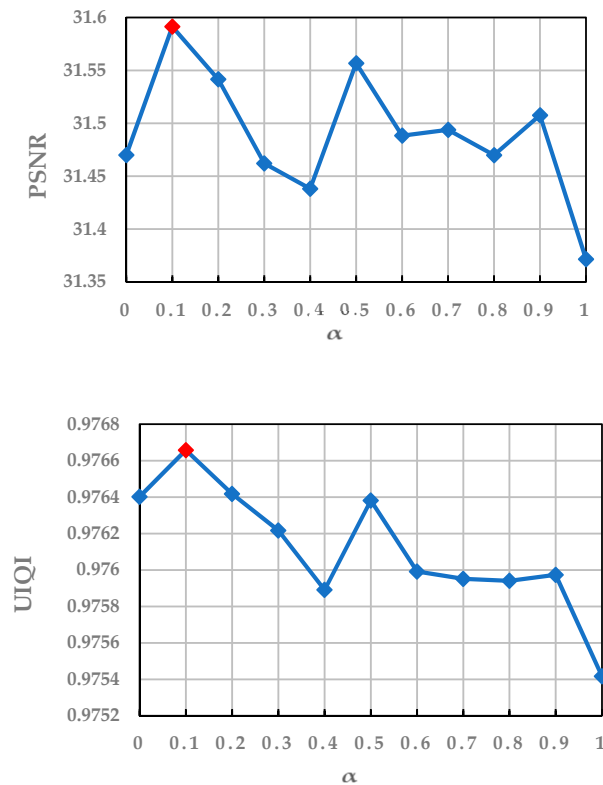
To further demonstrate the effectiveness of the multitask loss function, numerous experiments were performed using different  $\alpha$  values on 180 sets of the simulated dataset from Pléiades. The average PSNR and UIQI of PCDRN using different  $\alpha$  values on 180 groups of the simulated dataset from Pléiades are shown in Figure 6. As shown in Figure 6, we can observe that the best performances are obtained when  $\alpha$  is set to 0.1 in formula (4). The PCDRN is implemented under TensorFlow 1.8 and TensorLayer 1.8. The experiments were performed on an NVIDIA GeForce GT 1080Ti GPU.

#### 4.1.3. Compared Methods

In this paper, PCDRN is compared with one interpolated method and eleven popular pansharpening methods.

1. EXP: an interpolation method based on polynomial kernel [31];
2. AIHS: adaptive IHS [35];
3. ATWT: a Trous wavelet transform [9];
4. GSA: Gram Schmidt adaptive [36];
5. BT: Brovey transform [37];
6. MTF-GLP-CBD: generalized Laplacian Pyramid (GLP) [31] with MTF-matched filter [38] and regression based injection model [28];
7. MMT: a matting model and multiscale transform [16];
8. GS: Gram Schmidt [6];
9. MTF-GLP-HPM: GLP [31] with MTF-matched filter [38] and multiplicative injection model [39];
10. ASIM: adaptive spectral-intensity modulation [40];
11. DRPNN: a deep ResNet for pansharpening [41];
12. MSDCNN: a multiscale and multidepth CNN for pansharpening [12].

The AIHS, GS, GSA, and BT are four typical CS-based methods. The ATWT, MTF-GLP-CBD, and MTF-GLP-HPM are three MRA-based methods. MMMT and ASIM are two model-based methods. Both DRPNN and MSDCNN are deep learning-based methods.



**Figure 6.** Average PSNR and UIQI of PCDRN using different  $\alpha$  values on 180 groups of the simulated dataset from Pléiades.

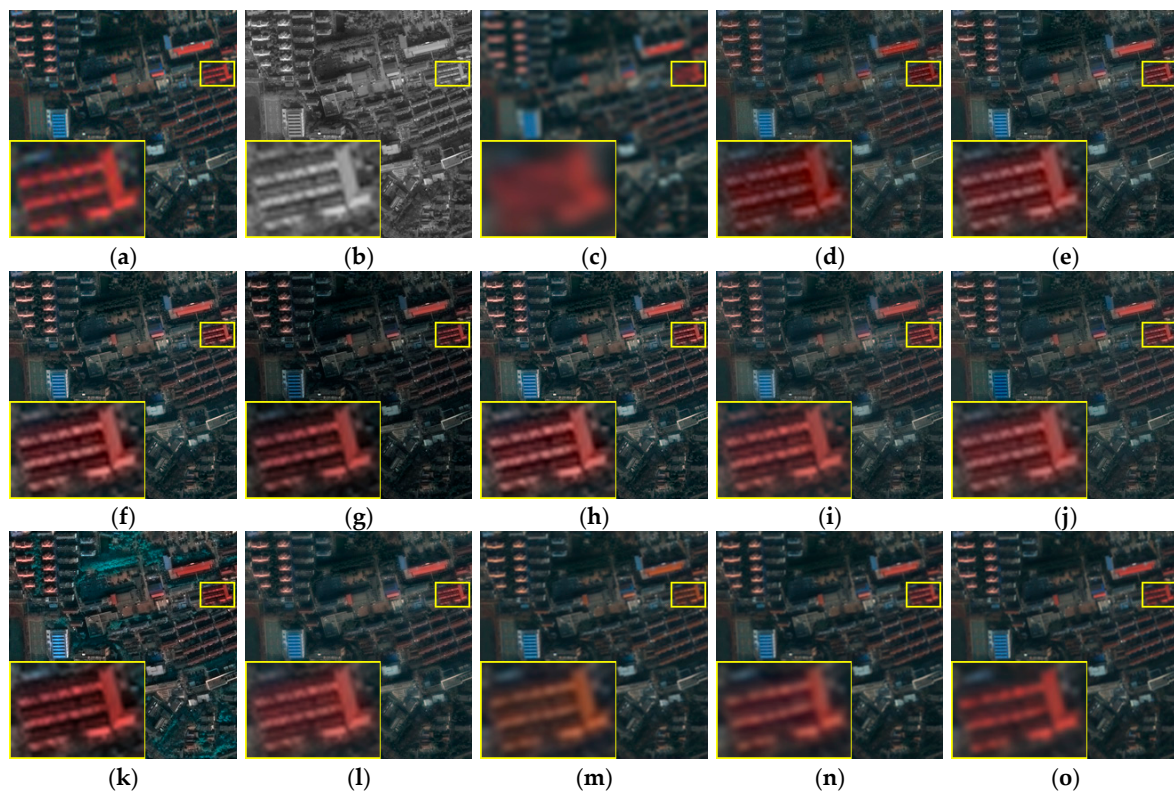
#### 4.2. Experiments on Simulated Data

In this subsection, to demonstrate the validity of PCDRN, we compared the visual effects and quantitative assessments of simulated experimental results obtained from different methods. Six metrics were used to evaluate the performance of PCDRN on simulated data from the Pléiades and WorldView-3 datasets, which include PSNR [24], CC [27], UIQI [23],  $Q2^n$  [30], SAM [28], and ERGAS [29].

