



Article

Remote Sensing Image Super-Resolution Based on Dense Channel Attention Network

Yunchuan Ma ¹, Pengyuan Lv ^{1,*} , Hao Liu ¹ , Xuehong Sun ¹ and Yanfei Zhong ²

¹ School of Information Engineering, Ningxia University, Yinchuan 750021, China; rs_student@nxu.edu.cn (Y.M.); liuhao@nxu.edu.cn (H.L.); sunxh@nxu.edu.cn (X.S.)

² State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China; zhongyanfei@whu.edu.cn

* Correspondence: lpydtc@nxu.edu.cn

Abstract: In the recent years, convolutional neural networks (CNN)-based super resolution (SR) methods are widely used in the field of remote sensing. However, complicated remote sensing images contain abundant high-frequency details, which are difficult to capture and reconstruct effectively. To address this problem, we propose a dense channel attention network (DCAN) to reconstruct high-resolution (HR) remote sensing images. The proposed method learns multi-level feature information and pays more attention to the important and useful regions in order to better reconstruct the final image. Specifically, we construct a dense channel attention mechanism (DCAM), which densely uses the feature maps from the channel attention block via skip connection. This mechanism makes better use of multi-level feature maps which contain abundant high-frequency information. Further, we add a spatial attention block, which makes the network have more flexible discriminative ability. Experimental results demonstrate that the proposed DCAN method outperforms several state-of-the-art methods in both quantitative evaluation and visual quality.



Citation: Ma, Y.; Lv, P.; Liu, H.; Sun, X.; Zhong, Y. Remote Sensing Image Super-Resolution Based on Dense Channel Attention Network. *Remote Sens.* **2021**, *13*, 2966. <https://doi.org/10.3390/rs13152966>

Academic Editor: Jon Atli Benediktsson

Received: 6 July 2021
Accepted: 23 July 2021
Published: 28 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: remote sensing images; super resolution; dense network; attention mechanism

1. Introduction

High-resolution (HR) remote sensing images provide detailed geometric information about land cover. Thus, HR remote sensing images are essential for many applications, such as object detection [1,2], urban planning [3], building extraction [4–6] and so on [7–15]. However, the spatial resolution of remote sensing images are influenced by the limitations of hardware and environmental factors [16–18]. Compared to current physical imaging technology, super resolution (SR) which recovers HR images from low-resolution (LR) images is more convenient and low cost. Thus, SR becomes an alternative method in the remote sensing field.

SR can be classified into single image super resolution (SISR) and multiple image super resolution (MISR) [19] by the number of input images. MISR methods utilize multiple LR images of the same area, to provide more information, to better reconstruct high spatial-frequency details and texture [20]. Nevertheless, it is difficult to obtain multiple remote sensing images of the same scene. Thus, this paper considers a study of SISR methods. The traditional SISR methods [21] are mainly grouped into interpolation-based methods and reconstruction-based methods [22]. The interpolation-based methods predict the unknown pixel using a simple linear or non-linear interpolation [23] operation. Although interpolation-based methods are convenient, they have limited performance for images which contain more details. The reconstruction-based methods utilize a certain kind of prior information to produce better results, such as local, nonlocal and sparse priors [23,24]. Although reconstruction-based methods [25] are flexible and allow consideration of different prior constraints [26–31], they also have difficulties with complicated remote sensing images. Recently, deep learning-based methods attracted a lot of attention

in the remote sensing images super resolution task. In 2015, a super resolution convolutional neural network (SRCNN) [32] was first proposed by Dong et al. to achieve the natural images super resolution. As the pioneer, SRCNN learned the mapping between the HR images and the corresponding LR images using a three layers network. While SRCNN had outperformed traditional-based methods, bicubic LR images made the network to operate in a high-dimensional space and largely increased the computational cost. To alleviate the problem, a fast super resolution convolutional neural network (FSRCNN) [33] and an efficient subpixel convolutional network (ESPCN) [34] were proposed which used a deconvolutional layer and a subpixel convolutional layer to achieve direct reconstruction in a low-dimensional space and save the computational cost, respectively. These networks were shallow and their performance were limited by the network depth. However, increasing the network depth leads to a vanishing gradient and an exploding gradient. To handle the problem, the skip connection operation [35] was proposed by He et al., which combined the low-level features and high-level features to effectively alleviate the gradient vanishing. Thus, the skip connection operation was gradually used in the SR networks. Among these networks, the very deep super resolution convolutional network (VDSR) [36] which used a global residual connection was proposed by Kim et al. to spread the LR information to the network end. It was the first network which introduced the residual learning to SR and succeeded training a 20-layer network. Besides, enhanced deep super resolution (EDSR) [37] and SRResNet [38] also used global residual connection. In addition, EDSR and SRResNet employed residual blocks as basic network module, which introduced local residual connection to ease the deep network training difficulty. Later, Zhang et al. [39] constructed a residual in residual structure where residual blocks compose residual groups using short and long skip connections. Further, cascading residual network (CARN) [40], dense deep back-projection network (D-DBPN) [41], SRDenseNet [42], residual dense network (RDN) [43], employed the dense skip connections or multiple skip connections to increase the training effect.

Deep learning-based SR methods in the field of remote sensing also developed fast in recent years. In 2017, a local-global combined network (LGC) [19] was first proposed by Lei et al. to enhance the spatial resolution of remote sensing images. LGC learns multi-level information including local details and global priors using the skip connection operation. In 2018, a residual dense backprojection network (RDBPN) [22] was proposed by Pan et al., which consists of several residual dense backprojection blocks that contain the upprojection module and the downprojection module. In 2020, Zhang et al. proposed a scene adaptive method [44] via a multi-scale attention network to enhance the SR reconstruction details under the different remote sensing scenes. Recently, an approach named dense-sampling super resolution network (DSSR) [45] presented a dense sampling mechanism which reuses an upscaler to upsample and overcome the large-scale remote sensing images SR reconstruction problem. However, the complex spatial distribution of remote sensing images need more attention. In 2020, a second-order multi-scale super resolution network (SMSR) [46] was proposed by Dong et al. to reuse the learned multi-level information to the high-frequency regions of remote sensing images. The multi-perception attention network (MPSR) [47] and the multi-scale residual neural network (MRNN) [48] are also doing some related work about using multi-scale information. In addition, the generative adversarial network (GAN)-based SR method is used to generate visually pleasing remote sensing images. In 2019, Jiang et al. presented an edge-enhancement generative adversarial network (EEGAN) [49], which introduces an edge enhancement module to improve the remote sensing images SR performance. In 2020, Lei et al. proposed a coupled-discriminated generative adversarial network (CDGAN) [50] for solving the discrimination-ambiguity problem for the low-frequency regions in the remote sensing images.

Although the above-mentioned methods have good performance, their results can be further improved. First, the distributions of remote sensing images are very complex; therefore, we need more high-frequency details and texture to better reconstruct HR images. Secondly, redundancy feature information are not beneficial to recover details and increase

computation cost. So, we propose a dense channel attention network (DCAN) which learns multi-level feature information and pays more attention to the important and useful regions in order to better reconstruct the final image. The major contributions are as follows:

- (1) We propose a DCAN for SR of the single remote sensing image, which makes full use of the features learned at different depths through densely using multi-level feature information and pay more attention to high-frequency regions. Both quantitative and qualitative evaluations demonstrate the superiority of DCAN over the state-of-the-art methods.
- (2) A dense channel attention mechanism (DCAM) is proposed to utilize the channel attention block through the dense skip connection manner. This mechanism can increase the flow of information through the network and improve the representation capacity of the network.
- (3) A spatial attention block (SAB) is added to the network. This helps the network have more flexible discriminative ability for different local regions. It contributes to reconstruct the final image. In addition, this helps the network have more flexible discriminative ability for global structure and focus on high-frequency information from the spatial dimension.

2. Method

2.1. Network Architecture

The architecture of our proposed DCAN is illustrated in Figure 1, which consists of three parts: shallow feature extraction, deep feature extraction, and reconstruction. The network first extracts the shallow features from the input LR image. Then, the second part extracts the deep features and increase the weights of important feature maps. Finally, the features which contain abundant useful information are sent to the third part of network for reconstruction. The network details will be introduced in the next part.

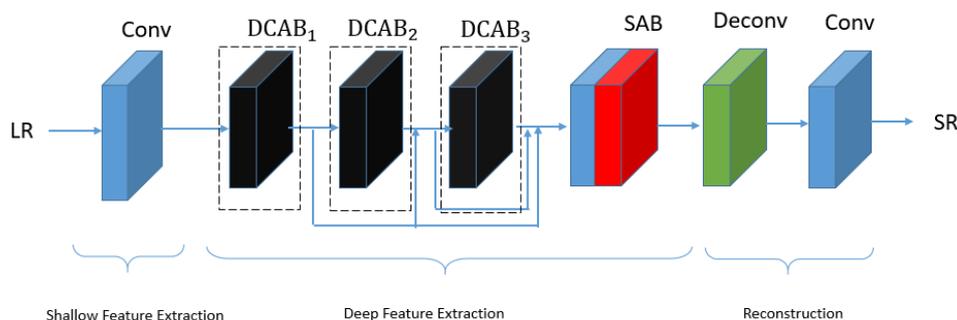


Figure 1. The structure of the proposed dense channel attention network (DCAN).

