



Article

A Lightweight Dense Connected Approach with Attention on Single Image Super-Resolution

Lei Zha ¹, Yu Yang ¹, Zicheng Lai ², Ziwei Zhang ¹ and Juan Wen ^{1,*}

- College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China; sy20203081527@cau.edu.cn (L.Z.); S20203081453@cau.edu.cn (Y.Y.); wuyoupikakhc@gmail.com (Z.Z.)
- School of Advanced Materials and Nanotechnology, Xidian University, Xian 710061, China; zicheng_lai@163.com
- * Correspondence: wenjuan@cau.edu.cn

Abstract: In recent years, neural networks for single image super-resolution (SISR) have applied more profound and deeper network structures to extract extra image details, which brings difficulties in model training. To deal with deep model training problems, researchers utilize dense skip connections to promote the model's feature representation ability by reusing deep features of different receptive fields. Benefiting from the dense connection block, SRDensenet has achieved excellent performance in SISR. Despite the fact that the dense connected structure can provide rich information, it will also introduce redundant and useless information. To tackle this problem, in this paper, we propose a Lightweight Dense Connected Approach with Attention for Single Image Super-Resolution (LDCASR), which employs the attention mechanism to extract useful information in channel dimension. Particularly, we propose the recursive dense group (RDG), consisting of Dense Attention Blocks (DABs), which can obtain more significant representations by extracting deep features with the aid of both dense connections and the attention module, making our whole network attach importance to learning more advanced feature information. Additionally, we introduce the group convolution in DABs, which can reduce the number of parameters to 0.6 M. Extensive experiments on benchmark datasets demonstrate the superiority of our proposed method over five chosen SISR methods.

Keywords: dense skip connections; single image super-resolution; deep features; channel attention mechanism



Citation: Zha, L.; Yang, Y.; Lai, Z.; Zhang, Z.; Wen, J. A Lightweight Dense Connected Approach with Attention for Single Image Super-Resolution. *Electronics* **2021**, *10*, 1234. https://doi.org/10.3390/electronics10111234

Academic Editor: Gemma Piella

Received: 11 April 2021 Accepted: 19 May 2021 Published: 22 May 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/).

1. Introduction

The single image super-resolution (SISR) image processing technique is essential in computer vision and aims to recover a high-resolution image from a single low-resolution (LR) counterpart. It has been used in a large number of computer vision applications, such as medical imaging [1], surveillance imaging [2,3], object recognition [4], remote sensing imaging [5] and image registration and fusion [6,7]. For example, image registration requires high-resolution (HR) images to provide richer details to transform different sets of data into one coordinate system. However, when the upscaling factor increases, the complexity of image registration increases. As a result, it is vital to design an appropriate architecture for the SISR technique.

Traditional SISR methods can be divided into three categories: reconstruction-based methods [8,9], interpolation methods [10–12], and learning-based methods [13–17]. Recently, with the rapid development of neural networks, Convolutional Neural Network-based (CNN-based) SR methods have achieved remarkable performances [18–24]. In 2015, Dong et al. [18] proposed the first Super-Resolution Convolutional Neural Network (SRCNN), introducing a three-layer convolution neural network for single image super-resolution. Afterwards, a Fast Super-Resolution Convolutional Neural Network (FSRCNN) [19] was proposed to accelerate SRCNN, speeding it up by more than 40 times with better restoration quality via a modified network structure

and smaller filter sizes. To deal with multi-scale SR problems, Lai et al. proposed a progressive upsampling framework named the Laplacian pyramid SR network (LapSRN) [20], which can progressively generate intermediate SR predictions. Since He et al. [25] presented a Residual Network (ResNet) to show that the network's depth can be of great significance for various computer vision tasks, some researchers began to increase the network depth to enhance the SR effect. Kim et al. [21] proposed the very deep convolutional network for SR (VDSR). They extended the depth of the network to 20 and achieved higher performance compared with SRCNN and FSRCNN. Soon after, some works [22,23] also successfully demonstrated that deepening CNN networks could further boost SR performance. To alleviate the vanishing-gradient problem brought by the deep network structure, T. Tong et al. [24] proposed a Densely Connected Convolutional Network for SR (SRDensenet) based on dense skip connections, which achieved significant improvement in the image SR task.