To better observe the texture details in the fused images, there are two rectangle boxes marked in yellow, and the bigger box is the enlarged image of the smaller one. Note that the bold represents the best performance of quantitative assessment in experiments.

##### 4.2.1. Experiments on Pléiades Dataset

A group of fusion results on simulated data from Pléiades are shown in Figure 7. Figure 7a is the reference image that was applied to assess the fused images. Figure 7b displays the degraded PAN. In Figure 7c, the upsampled image by the EXP has good spectral quality but exhibits serious spatial distortions. The corresponding fusion images obtained by twelve pansharpening methods are presented in Figure 7d–o, respectively. As shown in Figure 7, the fusion results of the BT and MTF\_GLP\_HPM methods suffer from serious spectral distortions. The fusion results of other compared methods have obvious spectral distortions, such as the roof of the building in the enlarged box. However, compared to all the comparison pansharpening methods, the fused result of PCDRN is closer to the reference image by observation. Furthermore, from the corresponding quantitative assessment of Figure 7 as shown in Table 4, PCDRN outperforms the other eleven pansharpening methods in six metrics.



**Figure 7.** Fusion results on simulated data from Pléiades. (a) Reference image; (b) PAN; (c) EXP; (d) AIHS; (e) ATWT; (f) GSA; (g) BT; (h) MTF\_GLP\_CBD; (i) MMTT; (j) GS; (k) MTF\_GLP\_HPM; (l) ASIM; (m) DRPNN; (n) MSDCNN; and (o) PCDRN.

**Table 4.** Quantitative assessment of results in Figure 7.

Methods	PSNR (↑)	CC (↑)	UIQI (↑)	Q2 <sup>n</sup> (↑)	SAM (↓)	ERGAS (↓)
EXP *	26.5538	0.9008	0.8900	0.8129	4.1067	5.0701
AIHS	27.4025	0.9231	0.9074	0.8491	4.5483	4.6003
ATWT	28.0016	0.9235	0.9279	0.8745	4.5283	4.2789
GSA	27.0666	0.9130	0.9265	0.8702	4.5274	4.7951
BT	20.6228	0.8936	0.8179	0.4855	4.3264	15.1748
MTF_GLP_CBD	27.2334	0.9129	0.9276	0.8712	4.4921	4.6934
MMMT	27.7634	0.9212	0.9225	0.8698	4.7816	4.3933
GS	26.7568	0.9046	0.8889	0.8366	4.5298	4.9587
MTF_GLP_HPM	22.3144	0.7899	0.8482	0.7274	4.7561	7.6598
ASIM	28.8910	0.9383	0.9448	0.9059	4.1985	3.8388
DRPNN	29.2947	0.9712	0.9500	0.9070	3.4037	3.9624
MSDCNN	28.0166	0.9608	0.9394	0.8660	3.4079	4.6946
PCDRN	<b>30.2689</b>	<b>0.9846</b>	<b>0.9681</b>	<b>0.9171</b>	<b>2.9567</b>	<b>3.5617</b>

\* The EXP method is an interpolation method for MS Image, which is not a pansharpening method.

Furthermore, we perform experiments on 180 sets of the simulated dataset from Pléiades. Table 5 tabulates the mean values of experimental results. As shown in Table 5, we can clearly observe that PCDRN produces the best quantitative assessment results of the most metrics among all pansharpening methods.

**Table 5.** Average quantitative results on 180 groups of the simulated dataset from Pléiades.

Methods	PSNR (↑)	CC (↑)	UIQI (↑)	Q2 <sup>n</sup> (↑)	SAM (↓)	ERGAS (↓)
EXP *	26.9511	0.9279	0.9257	0.8161	3.1093	3.9761
AIHS	28.2598	0.9502	0.9431	0.8659	3.4181	3.4507
ATWT	28.7868	0.9504	0.9550	0.8832	3.3870	3.2380
GSA	27.9760	0.9454	0.9530	0.8757	3.5046	3.5725
BT	19.8484	0.9355	0.8576	0.5249	3.2301	12.8501
MTF_GLP_CBD	27.9234	0.9437	0.9519	0.8717	3.5005	3.5770
MMMT	28.6014	0.9496	0.9528	0.8839	3.5549	3.2942
GS	27.5243	0.9400	0.9331	0.8536	3.5229	3.7475
MTF_GLP_HPM	23.8909	0.8678	0.8995	0.7632	3.9791	5.5050
ASIM	29.4570	0.9594	0.9643	<b>0.9067</b>	3.2244	2.9717
DRPNN	30.1216	0.9805	0.9680	0.8846	2.6809	2.9884
MSDCNN	29.2428	0.9739	0.9629	0.8638	2.7117	3.2713
PCDRN	<b>31.5914</b>	<b>0.9884</b>	<b>0.9767</b>	0.9004	<b>2.3488</b>	<b>2.6322</b>

\* The EXP method is an interpolation method for MS Image, which is not a pansharpening method.