- (1) Shallow Feature Extraction: Let $Conv(k, f, c)$ be a convolutional layer, where k, f, c represent the filter kernel size, the number of filters, and the number of filter channels, respectively. We use a 3×3 convolutional layer to extract the shallow feature F_0 from the input image I^{LR} , which contributes to the next feature extraction. The shallow feature extraction $f_{SF}(\cdot)$ operation can be formulated as follows:

$$F_0 = f_{SF}(I^{LR}) = Conv(k, f, c)(I^{LR}) \quad (1)$$

- (2) Deep Feature Extraction: After the shallow feature extraction, the backbone part which contains a series of Dense Channel Attention blocks ($DCAB_s$) and a Spatial Attention Block (SAB) is designed to extract deep features. The main block DCAB receives the feature maps from each single DCAB as input, the structure of which will

be given in Section 2.2. Then, the feature F_G which is the sum of output features of G DCAB_s is sent to a convolutional layer:

$$F_{G'} = \text{Conv}(k, f, c)(F_G) \quad (2)$$

where $F_{G'}$ is generated by the convolution operation, G denotes the number of DCAB_s. Then, the $F_{G'}$ is sent to a SAB as follows:

$$F_{\text{SAB}} = H_{\text{SAB}}(F_{G'}) \quad (3)$$

where H_{SAB} denotes the operation of the SAB. The operational details will be described in Section 2.3.

- (3) Reconstruction: The reconstruction part contains two convolutional layers and a deconvolution layer. The SR image I^{SR} is generated as follows:

$$I^{\text{SR}} = \text{Conv}(k, f, c)(H_{\text{Up}}(\text{Conv}(k, f, c)(F_{\text{SAB}}))) \quad (4)$$

where H_{up} denotes the operation of the deconvolution layer.

2.2. Dense Channel Attention Mechanism

As discussed in Section 1, most existing super-resolution models do not make full use of the information from the input LR image. So, we propose a novel dense channel attention mechanism to solve the problem. It concentrates on the high-frequency information and weakens the useless information. Figure 2 shows DCAM, which can be described that the G_{th} DCAB computes the feature map F_G from the outputs of the DCAB_{G-1}, DCAB_{G-2}, ..., DCAB₁ as follows:

$$F_G = H_{\text{DCAB}_G}(F_{G-1} + F_{G-2} + \dots + F_1) \quad (5)$$

where H_{DCAB_G} denotes the operation of the G_{th} DCAB, $F_{G-1}, F_{G-2}, \dots, F_1$ denote the outputs of the DCAB_{G-1}, DCAB_{G-2}, ..., DCAB₁. The purpose of DCAM is to focus on high-frequency components and make better use of the information.

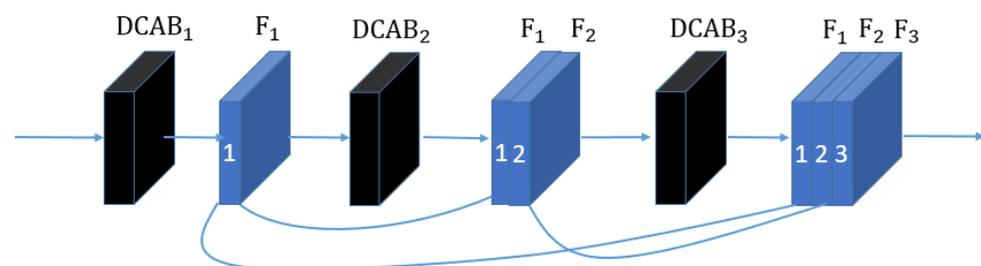


Figure 2. The structure of dense channel attention mechanism (DCAM).

As shown in Figure 2, the DCAB is the basic block of the proposed DCAM. It mainly contains two convolutional layers and a CA (Channel Attention) block. To be specific, in Figure 3, the first convolutional layer before Relu consists of n_f filters of size $n_c \times 3 \times 3$. The second convolutional layer after Relu contains n_f filters of size $n_f \times 3 \times 3$. Let F_{i-1} be the input of DCAB_i (the i -th DCAB in the DCAB_s), there is

$$X_i = W^{(3, n_f, n_f)} * \sigma(W^{(3, n_f, n_c)} * F_{i-1}) \quad (6)$$

where X_i is an intermediate feature containing n_f feature maps. $W^{(3, n_f, n_f)}$ and $W^{(3, n_f, n_c)}$ denote the weight matrix of n_f filters of size $n_f \times 3 \times 3$ and n_f filters of size $n_c \times 3 \times 3$. $\sigma(\cdot) = \max(0, x)$ indicates the Relu activation.

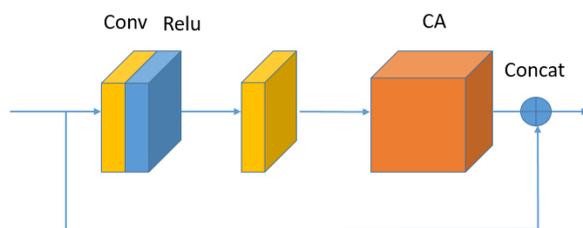


Figure 3. The structure of dense channel attention block (DCAB).

Then, CA is used to improve the discriminative learning ability. As shown in Figure 4, the mechanism of CA can be formulated as follows:

$$X'_i = X_i \otimes S_i = [X_1^i, \dots, X_{n_f}^i] \otimes [S_1^i, \dots, S_{n_f}^i] \tag{7}$$

where X'_i is the output of X_i , $X_1^i, \dots, X_{n_f}^i$ is the feature map of X_i , $S_1^i, \dots, S_{n_f}^i$ are the elements of S_i , \otimes indicates the elementwise product. S_i is a n_f dimensional channel statistical descriptor, which is used to update X_i . S_i is obtained by

$$S_i = f_{sigmoid}(f_{up}(\sigma(f_{down}(f_{gap}(X_i))))) \tag{8}$$

It contains several operations, including global average pooling $f_{gap}()$, channel-down $f_{down} = Conv(1, n_f/r, n_f)$, Relu activation $\sigma()$, channel-up $f_{up} = Conv(1, n_f, n_f/r)$, and sigmoid function $f_{sigmoid}$, on X_i (r is set to 16). The channel statistical descriptor can help express the different information among the feature maps of X_i .

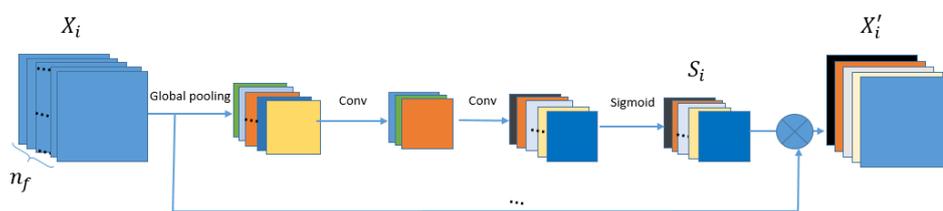


Figure 4. The structure of Channel Attention Block (CA).

As shown in Figure 3, the input of the i_{th} DCAB is from the output of i_1, \dots, i_{th-1} DCABs. In general, the complete operation of DCAB can be formulated as follows:

$$F_i = S_i \otimes (W^{(3, n_f, n_f)} * \sigma(W^{(3, n_f, n_c)} * F_{i-1})) + F_{i-1} \tag{9}$$

This operation increases the flow of information through the network and the representation capacity of the network.

2.3. Spatial Attention Block

Considering the complicated spatial information and distribution of the remote sensing images, we add a SAB to increase the discriminative ability of the network. It helps the network have discriminative ability for different local regions and pay more attention to the regions which are more important and more difficult to reconstruct. As shown in Figure 5, the operation of SAB can be formulated as follows:

$$F_{SAB} = F_{input} \otimes f_{sigmoid}(f_{conv}(f_{concat}(f_{Avgpooling}(F_{input}), f_{Maxpooling}(F_{input})))) \tag{10}$$

where F_{SAB} is obtained by several operations, including average pooling $f_{Avgpooling}()$, $f_{Maxpooling}()$, $concat$ $f_{concat}()$, channel-down $f_{conv}() = Conv(1, 1, 2)$, sigmoid function $f_{sigmoid}()$, and element-wise product σ . The operation of SAB can focus on the local regions of F_{input} which are useful to reconstruct.