SRDensenet has exhibited a good performance at ICCV 2017 due to its contribution to the effective integration of low-frequency and high-frequency features. However, as the network layer deepens, the number of model parameters grows exponentially, dramatically increasing computational complexity. In particular, SRDensenet could waste unnecessary calculations for low-frequency features. The network lacks the ability to identify and learn across feature channels in terms of staking dense blocks, ignoring the inherent feature relationship. Thus, SRDensenet will produce redundant and conflicting information in rich features, which is useless for reconstruction. To resolve the above-discussed issues, in this paper, we propose a lightweight dense connected approach with attention to single image super-resolution tasks, namely, LDCASR, which uses the attention mechanism to learn more effective channel-wise features and the group convolution structure to decrease model parameters. Experiments show that the proposed LDCASR can achieve a better reconstruction performance over the state-of-the-art SISR methods on public SR benchmarks while significantly reducing the model parameters. (About one-ninth of the parameters of SRDensenet.)

In summary, the main contributions of this paper can be summarized as follows.

- We introduce the attention mechanism to the dense connection structure as well as the reconstruction layer, which helps to suppress the less beneficial information during model training. Extensive experiments verify the effectiveness of this attention-based structure.
- 2. Our model can extract the important features by using a lightweight approach. By introducing the group convolution, we reduce the number of parameters to 0.6 M, which is around 1/9 of the original SRDensenet.

The remainder of this paper is organized as follows. In Section 2, we introduce the related work of super-resolution tasks and the attention mechanism. In Section 3, we describe the proposed network LDCASR architecture, including the details of its compositions: Recursive Dense Group (RDG), Dense Attention Block (DAB), and Channel Attention Unit (CAU). The experimental results and analysis on the comparison with other methods are provided in Section 4. Finally, we draw our conclusions in Section 5.

2. Related Work

2.1. CNN-Based SISR

Dong et al. unprecedentedly introduced a three-layer CNN framework into the SISR and proposed a super-resolution convolutional neural network (SRCNN) [18], which exhibited a remarkable performance compared to the traditional works [8–17] and opened the way for neural network-based SR research. After that, plenty of approaches based on convolution neural networks were proposed. A fast super-resolution convolutional neural network (FSRCNN) [19] introduced the deconvolution operation to the CNN model. It not only accelerated the speed but also enhanced the performance of the SRCNN. To further speed up the SR, Shi et al. [26] presented an effective method called efficient subpixel convolutional neural network (ESPCN), which completed the upscaling progress with the aid of subpixel convolution. Lai et al. [20] proposed a progressive upsampling framework

called the Laplacian pyramid network (LapSRN) to increase image size gradually. By further deepening the network structure, a very deep super-resolution, VDSR [21], achieved a better result using a deeper network, including almost 20 convolution layers. Afterwards, DRCN [22] and DRRN [23] also realized deep networks by employing recursive learning and sharing parameters. However, with the deepening of the network, the issue of the gradient vanishing appeared. Researchers found the skip connection [25] is a handy way to address the gradient vanishing. Regarding this problem, Ledig et al. [27] proposed residual neural network for SR (SRResNet) with more than 100 layers. They adopted the generator part of the SRGAN as the model structure and employed the residual connections between layers. After that, Lim et al. proposed two even deeper and wider networks: an enhanced deep SR network (EDSR) [28] and a multi-scale deep SR network (MDSR) [28], which both consisted of 1000 convolution layers. These deep SISR networks improve performance by simply stacking the different blocks. However, they ignore the channel-wise feature information. In fact, in addition to the dimension of length and width, the channel is another crucial dimension of an image. Channel attention assigns different weights to each channel to help the network pay attention to important features and suppress unimportant features. With channel attention, the model performance can be improved with a small amount of computation.