#### 4.2.2. Experiments on WorldView-3 Dataset

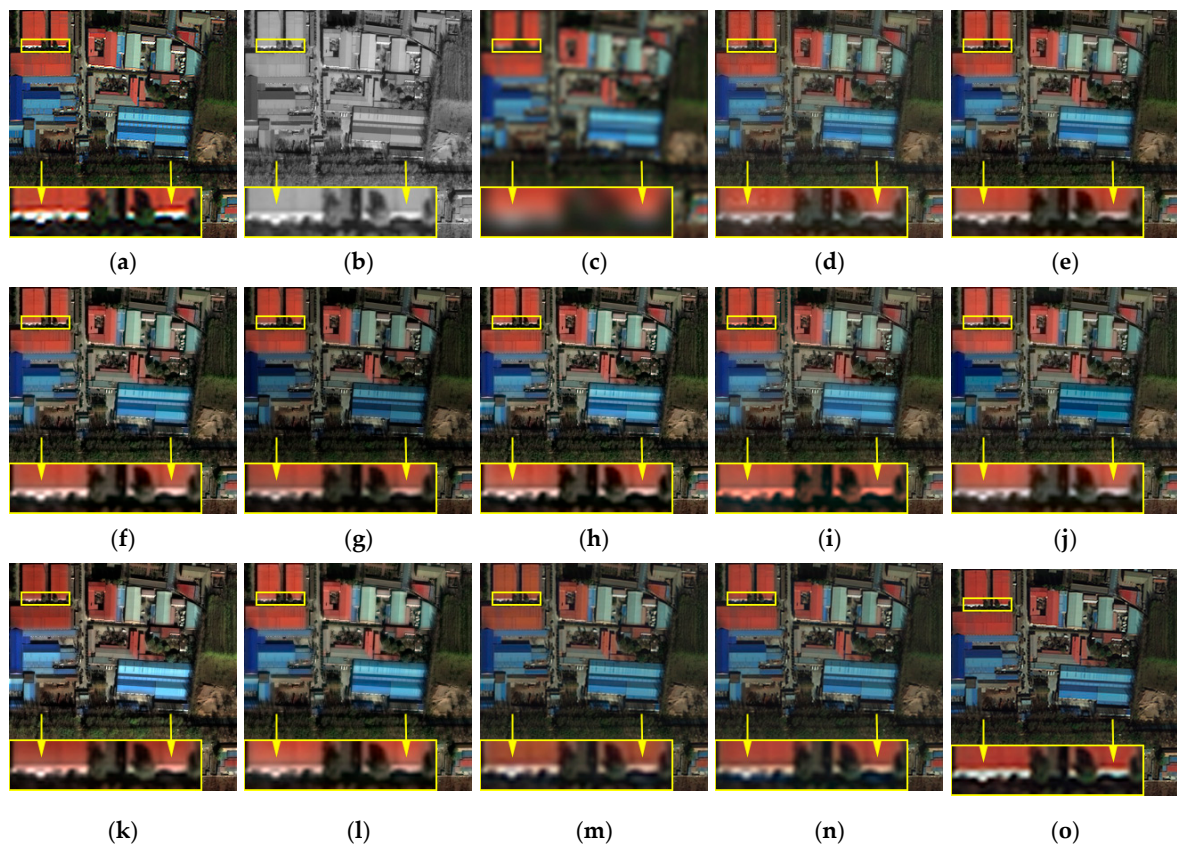
Figure 8 shows an example of the simulated experiment performed on the WorldView-3 dataset. The reference and Pan images are shown in Figure 8a,b, respectively. The upsampled LRMS using EXP method is shown in Figure 8c, which is also blurred as can be seen from the enlarged box. The 12 fused results are given in Figure 8d–o. From Figure 8d–o, we can observe that the results of the BT, MMT, and DRPNN methods have serious spectral distortions. The fused results of the ATWT, GSA, MTF\_GLP\_CBD, GS, MTF\_GLP\_HPM, and ASIM methods exhibit some spectral distortions on the roof of buildings in the enlarged box. The AIHS method yields obvious spectral distortion and some artifacts that do not exist in the original image. Although the results of MSDCNN method have good spatial quality, they show some spectral distortions in the enlarged box. In Figure 8, we can observe a red line between the white and red areas, which can be clearly observed in the results of PCDRN. However, in the results of other methods, the red line is very blurred or even nonexistent. We can also observe that PCDRN is superior to other methods in preserving the spectral fidelity in the enlarged box. For the corresponding quantitative evaluation of Figure 8 (see Table 6), PCDRN gains the best fusion values in most metrics.

**Table 6.** Quantitative assessment of results in Figure 8.

Methods	PSNR (↑)	CC (↑)	UIQI (↑)	Q2 <sup>n</sup> (↑)	SAM (↓)	ERGAS (↓)
EXP *	21.7015	0.8676	0.8758	0.6946	4.9591	7.5501
AIHS	23.4222	0.9412	0.9097	0.8054	5.9824	5.6479
ATWT	25.3147	0.9466	0.9433	0.8957	5.7960	4.9356
GSA	25.6204	0.9480	0.9502	0.9045	7.0358	4.7366
BT	20.9394	0.9353	0.8793	0.6533	5.0673	10.7228
MTF_GLP_CBD	25.6786	0.9485	0.9490	0.9053	6.7749	<b>4.6830</b>
MMMT	24.9489	0.9422	0.9367	0.8862	6.0845	5.1410
GS	23.5740	0.9267	0.9081	0.8455	6.6026	6.1103
MTF_GLP_HPM	23.1149	0.9185	0.9277	0.8365	5.2112	5.9757
ASIM	25.6169	0.9473	0.9488	0.9083	5.9963	4.7283
DRPNN	24.5411	0.9604	0.9258	0.8676	5.8623	5.6337
MSDCNN	23.9079	0.9508	0.9175	0.8470	5.8183	6.1578
PCDRN	<b>26.0631</b>	<b>0.9817</b>	<b>0.9566</b>	<b>0.9206</b>	<b>4.7731</b>	4.9920

\* The EXP method is an interpolation method for MS Image, which is not a pansharpening method.





**Figure 8.** Fusion results on simulated data from WorldView-3. (a) Reference image; (b) PAN; (c) EXP; (d) AIHS; (e) ATWT; (f) GSA; (g) BT; (h) MTF\_GLP\_CBD; (i) MMTT; (j) GS; (k) MTF\_GLP\_HPM; (l) ASIM; (m) DRPNN; (n) MSDCNN; and (o) PCDRN.