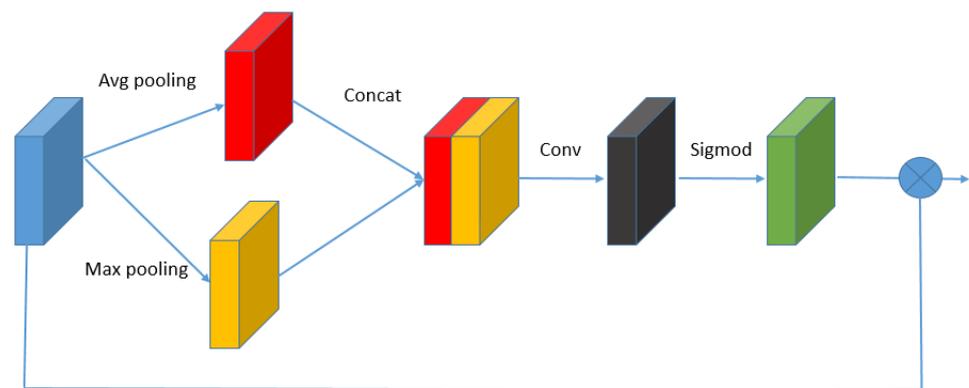


Figure 5. The structure of Spatial Attention Block (SAB).

2.4. Loss Function

We use the L_1 loss as the total loss because the L_1 loss can support better convergence. In addition, the L_1 loss can be described as follows:

$$L_1(\theta) = \frac{1}{n} \sum_{i=1}^n \|f_{\text{DCAN}}(I_i^{\text{LR}}; \theta) - I_i^{\text{HR}}\| \quad (11)$$

where θ represents the whole parameter of the DCAN network, n represents the number of training images. The purpose of the L_1 loss function is that the reconstructed HR image $I_i^{\text{SR}} = f_{\text{DCAN}}(I_i^{\text{LR}}; \theta)$ can be similar to its corresponding ground truth image I_i^{HR} .

3. Results

3.1. Experimental Settings

The relevant experimental settings on the experimental datasets, degradation method, and evaluation metrics are detailed in this section.

(1) Datasets: We use UC Merced dataset and RSSCN7 dataset for qualitative and quantitative analysis. The UC Merced dataset [51] is a classification dataset of remote sensing images containing 21 land use classes, and each class contains 100 images with size 256×256 and RGB channels. Figure 6 shows some images. The UC Merced dataset is mainly used as the experimental data and divided into two sections. The first section includes images in 21 classes of agricultural, baseball diamond, beach, building and so on. For each category, 90 images are taken to create the training set. Therefore, the second section includes the remaining ten images for each class which are taken to construct the test set. We can validate the performance of our model for each class. The results are discussed in Section 3.2.

In addition, the other dataset named RSSCN7 [52] is also used to train our method and verify the effectiveness of our method. The RSSCN7 dataset is a classification dataset of remote sensing images containing 7 land use classes, and each class contains 400 images with size 400×400 and RGB channels. Figure 7 shows some images. This dataset is divided into two sections. The first section contains images in 7 classes. For each class, 360 images are used to train the model. Thus, the second section includes the remaining 40 images for each class which are used to construct the test set. The test results are discussed in Section 3.2.



Figure 6. Some images in the UC Merced dataset: 21 land use classes, including buildings, agricultural, airplane, baseball diamond, beach, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis court. The above images correspond to the above categories respectively.

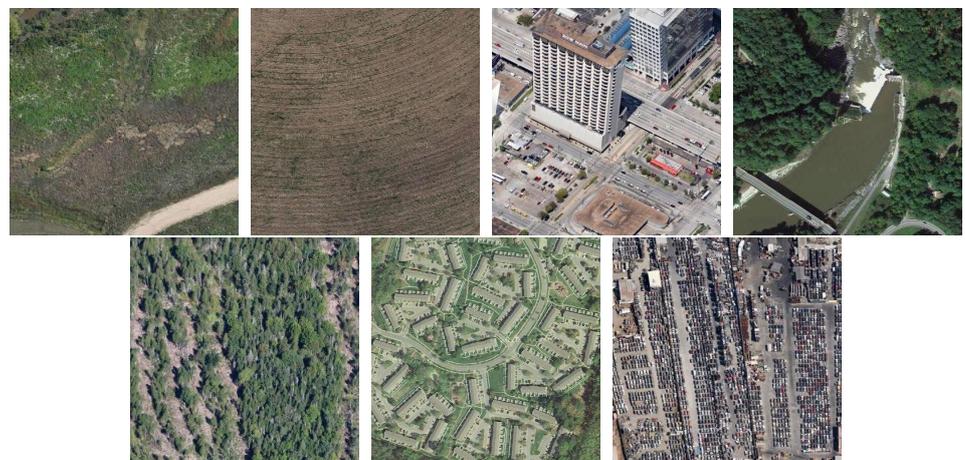


Figure 7. Some images in the RSSCN7 dataset: 7 land use classes, including grass, field, industry, river lake, forest, residential and parking. The above images correspond to the above categories respectively.

To validate the robustness of the proposed method, the real world data from GaoFen-2 in Ningxia, China, are used to test our model in Section 3.4. Experiments are designed by the $\times 4$ scale factor and the $\times 8$ scale factor, respectively. We obtain LR images by down sampling HR images using the bicubic operation.

All the parameter settings are same on the UC Merced dataset and RSSCN7 dataset. The training step is performed on the three channels of the RGB space. The channel number of the input image n_c is 3, the filter size k is set as 3. The Adam [53] optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ is used to train the proposed models under the batchsize of 16. The weights of the model are initialized using the method in [54]. We initialize the learning rate as 10^{-4} and halve it at every 2×10^5 batches updates. We implement the proposed algorithm with the PyTorch [55] framework, and we train the DCAN model using one NVIDIA Tesla V100 GPU.

(2) Evaluation Metrics: Two widely used image quality assessment metrics, the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used as evaluation metrics. They are used as the criteria for evaluating the quality of reconstructed HR images. Given the ground truth HR image x and its corresponding super-resolved image y , the PSNR of the super-resolved image y is computed as follows:

$$\text{MSE}(x, y) = \frac{1}{n_s} \sum_{i=1}^{n_s} (x_i - y_i)^2 \quad (12)$$

$$\text{PSNR}(x, y) = 10 \log_{10} \frac{255^2}{\text{MSE}(x, y)} \quad (13)$$

where x_i and y_i denote the i th pixel in x and y values, respectively. n_s represents the image pixel number. The higher PSNR indicates that the quality of the SR image is better. In addition to using PSNR to measure image quality, we also use SSIM as the image quality assessment, the SSIM of the super-resolved image y is defined as follows:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (14)$$

where μ_x and μ_y denote the mean values of x and y , and σ_x and σ_y denote the standard deviation values of x and y , and C_1 and C_2 are constants. The higher SSIM indicates that the quality of the SR image is better.

3.2. Comparisons with the Other Methods

The proposed DCAN method is compared with four state-of-the-art methods: Bicubic, SRCNN [32], VDSR [36], SRResNet [38]. In this experiment, we use the same training datasets and test datasets to compare these methods fairly. In addition, these models are trained well to test under the suitable conditions. Figures 8–15 show some reconstructed images of the UC Merced test set using the corresponding algorithm and the ground truth HR images in the visual quality.

Figure 8 shows the super-resolved HR images obtained through different models for “airplane89.jpg” with a scale factor of $\times 4$. The figure demonstrates that the reconstructed result of our DCAN model is better than those of the other algorithms, such as the edge of the plane. The reconstructed images for “overpass18.jpg” with a scale factor of $\times 4$ are shown in Figure 9. It is obvious that DCAN reconstructed more details and recovered sharper edges, such as the traffic index line of the road and the car. Comparison results for “harbor32.jpg” of a $\times 8$ scale factor are shown in Figure 10. We can get the conclusion that our method successfully reconstructed the main part of the bridge. The reconstructed HR images of “storagetanks68.jpg” are compared in Figure 12. The results demonstrate that the DCAN method provided more enhanced textures and recovered more details in comparison with other algorithms. Figure 13 shows the reconstructed images of “mediumresidential34.jpg” of a $\times 4$ scale factor. It can be seen that the result of our proposed method reconstructed the edges sharper than the other methods. Figure 14 compares the reconstructed images of “runway39.jpg” with other methods. It is obvious that the proposed methods reconstructed the details better than the other algorithms, such as the thin strip lines at the right of the image. As shown in Figures 11 and 15, we enlarge some regions of super-resolved images to better compare the textures and details. The texture and details of our proposed method are superior than those of other methods. The above results show that our method achieves better results compared with other methods in terms of the visual quality.



Figure 8. \times 4 super resolved-images of “airplane89.jpg” on the UC Merced dataset via different algorithms, and the numbers under these images represent the PSNR and SSIM values. (a) represent HR image, (b–f) represent the super-resolved image of Bicubic, SRCNN, VDSR, SRResNet, DCAN.



Figure 9. \times 4 super-resolved images of “overpass18.jpg” on the UC Merced dataset via different algorithms, and the numbers under these images represent the PSNR and SSIM values. (a) represent HR image, (b–f) represent the super-resolved image of Bicubic, SRCNN, VDSR, SRResNet, DCAN.

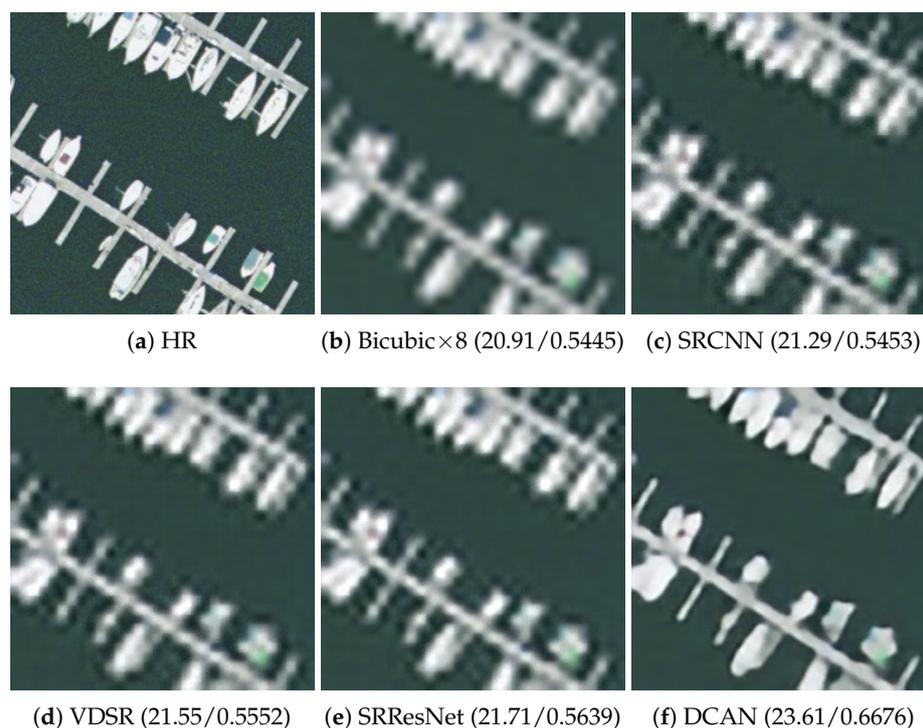


Figure 10. \times 8 super-resolved images of “harbor32.jpg” on the UC Merced dataset via different algorithms, and the numbers under these images represent the PSNR and SSIM values. (a) represent HR image, (b–f) represent the super-resolved image of Bicubic, SRCNN, VDSR, SRResNet, DCAN.

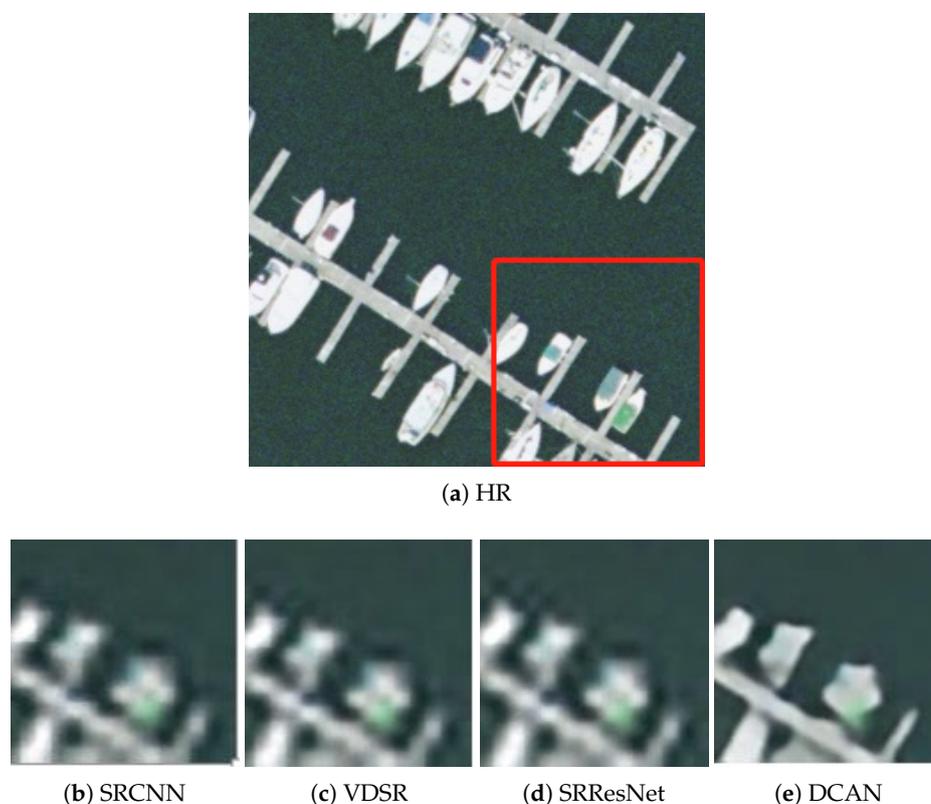


Figure 11. The detail of super-resolved “harbor32.jpg” region with a scale factor of \times 8. (a) represent HR image, (b–e) represent the super-resolved image patch of SRCNN, VDSR, SRResNet, DCAN.

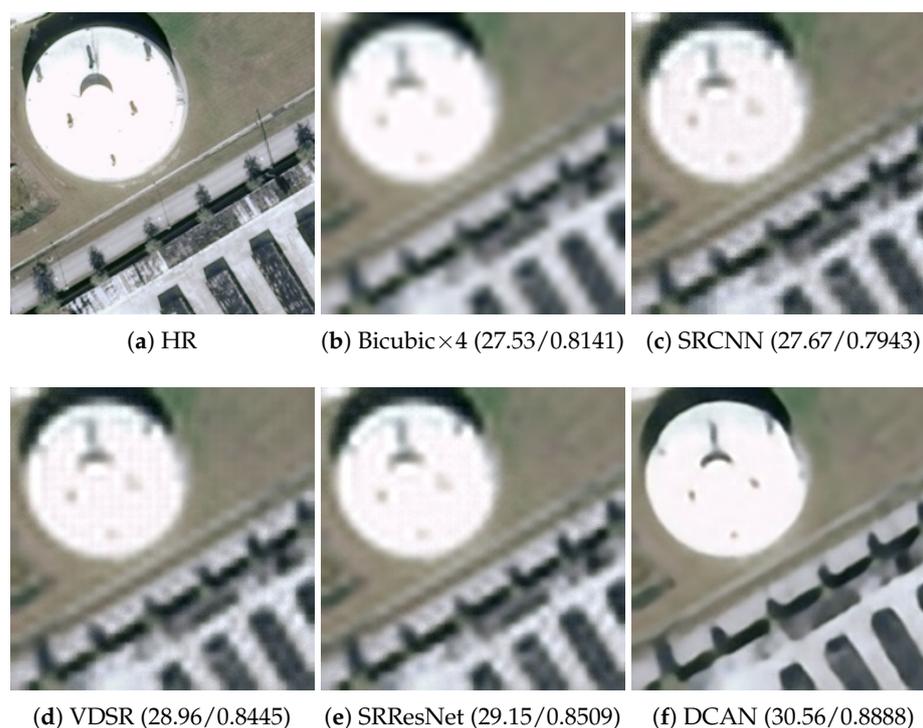


Figure 12. $\times 4$ super-resolved images of “storagetanks68.jpg” on the UC Merced dataset via different algorithms, and the numbers under these images represent the PSNR and SSIM values. (a) represent HR image, (b–f) represent the super-resolved image of Bicubic, SRCNN, VDSR, SRResNet, DCAN.



Figure 13. $\times 8$ super-resolved images of “mediumresidential34.jpg” on the UC Merced dataset via different algorithms, and the numbers under these images represent the PSNR and SSIM values. (a) represent HR image, (b–f) represent the super-resolved image of Bicubic, SRCNN, VDSR, SRResNet, DCAN.