Table 7 tabulates the average quantitative results on 108 groups of the simulated dataset from WorldView-3. As can be observed, PCDRN achieves the best fusion results in most metrics.

**Table 7.** Average quantitative results on 108 groups of the simulated dataset from WorldView-3.

Methods	PSNR (↑)	CC (↑)	UIQI (↑)	Q2 <sup>n</sup> (↑)	SAM (↓)	ERGAS (↓)
EXP *	24.4725	0.8696	0.8754	0.6853	4.5627	6.2634
AIHS	26.0239	0.9431	0.9087	0.7960	5.1772	4.9621
ATWT	27.7866	0.9460	0.9419	0.8771	4.9188	4.1494
GSA	28.3915	0.9501	0.9534	<b>0.8962</b>	5.3194	3.8844
BT	21.8128	0.9367	0.8728	0.5729	4.6377	10.9669
MTF_GLP_CBD	28.2251	0.9483	0.9510	0.8906	5.2584	3.9354
MMMT	27.5249	0.9433	0.9355	0.8650	5.0332	4.2669
GS	26.8001	0.9399	0.9183	0.8504	5.2075	4.7135
MTF_GLP_HPM	23.6310	0.8831	0.8986	0.7906	5.7262	6.3586
ASIM	28.1808	0.9475	0.9484	0.8869	4.8056	3.9356
DRPNN	27.0167	0.9566	0.9337	0.8449	4.9077	4.5371
MSDCNN	26.5101	0.9506	0.9263	0.8300	5.0109	4.8904
PCDRN	<b>28.8141</b>	<b>0.9781</b>	<b>0.9571</b>	0.8872	<b>4.2326</b>	<b>3.7930</b>

\* The EXP method is an interpolation method for MS Image, which is not a pansharpening method.

Therefore, according to the above experimental results on the simulated data from Pléiades and WorldView-3 satellites, we find that PCDRN is better than the other 11 pansharpening methods.

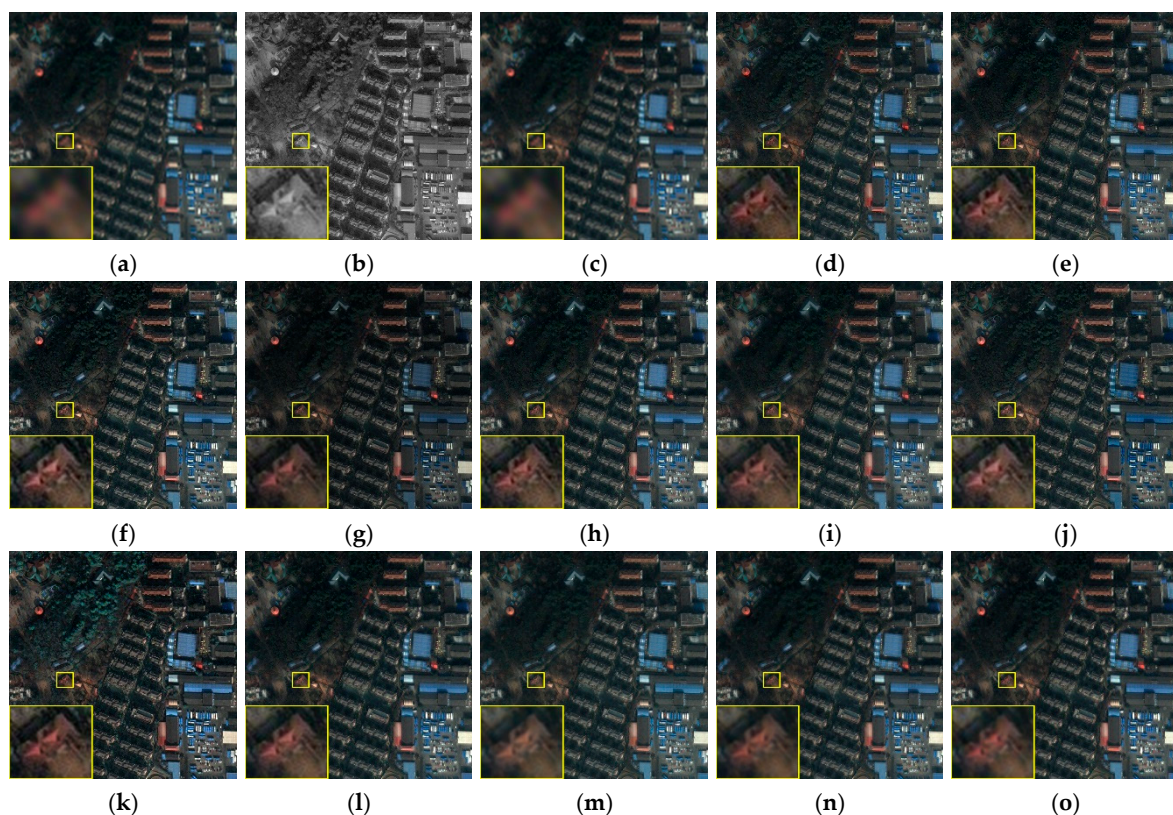
### 4.3. Experiments on Real Data

To demonstrate the effectiveness of our proposed method, we also compare the visual effects and quantitative evaluations of real data experimental results obtained from 12 pansharpener methods. The twelve pansharpener methods are evaluated on the real dataset from Pléiades and WorldView-3 in terms of quality with no reference (QNR) index [42]. The QNR metric is composed of the spectral distortion index  $D_\lambda$  and the spatial distortion index  $D_S$ . The optimum values of QNR,  $D_\lambda$ , and  $D_S$  are 1, 0, and 0 respectively.

Similar to what was demonstrated on simulated data, two yellow rectangle boxes are marked in each fused image to better observe the texture details. The bigger box is the enlarged image of the smaller one.