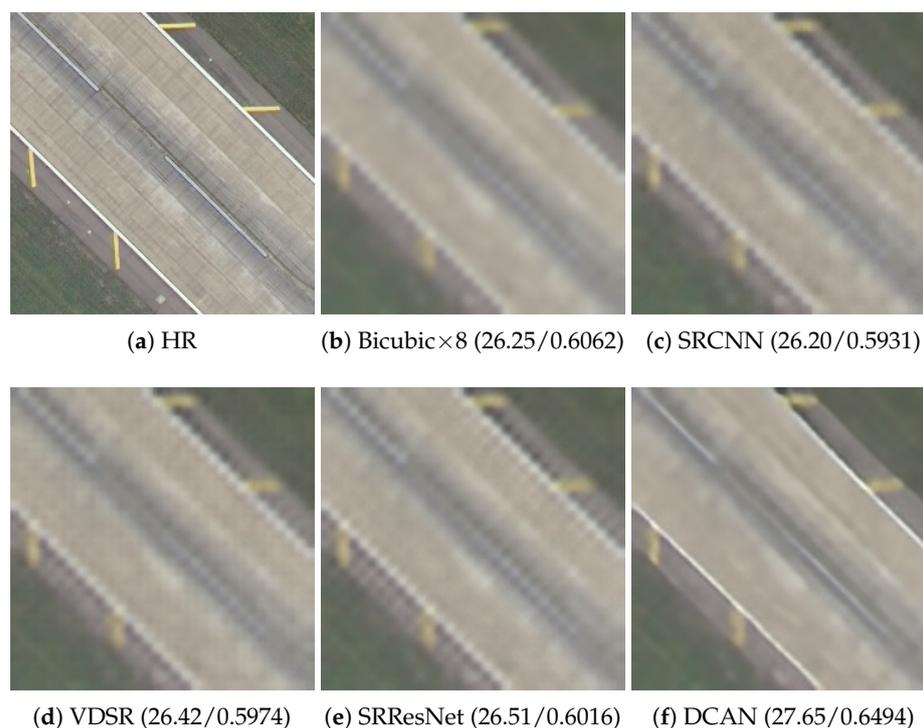
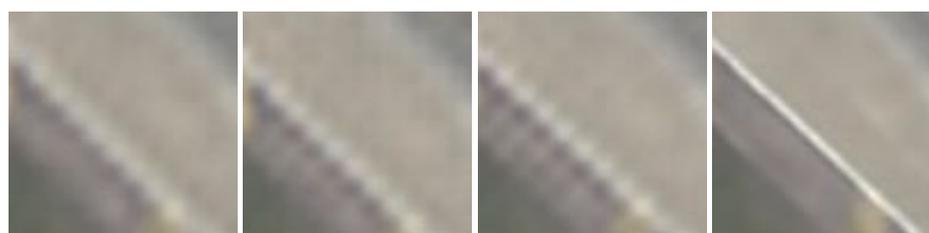


Figure 14. $\times 8$ super-resolved images of “runway39.jpg” on the UC Merced dataset via different algorithms, and the numbers under these images represent the PSNR and SSIM values. (a) represent HR image, (b–f) represent the super-resolved image of Bicubic, SRCNN, VDSR, SRResNet, DCAN.



(a) HR



(b) SRCNN

(c) VDSR

(d) SRResNet

(e) DCAN

Figure 15. The detail of super-resolved “runway39.jpg” region with a scale factor of $\times 8$. (a) represent HR image, (b–e) represent the super-resolved image patch of SRCNN, VDSR, SRResNet, DCAN.

Tables 1 and 2 show the PSNR and SSIM of the reconstructed HR images of the UC Merced test set on the $\times 4$ and $\times 8$ enlargement, respectively.

Table 1. The PSNR and SSIM of UC Merced test dataset with a scale factor of 4.

Data	Scale	Bicubic	SRCNN	VDSR	SRResNet	DCAN
Agricultural	$\times 4$	25.24/0.4526	25.66/0.4728	25.87/0.0.4753	25.74/0.4811	26.20/0.5332
Airplane	$\times 4$	25.83/0.7524	26.40/0.7654	27.93/0.8311	27.95/0.8192	28.57/0.8627
Baseball diamond	$\times 4$	30.32/0.7754	30.73/0.8003	31.48/0.8166	31.78/0.8195	32.37/0.8314
Beach	$\times 4$	33.16/0.8375	33.47/0.8507	33.85/0.8611	34.05/0.8665	34.26/0.8678
Buildings	$\times 4$	21.44/0.6602	22.11/0.6687	22.55/0.7449	23.62/0.7814	24.12/0.8217
Chaparral	$\times 4$	24.15/0.6512	24.78/0.6978	24.86/0.7146	25.21/0.7431	25.54/0.7438
Dense residential	$\times 4$	23.34/0.6721	23.85/0.6742	24.27/0.7211	25.43/0.7785	25.92/0.8014
Forest	$\times 4$	26.28/0.6010	26.47/0.6413	26.63/0.6572	26.89/0.6723	27.11/0.6914
Freeway	$\times 4$	26.23/0.6773	26.69/0.6914	27.11/0.7411	27.91/0.7708	28.46/0.7928
Golf course	$\times 4$	30.76/0.7711	31.22/0.7834	31.24/0.7912	31.53/0.8003	32.31/0.8077
Harbor	$\times 4$	17.38/0.6906	17.90/0.6991	18.58/0.8064	19.79/0.8453	21.03/0.8603
Intersection	$\times 4$	24.37/0.6951	24.81/0.6983	24.89/0.7437	25.82/0.7763	26.11/0.7883
Medium residential	$\times 4$	23.47/0.6534	23.96/0.6864	24.51/0.7281	25.42/0.7533	25.63/0.7682
Mobile homepark	$\times 4$	21.83/0.6631	22.45/0.6943	23.19/0.7516	24.31/0.7815	24.93/0.8006
Overpass	$\times 4$	23.14/0.6417	23.51/0.6463	24.06/0.6944	25.41/0.7469	25.94/0.7664
Parking lot	$\times 4$	19.21/0.6011	19.83/0.6213	19.52/0.6617	20.04/0.7148	21.08/0.7760
River	$\times 4$	26.41/0.6551	26.82/0.6904	26.99/0.7038	27.07/0.7109	27.37/0.7162
Runway	$\times 4$	26.56/0.7177	27.24/0.7401	28.21/0.7753	29.64/0.7962	30.53/0.8093
Sparse residential	$\times 4$	26.15/0.6704	26.64/0.7038	27.08/0.7263	27.65/0.7384	27.81/0.7463
Storage tanks	$\times 4$	23.57/0.6970	24.01/0.6998	24.43/0.7607	25.26/0.7875	25.53/0.8211
Tennis court	$\times 4$	28.03/0.7812	28.67/0.7816	28.99/0.8324	30.02/0.8514	30.46/0.8601

Table 2. The PSNR and SSIM of UC Merced test dataset with a scale factor of 8.

Data	Scale	Bicubic	SRCNN	VDSR	SRResNet	DCAN
Agricultural	$\times 8$	23.62/0.2711	23.74/0.2813	23.78/0.2817	23.61/0.2806	23.77/0.2874
Airplane	$\times 8$	22.98/0.6152	23.45/0.6321	23.89/0.6564	24.02/0.6583	24.33/0.6759
Baseball diamond	$\times 8$	27.15/0.6514	27.87/0.6661	27.98/0.6825	28.17/0.6856	28.41/0.6933
Beach	$\times 8$	30.31/0.7385	30.56/0.7383	30.94/0.7524	30.68/0.7574	31.02/0.7614
Buildings	$\times 8$	18.74/0.4497	19.14/0.4623	19.45/0.5209	19.92/0.5497	20.11/0.5687
Chaparral	$\times 8$	20.61/0.3165	20.53/0.3395	20.68/0.3567	20.93/0.3741	20.99/0.3977
Dense residential	$\times 8$	20.44/0.4617	20.92/0.4775	21.03/0.5236	21.55/0.5491	21.73/0.5702
Forest	$\times 8$	23.97/0.3741	24.14/0.4011	24.21/0.4018	24.13/0.4019	24.27/0.4147
Freeway	$\times 8$	23.84/0.4946	24.27/0.5271	24.59/0.5497	24.71/0.5533	24.95/0.5841
Golf course	$\times 8$	27.65/0.6491	28.15/0.6601	28.56/0.6753	28.63/0.6817	28.96/0.6893
Harbor	$\times 8$	14.84/0.5089	15.16/0.4997	15.29/0.5849	15.72/0.6243	15.91/0.6704
Intersection	$\times 8$	21.46/0.4880	21.89/0.5112	21.82/0.5194	22.36/0.5437	22.63/0.5722

Table 2. *Cont.*

Data	Scale	Bicubic	SRCNN	VDSR	SRResNet	DCAN
Medium residential	×8	20.40/0.4227	20.77/0.4416	20.84/0.4663	21.36/0.4999	21.55/0.5208
Mobile homepark	×8	18.27/0.4143	18.59/0.4361	18.78/0.4698	19.42/0.5068	19.61/0.5301
Overpass	×8	20.72/0.4182	20.97/0.4424	21.25/0.4659	21.67/0.4932	22.86/0.5321
Parking lot	×8	16.98/0.3771	17.21/0.3889	16.95/0.4057	17.19/0.4187	17.29/0.4416
River	×8	24.41/0.5025	24.75/0.5269	24.74/0.5311	24.77/0.5335	24.94/0.5435
Runway	×8	23.25/0.5567	23.51/0.5498	24.03/0.5849	24.84/0.6136	25.63/0.6371
Sparse residential	×8	22.93/0.4678	23.28/0.4915	23.41/0.5022	23.72/0.5168	23.95/0.5298
Storage tanks	×8	21.22/0.5356	21.68/0.5439	21.87/0.5835	22.16/0.6008	22.79/0.6451
Tennis court	×8	24.43/0.5968	25.13/0.6194	25.41/0.6456	25.63/0.6542	26.13/0.8078

Our proposed method is also better than other methods in the quantitative evaluation. Tables 3 and 4 show the PSNR and SSIM of the reconstructed HR images of RSSCN7 test set on the ×4 enlargement and ×8 enlargement, respectively. Experimental results demonstrate the effectiveness of our method.

Table 3. The PSNR and SSIM of RSSCN7 test dataset with a scale factor of 4.

Data	Scale	Bicubic	SRCNN	VDSR	SRResNet	DCAN
Grass	×4	32.69/0.8073	33.48/0.8132	34.04/0.8231	34.27/0.8294	34.39/0.8342
Industry	×4	24.42/0.6617	24.85/0.6651	25.76/0.7185	26.39/0.7414	26.61/0.7536
River lake	×4	29.89/0.8014	30.33/0.8062	30.91/0.8231	31.17/0.8294	31.35/0.8346
Filed	×4	31.83/0.7228	32.59/0.7332	33.10/0.7527	33.37/0.7528	33.51/0.7683
Forest	×4	27.79/0.5996	28.06/0.6072	28.27/0.6271	28.38/0.6403	28.53/0.6421
Resident	×4	23.82/0.6431	24.11/0.6453	24.87/0.6931	25.39/0.7182	25.59/0.7287
Parking	×4	23.76/0.6187	24.34/0.6332	25.12/0.6761	25.57/0.7064	25.82/0.7138

Table 4. The PSNR and SSIM of RSSCN7 test dataset with a scale factor of 8.

Data	Scale	Bicubic	SRCNN	VDSR	SRResNet	DCAN
Grass	×8	29.67/0.7214	31.07/0.7361	31.41/0.7513	31.55/0.7496	31.76/0.7552
Industry	×8	21.85/0.4853	22.04/0.4872	22.44/0.5131	22.58/0.5164	22.69/0.5273
River lake	×8	27.42/0.7011	27.91/0.7154	28.23/0.7263	28.35/0.7295	28.51/0.7353
Filed	×8	30.04/0.6771	30.97/0.6841	31.25/0.6874	31.36/0.6893	31.44/0.6936
Forest	×8	25.56/0.4418	25.91/0.4433	26.11/0.4554	26.17/0.4585	26.31/0.4615
Resident	×8	20.78/0.4239	21.13/0.4251	21.44/0.4482	21.54/0.4563	21.70/0.4624
Parking	×8	21.77/0.4531	22.05/0.4681	22.35/0.4873	22.46/0.4942	22.59/0.4998

3.3. Model Analysis

In this section, we use three groups of experiments to analyze the proposed method using the UC Merced dataset, including the number of DCAB, the dense channel attention mechanism, and the SAB. In addition, we use PSNR as quantitative evaluation.

- (1) DCAB: A small DCAN which contains four DCAB_s was used to learn the effect of n_f . The experimental results about the effect of n_f are presented in Table 5. We use three widely used values, including 32, 64 and 128. When $n_f = 64$, the average PSNR values on the test dataset are 0.27 dB higher than $n_f = 32$, and it also are 0.16 dB higher than $n_f = 128$. So, we set the $n_f = 64$ in the rest of the experiments.

Table 5. The PSNR of DCAN with a scale factor of 4 under different n_f on the UC Merced dataset.

n_f	PSNR
32	28.41
64	28.68
128	28.52

- (2) Dense Channel Attention Mechanism: We use several DCAN which contain different DCAB_s to study the effect of the number of DCAB_s. The experimental results about the effect of DCAB_s are provided in Table 6. We set the number of DCAB_s to 4, 8 and 10, respectively. As shown in the table, we conclude that the PSNR of ten DCAB_s are higher than 4 and 8.

Table 6. The PSNR of DCAN which have different DCAB_s with a scale factor of 4 on the UC Merced dataset.

DCAB _s	PSNR
4	28.68
8	28.70
10	28.71

- (3) SAB: To study the effect of the SAB, we compare the DCAN-s which contain a SAB and the DCAN without SAB. The experimental results are provided in Table 7. It can be seen that the PSNR of DCAN-s is higher than DCAN. It demonstrates that SAB improves the network performance.

Table 7. The PSNR of DCAN and DCAN-s with a scale factor of 4 on the UC Merced dataset.

Model	PSNR
DCAN	28.63
DCAN-S	28.70

In addition, we explore the relationship between PSNR and epochs of network training. As shown in Figure 16, it can be seen that the PSNR of around 350 epochs begins to converge with a scale factor of $\times 4$. As shown in Figure 17, at around 50 epochs, it begins to converge with a scale factor of $\times 8$. We can draw the conclusion that, with our model, it is easy to achieve high performance.

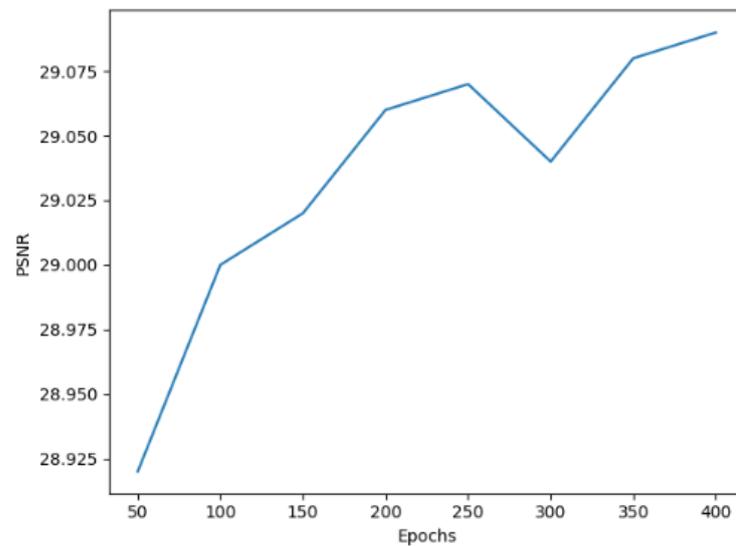


Figure 16. PSNR of DCAN with a scale factor of $\times 4$ on the UC Merced dataset.

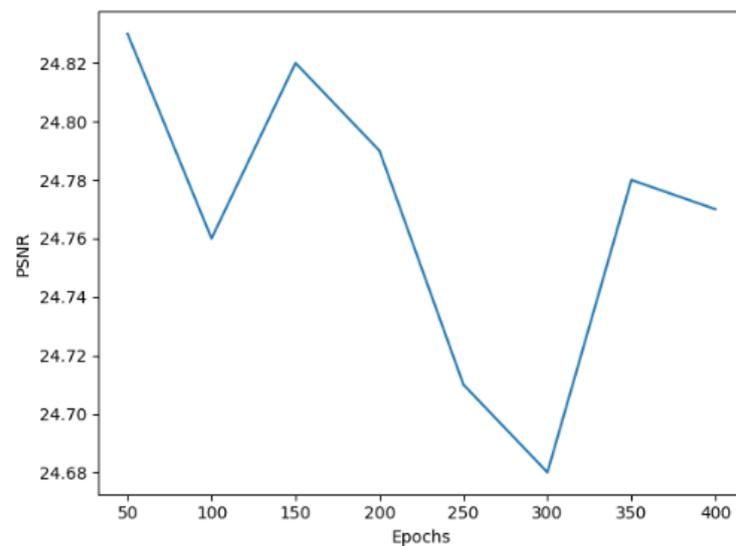


Figure 17. PSNR of DCAN with a scale factor of $\times 8$ on the UC Merced dataset.

3.4. Super-Resolving the Real-World Data

In this section, we use the real data to validate the robustness of our proposed method. The model is trained by UC Merced dataset and tested by some remote sensing images from GaoFen-2 in Ningxia, China, as the real-world data. The size and band number of real-world images are 400×400 and 3, respectively. Figures 18 and 19 show one LR image on the $\times 4$ and $\times 8$ enlargement, respectively. The reconstructed results show that the DCAN method achieves good results in terms of the visual quality.