#### 4.3.1. Experiments on Pléiades Dataset

A set of fusion results on real data from Pleiades are shown in Figure 9. Figure 9a is the upsampled LRMS using bicubic interpolation. Figure 9b displays the PAN image. Figure 9c is the upsampled LRMS using the EXP method. As shown in Figure 9c, the upsampled image by the EXP method has good spectral quality but exhibits serious spatial distortions. As we can see in Figure 9d–o, the result of AIHS method has some artifacts in the enlarged box. BT and MTF\_GLP\_HPM methods yield serious spectral distortions. The results obtained by ATWT, GSA, MTF\_GLP\_CBD, MMTT, GS, ASIM, and DRPNN methods have some spectral distortions because they are oversharpened in the process of pansharpener. The MSDCNN method produces slight spectral distortions as seen in the enlarged box. Table 8 also shows the quantitative assessment of the experimental results with real data in Figure 9. As it can be observed, PCDRN obtains the best values in QNR and  $D_\lambda$  metrics.



**Figure 9.** Fusion results on real data from Pleiades. (a) MS image; (b) PAN; (c) EXP; (d) AIHS; (e) ATWT; (f) GSA; (g) BT; (h) MTF\_GLP\_CBD; (i) MMTT; (j) GS; (k) MTF\_GLP\_HPM; (l) ASIM; (m) DRPNN; (n) MSDCNN; and (o) PCDRN.

**Table 8.** Quantitative assessment of results in Figure 9.

Methods	$D_\lambda$ (↓)	$D_S$ (↓)	QNR (↑)
EXP *	0.0012	0.1961	0.8029
AIHS	0.0954	0.1035	0.8110
ATWT	0.1055	0.1080	0.7978
GSA	0.1447	0.1621	0.7167
BT	0.1509	0.1192	0.7479
MTF_GLP_CBD	0.1145	0.1042	0.7932
MMMT	0.0997	0.0908	0.8185
GS	0.1110	0.1452	0.7599
MTF_GLP_HPM	0.1284	0.1497	0.7410
ASIM	0.1008	0.0996	0.8096
DRPNN	0.0581	<b>0.0162</b>	0.9266
MSDCNN	0.0495	0.0243	0.9274
PCDRN	<b>0.0376</b>	0.0276	<b>0.9358</b>

\* The EXP method is an interpolation method for MS Image, which is not a pansharpening method.

The average quantitative results on 165 groups of the real dataset from Pleiades are listed in Table 9. From Table 9, we can testify that PCDRN not only obtains the best results of QNR and  $D_\lambda$  but also obtains the second best results of  $D_S$  in all the comparison methods.

**Table 9.** Average quantitative results on 165 Groups of the real dataset from Pléiades.

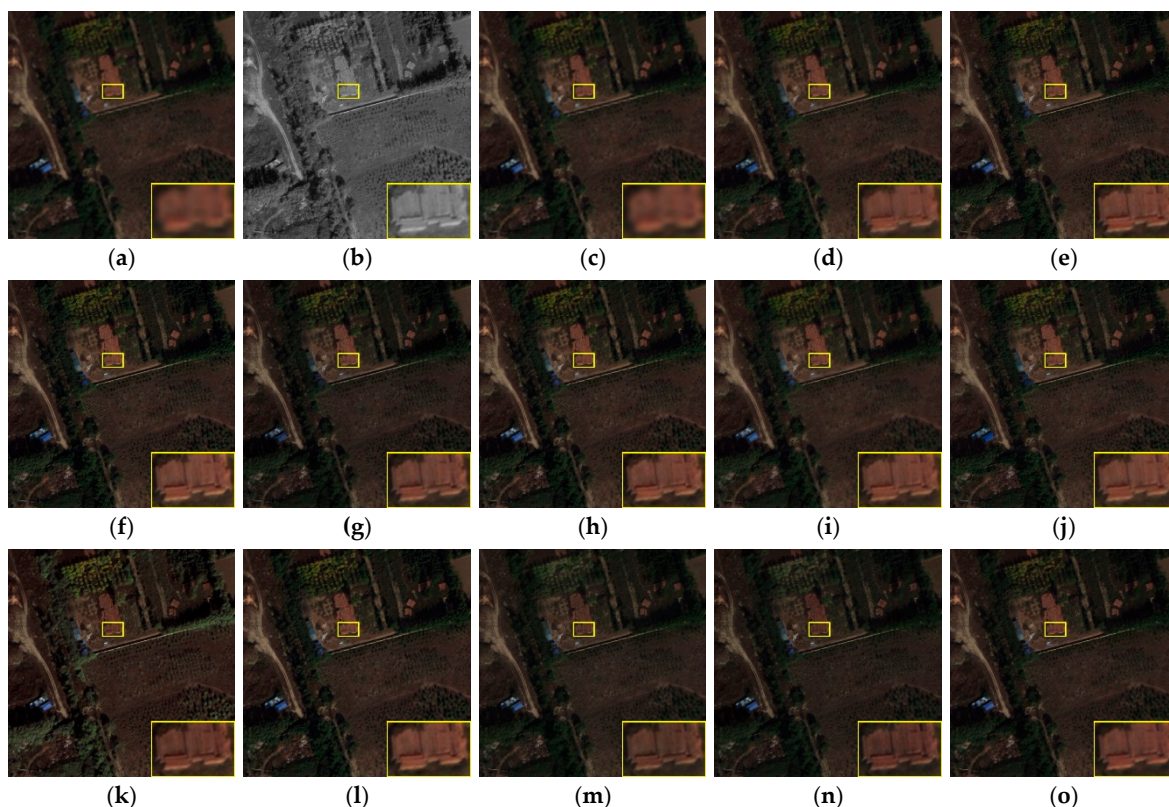
Methods	$D_\lambda$ (↓)	$D_S$ (↓)	QNR (↑)
EXP *	0.0026	0.2055	0.7924
AIHS	0.1021	0.1142	0.7961
ATWT	0.1121	0.1172	0.7848
GSA	0.1347	0.1619	0.7259
BT	0.1660	0.1433	0.7153
MTF_GLP_CBD	0.1098	0.1011	0.8009
MMMT	0.1013	0.0982	0.8107
GS	0.1076	0.1465	0.7625
MTF_GLP_HPM	0.1424	0.1427	0.7368
ASIM	0.1105	0.1142	0.7887
DRPNN	0.0533	<b>0.0363</b>	0.9123
MSDCNN	0.0434	0.0462	0.9123
PCDRN	<b>0.0375</b>	0.0382	<b>0.9256</b>

\* The EXP method is an interpolation method for MS Image, which is not a pansharpening method.