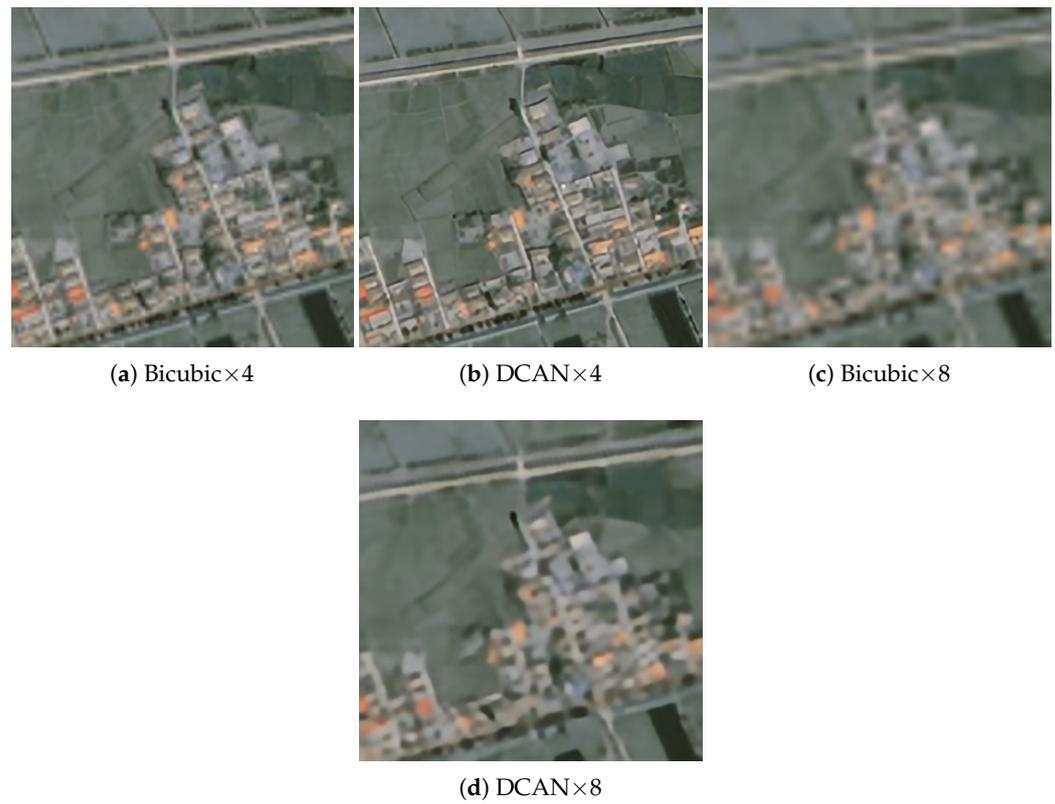


Figure 18. SR results of real data of $\times 4$ and $\times 8$ scale factor. (a–d) represent the result of Bicubic $\times 4$, DCAN $\times 4$, Bicubic $\times 8$, DCAN $\times 8$, respectively.

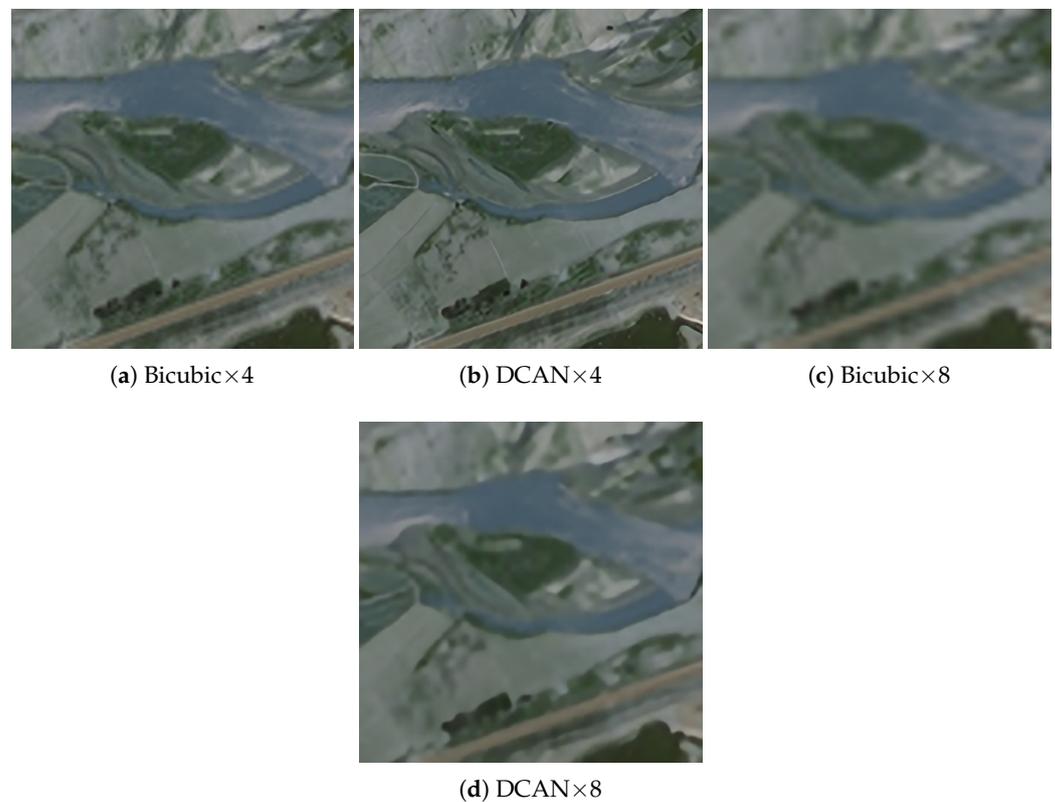


Figure 19. SR results of real data of $\times 4$ and $\times 8$ scale factor. (a–d) represent the result of Bicubic $\times 4$, DCAN $\times 4$, Bicubic $\times 8$, DCAN $\times 8$, respectively.

4. Discussion

The proposed DCAN method is proven to have a good performance with experimental results. In Section 3.2, our method outperforms several state-of-the-art methods in both quantitative evaluation and visual quality. In Section 3.3, the increase of PSNR after adding DCAB and SAB demonstrates the effectiveness of our approach. In Section 3.4, we deal with real satellite images from GaoFen-2 using the DCAN model and obtain satisfactory SR results. In addition, we discussed the experimental results in combination with theoretical analysis.

- (1) Effect of Dense Channel Attention Block: As shown in Table 6, the PSNR of super-resolved images increases after we add the DCAB. It can be seen that when the number of DCABs increases from 4 to 10, it improves the performance of network. It proves that when we increase appropriately the number of DCABs, the capacity of the network can be improved.
- (2) Effect of Spatial Attention Block: As shown in Table 7, it can be seen that the PSNR value ranges from 28.63 to 28.70 after we add the SAB. Thus, we conclude that SAB can improve the performance of network and help the network have more flexible discriminative ability for global structure and focus on high-frequency information from spatial dimension.
- (3) Effect of Scale Factor: As shown in Tables 1 and 2, due to the increasing scale factor, the improvement of our proposed DCAN method reduces. It indicates that the large-scale super resolution is still a hard problem.

5. Conclusions

This article develops a DCAN network which achieves good performance in super-resolving the remote sensing images with complicated spatial distribution. Specifically, we design a network which densely uses the multi-level feature information and strengthens the effective information. In addition, we propose a dense channel attention mechanism which makes better use of multi-level feature maps which contain abundant high-frequency information. Further, we add a spatial attention block to pay more attention to the regions which are more important and more difficult to reconstruct. Results of the extensive experiments demonstrate the superiority of our method over the other compared algorithms.

Author Contributions: Funding acquisition, P.L. and X.S.; Investigation, Y.M.; Methodology, Y.M., P.L. and H.L.; Resources, P.L., H.L., X.S. and Y.Z.; Supervision, P.L.; Validation, Y.M.; Writing—original draft, Y.M. and P.L.; Writing—review & editing, P.L. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 42001307, and Grant 62061038, in part by the Ningxia Key RD Program under Grant 2020BFG02013, and in part by the Natural Science Foundation of Ningxia under Grant 2020AAC02006.

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

References

1. Dong, Z.; Wang, M.; Wang, Y.; Zhu, Y.; Zhang, Z. Object detection in high resolution remote sensing imagery based on convolutional neural networks with suitable object scale features. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2104–2114. [[CrossRef](#)]
2. Amit, S.N.K.B.; Shiraishi, S.; Inoshita, T.; Aoki, Y. Analysis of satellite images for disaster detection. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 5189–5192.
3. Mathieu, R.; Freeman, C.; Aryal, J. Mapping private gardens in urban areas using object-oriented techniques and very high-resolution satellite imagery. *Landsc. Urban Plan.* **2007**, *81*, 179–192. [[CrossRef](#)]
4. Li, W.; He, C.; Fang, J.; Zheng, J.; Fu, H.; Yu, L. Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data. *Remote Sens.* **2019**, *11*, 403. [[CrossRef](#)]
5. Zhao, K.; Kang, J.; Jung, J.; Sohn, G. Building extraction from satellite images using mask R-CNN with building boundary regularization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 247–251.

6. Yuan, S.; Dong, R.; Zheng, J.; Wu, W.; Zhang, L.; Li, W.; Fu, H. Long time-series analysis of urban development based on effective building extraction. In Proceedings of the Geospatial Informatics X, International Society for Optics and Photonics, Online, 21 April 2020 ; Volume 11398, p. 113980M.
7. Lu, T.; Ming, D.; Lin, X.; Hong, Z.; Bai, X.; Fang, J. Detecting building edges from high spatial resolution remote sensing imagery using richer convolution features network. *Remote Sens.* **2018**, *10*, 1496. [[CrossRef](#)]
8. Zhao, P.; Liu, K.; Zou, H.; Zhen, X. Multi-stream convolutional neural network for SAR automatic target recognition. *Remote Sens.* **2018**, *10*, 1473. [[CrossRef](#)]
9. Zhang, W.; Witharana, C.; Liljedahl, A.K.; Kanevskiy, M. Deep convolutional neural networks for automated characterization of arctic ice-wedge polygons in very high spatial resolution aerial imagery. *Remote Sens.* **2018**, *10*, 1487. [[CrossRef](#)]
10. Xu, Y.; Zhu, M.; Li, S.; Feng, H.; Ma, S.; Che, J. End-to-end airport detection in remote sensing images combining cascade region proposal networks and multi-threshold detection networks. *Remote Sens.* **2018**, *10*, 1516. [[CrossRef](#)]
11. Ma, J.; Zhao, J.; Jiang, J.; Zhou, H.; Guo, X. Locality preserving matching. *Int. J. Comput. Vis.* **2019**, *127*, 512–531. [[CrossRef](#)]
12. Lv, P.; Zhong, Y.; Zhao, J.; Zhang, L. Unsupervised change detection based on hybrid conditional random field model for high spatial resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4002–4015. [[CrossRef](#)]
13. Kakareko, G.; Jung, S.; Ozguven, E.E. Estimation of tree failure consequences due to high winds using convolutional neural networks. *Int. J. Remote Sens.* **2020**, *41*, 9039–9063. [[CrossRef](#)]
14. Song, Y.; Zhang, Z.; Baghbaderani, R.K.; Wang, F.; Qu, Y.; Stuttsy, C.; Qi, H. Land cover classification for satellite images through 1d cnn. In Proceedings of the 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 24–26 September 2019; pp. 1–5.
15. Kocatepe, A.; Ulak, M.B.; Kakareko, G.; Ozguven, E.E.; Jung, S.; Arghandeh, R. Measuring the accessibility of critical facilities in the presence of hurricane-related roadway closures and an approach for predicting future roadway disruptions. *Nat. Hazards* **2019**, *95*, 615–635. [[CrossRef](#)]
16. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [[CrossRef](#)]
17. Shen, H.; Zhang, L.; Huang, B.; Li, P. A MAP approach for joint motion estimation, segmentation, and super resolution. *IEEE Trans. Image Process.* **2007**, *16*, 479–490. [[CrossRef](#)]
18. Köhler, T.; Huang, X.; Schebesch, F.; Aichert, A.; Maier, A.; Hornegger, J. Robust multiframe super-resolution employing iteratively re-weighted minimization. *IEEE Trans. Comput. Imaging* **2016**, *2*, 42–58. [[CrossRef](#)]
19. Lei, S.; Shi, Z.; Zou, Z. Super-resolution for remote sensing images via local–global combined network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1243–1247. [[CrossRef](#)]
20. Li, F.; Jia, X.; Fraser, D.; Lambert, A. Super resolution for remote sensing images based on a universal hidden Markov tree model. *IEEE Trans. Geosci. Remote Sens.* **2009**, *48*, 1270–1278.
21. Garzelli, A. A review of image fusion algorithms based on the super-resolution paradigm. *Remote Sens.* **2016**, *8*, 797. [[CrossRef](#)]
22. Pan, Z.; Ma, W.; Guo, J.; Lei, B. Super-resolution of single remote sensing image based on residual dense backprojection networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7918–7933. [[CrossRef](#)]
23. Chang, K.; Ding, P.L.K.; Li, B. Single image super-resolution using collaborative representation and non-local self-similarity. *Signal Process.* **2018**, *149*, 49–61. [[CrossRef](#)]
24. Lu, X.; Yuan, H.; Yuan, Y.; Yan, P.; Li, L.; Li, X. Local learning-based image super-resolution. In Proceedings of the 2011 IEEE 13th International Workshop on Multimedia Signal Processing, Hangzhou, China, 17–19 October 2011; pp. 1–5.
25. Zhong, Y.; Zhang, L. Remote sensing image subpixel mapping based on adaptive differential evolution. *IEEE Trans. Syst. Man Cybern. Part B* **2012**, *42*, 1306–1329. [[CrossRef](#)]
26. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873. [[CrossRef](#)] [[PubMed](#)]
27. Lu, X.; Yan, P.; Yuan, Y.; Li, X.; Yuan, H. Utilizing homotopy for single image superresolution. In Proceedings of the First Asian Conference on Pattern Recognition, Beijing, China, 28 November 2011; pp. 316–320.
28. Lu, X.; Yuan, H.; Yan, P.; Yuan, Y.; Li, X. Geometry constrained sparse coding for single image super-resolution. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1648–1655.
29. Lu, X.; Yuan, Y.; Yan, P. Image super-resolution via double sparsity regularized manifold learning. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 2022–2033. [[CrossRef](#)]
30. Lu, X.; Yuan, Y.; Yan, P. Alternatively constrained dictionary learning for image superresolution. *IEEE Trans. Cybern.* **2013**, *44*, 366–377. [[PubMed](#)]
31. Dong, W.; Zhang, L.; Shi, G.; Wu, X. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Trans. Image Process.* **2011**, *20*, 1838–1857. [[CrossRef](#)]
32. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [[CrossRef](#)]
33. Dong, C.; Loy, C.C.; Tang, X. Accelerating the super-resolution convolutional neural network. In *Lecture Notes in Computer Science, Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016*; Springer: Cham, Switzerland, 2016; pp. 391–407.

34. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1874–1883.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
36. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1646–1654.
37. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 June 2017; pp. 136–144.
38. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 June 2017; pp. 4681–4690.
39. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 286–301.
40. Ahn, N.; Kang, B.; Sohn, K.A. Fast, accurate, and lightweight super-resolution with cascading residual network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 252–268.
41. Haris, M.; Shakhnarovich, G.; Ukita, N. Deep back-projection networks for super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1664–1673.
42. Tong, T.; Li, G.; Liu, X.; Gao, Q. Image super-resolution using dense skip connections. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4799–4807.
43. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2472–2481.
44. Zhang, S.; Yuan, Q.; Li, J.; Sun, J.; Zhang, X. Scene-adaptive remote sensing image super-resolution using a multiscale attention network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4764–4779. [[CrossRef](#)]
45. Dong, X.; Sun, X.; Jia, X.; Xi, Z.; Gao, L.; Zhang, B. Remote sensing image super-resolution using novel dense-sampling networks. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 1618–1633. [[CrossRef](#)]
46. Dong, X.; Wang, L.; Sun, X.; Jia, X.; Gao, L.; Zhang, B. Remote Sensing Image Super-Resolution Using Second-Order Multi-Scale Networks. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 3473–3485. [[CrossRef](#)]
47. Dong, X.; Xi, Z.; Sun, X.; Gao, L. Transferred multi-perception attention networks for remote sensing image super-resolution. *Remote Sens.* **2019**, *11*, 2857. [[CrossRef](#)]
48. Lu, T.; Wang, J.; Zhang, Y.; Wang, Z.; Jiang, J. Satellite image super-resolution via multi-scale residual deep neural network. *Remote Sens.* **2019**, *11*, 1588. [[CrossRef](#)]
49. Jiang, K.; Wang, Z.; Yi, P.; Wang, G.; Lu, T.; Jiang, J. Edge-enhanced GAN for remote sensing image superresolution. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5799–5812. [[CrossRef](#)]
50. Lei, S.; Shi, Z.; Zou, Z. Coupled adversarial training for remote sensing image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 3633–3643. [[CrossRef](#)]
51. Yang, Y.; Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 270–279.
52. Zou, Q.; Ni, L.; Zhang, T.; Wang, Q. Deep learning based feature selection for remote sensing scene classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2321–2325. [[CrossRef](#)]
53. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
54. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1026–1034.
55. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic Differentiation in Pytorch. 2017. Available online: <https://openreview.net/forum?id=BJJsrmfCZ> (accessed on 28 October 2017).