#### 4.3.2. Experiments on WorldView-3 Dataset

Figure 10 displays a group of fusion results on real data from WorldView-3. The upsampled LRMS using bicubic interpolation is presented in Figure 10a, and Figure 10b is the PAN image. The upsampled LRMS using EXP method is given in Figure 10c, from which we can observe that the EXP method yields obvious spatial distortions. From Figure 10d–o, we can observe that the AIHS method yields some artifacts in the enlarged box. MTF\_GLP\_HPM, DRPNN and MSDCNN methods produce obvious spectral distortions. ATWT, GSA, BT, MTF\_GLP\_CBD, MMT, GS, and ASIM methods yield some spectral distortions in the enlarged box. For the corresponding quantitative evaluation of Figure 10 (see Table 10), PCDRN gains the best fusion values in QNR and  $D_\lambda$  metrics.





**Figure 10.** Fusion results on real data from WorldView-3. (a) MS; (b) PAN; (c) EXP; (d) AIHS; (e) ATWT; (f) GSA; (g) BT; (h) MTF\_GLP\_CBD; (i) MMMT; (j) GS; (k) MTF\_GLP\_HPM; (l) ASIM; (m) DRPNN; (n) MSDCNN; and (o) PCDRN.

**Table 10.** Quantitative assessment of results in Figure 10.

Methods	$D_{\lambda}$ (↓)	$D_S$ (↓)	QNR (↑)
EXP *	0.0005	0.1082	0.8913
AIHS	0.0391	<b>0.0457</b>	0.9169
ATWT	0.0676	0.0730	0.8643
GSA	0.0881	0.1130	0.8089
BT	0.1255	0.1113	0.7772
MTF_GLP_CBD	0.0749	0.0738	0.8568
MMMT	0.0493	0.0602	0.8935
GS	0.0534	0.0633	0.8866
MTF_GLP_HPM	0.1167	0.1460	0.7544
ASIM	0.0857	0.0907	0.8314
DRPNN	0.0589	0.0586	0.8859
MSDCNN	0.0554	0.0683	0.8801
PCDRN	<b>0.0281</b>	0.0480	<b>0.9253</b>

\* The EXP method is an interpolation method for MS Image, which is not a pansharpening method.

Table 11 tabulates the average quantitative results on 100 groups of the real dataset from WorldView-3. As we can see in Table 11, PCDRN obtains the optimal fusion results in all metrics.

Thus, from the above-mentioned experimental results on real data from Pléiades and WorldView-3 satellites, we can conclude that PCDRN outperforms other fusion methods in balancing the spectral preservation and spatial enhancement.

**Table 11.** Average quantitative results on 100 Groups of the real dataset from WorldView-3.

Methods	$D_{\lambda}$ (↓)	$D_S$ (↓)	QNR (↑)
EXP *	0.0028	0.1163	0.8813
AIHS	0.0585	0.1032	0.8451
ATWT	0.0679	0.1113	0.8292
GSA	0.0778	0.1324	0.8011
BT	0.0994	0.1066	0.8058
MTF_GLP_CBD	0.0778	0.1041	0.8272
MMMT	0.0689	0.1024	0.8366
GS	0.0507	0.1106	0.8452
MTF_GLP_HPM	0.1011	0.1363	0.7782
ASIM	0.0803	0.1046	0.8241
DRPNN	0.0644	0.0836	0.8583
MSDCNN	0.0554	0.0849	0.8649
PCDRN	<b>0.0279</b>	<b>0.0450</b>	<b>0.9285</b>

\* The EXP method is an interpolation method for MS Image, which is not a pansharpening method.

## 5. Further Discussion

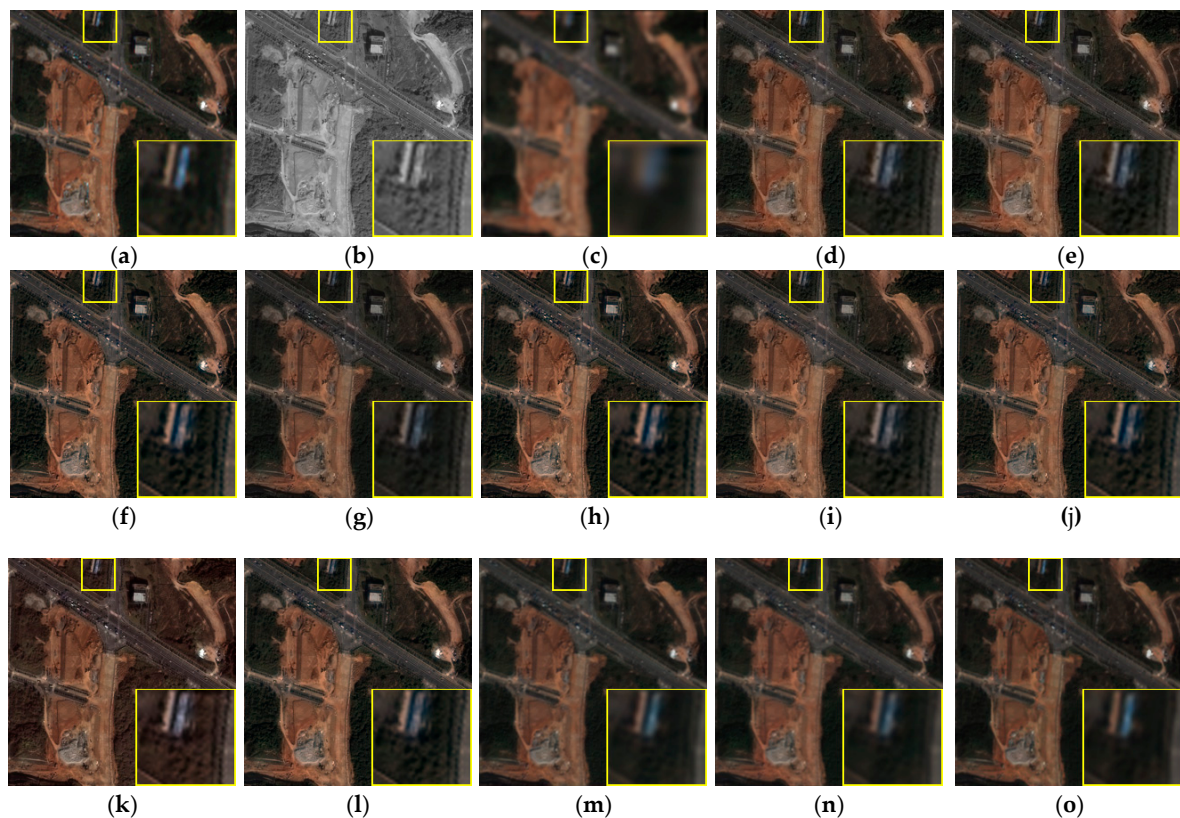
To further verify the performance of PCDRN, experiments were performed on 90 groups of the simulated images, which were selected from another scene of the Pléiades dataset. It should be noted that these images were not used to train the PCDRN. An example of the simulated experiment performed on the Pléiades dataset is shown in Figure 11. Figure 11a,b are the reference image and the degraded PAN image, respectively. The upsampled LRMS using EXP method is shown in Figure 11c, which is blurred because of not using the PAN image. Figure 11d–o displays the corresponding fused results. In Figure 11d–o, the results of AIHS, ATWT, GSA, BT, MTF\_GLP\_CBD, MMT, GS, MTF\_GLP\_HPM, and ASIM methods not only suffer from serious spectral distortion but also exhibit some artifacts. DRPNN and MSDCNN methods produce some spectral distortion. As we can observe, the result of PCDRN is still closest to the reference image in 12 pansharpening methods by visual inspection.

Table 12 tabulates the average quantitative results on 90 groups of the simulated dataset from Pléiades. In Table 12, we can also find that PCDRN achieves the best values in 6 image quality indexes, which can again prove the proposed PCDRN is effective.

**Table 12.** Average quantitative results on 90 groups of the simulated dataset from Pléiades.

Methods	PSNR (↑)	CC (↑)	UIQI (↑)	Q2 <sup>n</sup> (↑)	SAM (↓)	ERGAS (↓)
EXP *	27.1185	0.9287	0.9286	0.7936	3.3655	4.9278
AIHS	27.2114	0.9276	0.9273	0.7744	4.2540	5.0893
ATWT	27.3592	0.9231	0.9343	0.7742	4.5509	5.0941
GSA	26.2029	0.9143	0.9286	0.7534	5.3822	5.8476
BT	23.2679	0.9145	0.8656	0.6511	5.0901	9.0333
MTF_GLP_CBD	26.4855	0.9177	0.9313	0.7586	5.2536	5.6210
MMMT	27.2816	0.9275	0.9353	0.8046	4.5076	4.9474
GS	26.4368	0.9136	0.9164	0.7423	4.8918	5.6067
MTF_GLP_HPM	21.9264	0.8428	0.8587	0.6403	6.1466	7.9485
ASIM	27.6697	0.9336	0.9400	0.8234	4.2760	4.8197
DRPNN	28.7108	0.9692	0.9525	0.8474	3.5782	4.3122
MSDCNN	27.4453	0.9538	0.9405	0.8234	3.7876	4.9131
PCDRN	<b>29.5375</b>	<b>0.9853</b>	<b>0.9603</b>	<b>0.8691</b>	<b>3.2530</b>	<b>4.0998</b>

\* The EXP method is an interpolation method for MS Image, which is not a pansharpening method.



**Figure 11.** An example of the simulated experiment on Pléiades dataset. (a) Reference image; (b) PAN; (c) EXP; (d) AIHS; (e) ATWT; (f) GSA; (g) BT; (h) MTF\_GLP\_CBD; (i) MMMT; (j) GS; (k) MTF\_GLP\_HPM; (l) ASIM; (m) DRPNN; (n) MSDCNN; and (o) PCDRN.

## 6. Conclusions

In this paper, we presented a new deep learning-based approach for pansharpening, i.e., PCDRN. Different from other pansharpening approaches, we interpolated LRMS images twice and extracted features from source images using two residual subnetworks to inject details in two sizes. To avoid the over-smooth phenomenon, we design a multitask loss function to train our network to achieve high-quality remote sensing image fusion. To eliminate checkerboard artifacts, a resize-convolution consisting of a nearest-neighbor interpolation and a convolution layer was employed instead of transposed convolution for upsampling. Compared with other pansharpening methods, the experimental results demonstrated that PCDRN exhibits the best performance. In the future, we intend to develop an adaptive method of the multitask loss function to preserve additional spatial information.

**Author Contributions:** Y.Y. and S.H. conceived and designed the study; Y.Y. and W.T. drafted the manuscript; W.T. and H.L. conducted the experiments; Y.Y. and S.H. performed manuscript review. All authors have read and agree to the published version of the manuscript.

**Funding:** This work is supported by the National Natural Science Foundation of China (No. 61662026, No.61862030, and No.61562031), by the Natural Science Foundation of Jiangxi Province (No. 20182BCB22006, No. 20181BAB202010, No.20192ACB20002, and No.20192ACBL21008), and by the Project of the Education Department of Jiangxi Province (No. YC2018-B065).

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



## References

1. Ghassemian, H. A review of remote sensing image fusion methods. *Inf. Fusion* **2016**, *32*, 75–89. [[CrossRef](#)]
2. Zhang, Y.; Hong, G. An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural colour IKONOS and QuickBird images. *Inf. Fusion* **2005**, *6*, 225–234. [[CrossRef](#)]
3. Tu, T.M.; Su, S.C.; Shyu, H.C.; Huang, P.S. A new look at IHS like image fusion methods. *Inf. Fusion* **2001**, *2*, 177–186. [[CrossRef](#)]
4. Vivone, G.; Alparone, L.; Chanussot, J.; Mura, M.D.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A Critical Comparison Among Pansharpening Algorithms. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2565–2586. [[CrossRef](#)]
5. Wang, Z.; Ziou, D.; Armenakis, C.; Li, D.; Li, Q. A comparative analysis of image fusion methods. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1391–1402. [[CrossRef](#)]
6. Laben, C.A.; Brower, B.V. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpener. US Patent 6,011,875, 4 January 2000.
7. Loncan, L.; Almeida, L.B.; Bioucas-Dias, J.M.; Briottet, X.; Chanussot, J.; Dobigeon, N.; Fabre, S.; Liao, W.; Licciardi, G.A.; Simoes, M.; et al. Hyperspectral Pansharpening: A Review. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 27–46. [[CrossRef](#)]
8. Miao, Q.; Wang, B. Multi-sensor image fusion based on improved laplacian pyramid transform. *Acta Opt. Sin.* **2007**, *27*, 1605–1610.
9. Nunez, J.; Otazu, X.; Fors, O.; Prades, A.; Pala, V.; Arbiol, R. Multiresolution-based image fusion with additive wavelet decomposition. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1204–1211. [[CrossRef](#)]
10. Li, S.; Kwok, J.T.; Wang, Y. Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images. *Inf. Fusion* **2002**, *3*, 17–23. [[CrossRef](#)]
11. Yang, Y.; Tong, S.; Huang, S.; Lin, P. Multi-focus Image Fusion Based on NSCT and Focused Area Detection. *IEEE Sens. J.* **2015**, *15*, 2824–2838. [[CrossRef](#)]
12. Yuan, Q.; Wei, Y.; Meng, X.; Shen, H.; Zhang, L. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2018**, *11*, 978–989. [[CrossRef](#)]
13. Fasbender, D.; Radoux, J.; Bogaert, P. Bayesian data fusion for adaptable image pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1847–1857. [[CrossRef](#)]
14. Joshi, M.; Jalobeanu, A. MAP estimation for multiresolution fusion in remotely sensed images using an IGMRF prior model. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 1245–1255. [[CrossRef](#)]
15. Cheng, J.; Liu, H.; Liu, T.; Wang, F.; Li, H. Remote sensing image fusion via wavelet transform and sparse representation. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 158–173. [[CrossRef](#)]
16. Yang, Y.; Wan, W.; Huang, S.; Lin, P.; Que, Y. A Novel Pan-Sharpener Framework Based on Matting Model and Multiscale Transform. *Remote Sens.* **2017**, *9*, 391. [[CrossRef](#)]
17. Masi, G.; Cozzolino, D.; Verdoliva, L.; Scarpa, G. Pansharpening by convolutional neural networks. *Remote Sens.* **2016**, *8*, 594. [[CrossRef](#)]
18. Yang, J.; Fu, X.; Hu, Y.; Huang, Y.; Ding, X.; Paisley, J. PanNet: A deep network architecture for pan-sharpening. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
19. Wei, Y.; Yuan, Q. Deep residual learning for remote sensed imagery pansharpening. In Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 18–21 May 2017.
20. Shao, Z.; Cai, J. Remote sensing image fusion with deep convolutional neural network. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2018**, *11*, 1656–1669. [[CrossRef](#)]
21. He, L.; Rao, Y.; Li, J.; Chanussot, J.; Plaza, A.; Zhu, J.; Li, B. Pansharpening via Detail Injection Based Convolutional Neural Networks. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2019**, *12*, 1188–1204. [[CrossRef](#)]
22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
23. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [[CrossRef](#)]

24. Nezhad, Z.H.; Karami, A.; Heylen, R.; Scheunders, P. Fusion of hyperspectral and multispectral images using spectral unmixing and sparse coding. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2016**, *9*, 2377–2389. [[CrossRef](#)]
25. Dumoulin, V.; Visin, F. A Guide to Convolution Arithmetic for Deep Learning [Online]. Available online: <https://arxiv.org/abs/1603.07285> (accessed on 5 February 2020).
26. Chong, N.; Wong, L.; See, J. GANmera: Reproducing aesthetically pleasing photographs using deep adversarial networks. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–20 June 2016.
27. 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](#)]
28. Alparone, L.; Wald, L.; Chanussot, J.; Thomas, C.; Gamba, P.; Bruce, L.M. Comparison of pansharpening algorithms: Outcome of the 2006 GRSS data fusion contest. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3012–3021. [[CrossRef](#)]
29. Wald, L. Quality of high resolution synthesised images: Is there a simple criterion? In Proceedings of the 3rd Conference Fusion Earth Data: Merging Point Measurement, Raster Maps and Remotely Sensed Images, Sophia Antipolis, France, 26–28 January 2000.
30. Garzelli, A.; Nencini, F. Hypercomplex quality assessment of multi-/hyper-spectral images. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 662–665. [[CrossRef](#)]
31. 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](#)]
32. 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.
33. Selva, M.; Santurri, L.; Baronti, S. On the use of the expanded image in quality assessment of pansharpened images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 320–324. [[CrossRef](#)]
34. Yang, Y.; Nie, Z.; Huang, S.; Lin, P.; Wu, J. Multilevel features convolutional neural network for multifocus image fusion. *IEEE Trans. Comput. Imag.* **2019**, *5*, 262–273. [[CrossRef](#)]
35. Rahmani, S.; Strait, M.; Merkurjev, D.; Moeller, M.; Wittman, T. An adaptive IHS pan-sharpening method. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 746–750. [[CrossRef](#)]
36. 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](#)]
37. Gillespie, A.; Kahle, A.; Walker, R. Color enhancement of highly correlated images. II. Channel ratio and “Chromaticity” transformation techniques. *Remote Sens. Environ.* **1987**, *22*, 343–365. [[CrossRef](#)]
38. 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](#)]
39. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A.; Selva, M. An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas. In Proceedings of the 2nd GRSS/ISPRS Joint Workshop Remote Sens. Data Fusion URBAN Areas, Berlin, Germany, 22–23 May 2003.
40. Yang, Y.; Wu, L.; Huang, S.; Tang, Y.; Wan, W. Pansharpening for multiband images with adaptive spectral-intensity modulation. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2018**, *11*, 3196–3208. [[CrossRef](#)]
41. 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](#)]
42. 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](#)]