2.2. Dense Skip Connections in SISR

To deal with deep model training problems, researchers utilized dense skip connections to promote the model's feature representation ability by reusing deep features of different receptive fields. The dense skip connections were first proposed in DenseNet [29], which was the subject of best paper in CVPR 2017. Afterward, SRDenseNet [24] exhibited a good performance in SISR by introducing the dense skip connections. Then, many networks employed the dense connected structure in the SR task and exhibited remarkable performances. However, the dense connected structure simultaneously introduces redundant and useless information, which is harmful to the image's super-resolution. Different from these methods, we combined the dense connected blocks with the attention mechanism to focus on learning important information.

2.3. Attention Mechanism

The attention mechanism was derived from a study of human vision, which was first proposed in the field of visual images. Google Mind team [30] proposed Recurrent Models of Visual Attention in 2014, which used the attention mechanism for image classification on the RNN model. In recent years, attention-based methods have yielded attractive results in various tasks, for instance, image recognition [31] and natural language processing [32]. Researchers found that attention mechanism can not only reduce useless information by discriminating the effective feature information but emphasize the importance in various dimensions. Wang et al. [33] designed a stackable network structure with a trunk-andmask attention mechanism for image classification tasks. Hu et al. presented a novel block called squeeze and excitation (SE) [34], which employs channel-wise associations using average-pooled features to increase the representational power of a CNN network and ensure accurate image classification. Specifically, the SE block was introduced to deep convolutional neural networks to enhance performance more [35]. Recently, Dai et al. [36] proposed a novel module called second-order channel attention (SOCA) to gain more useful feature expression as well as feature correlation learning. All the methods mentioned above undoubtedly obtained significant results. Inspired by these works, we introduce the attention mechanism to reinforce our network and improve the effect.

3. Our Model

In this section, we first describe the entire architecture of the proposed LDCASR. Then, we introduce the details of different components of the proposed network, including the Recursive Dense Group (RDG) and Dense Attention Block (DAB) with a Channel Attention Unit (CAU).

Electronics **2021**, 10, 1234 4 of 14

3.1. Network Architecture

As illustrated in Figure 1, our LDCASR mainly includes three modules: feature extraction module, upscale module, and reconstruction module. We define the original LR input as I_{LR} and the output as I_{SR} .

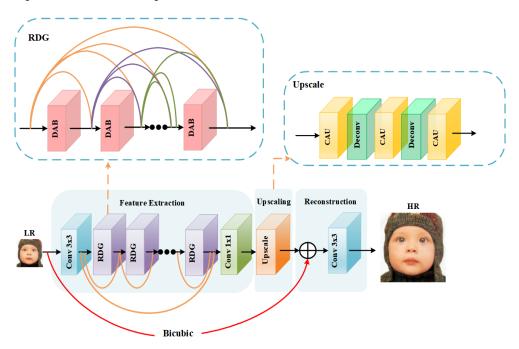


Figure 1. The overall architecture of our model. \oplus denotes the element-wise product.

3.1.1. Feature Extraction Module

The feature extraction module includes a 3×3 convolutional layer for shallow feature extraction and several Recursive Dense Groups (RDGs) for deep feature extraction. First, the original low-resolution images were fed into the 3×3 convolutional layer straightway to extract the shallow feature. This part aims to transfer the inputs from color space to feature space. The process can be expressed as:

$$F_0 = H_{SF}(I_{LR}), \tag{1}$$

where $H_{SF}(\cdot)$ represents the function of the shallow feature extraction process, including only one 3 × 3 convolution layer.

Then, the extracted shallow feature F_0 passed through the stacked Recursive Dense Groups, as well as a 1×1 convolution layer to change the number of channels. This process produces the deep image feature, denoted as:

$$F_{DF} = H_{DF}(H_{RDG}(F_0)), \tag{2}$$

where $H_{RDG}(\cdot)$ represents the feature extraction operation by RDGs, and $H_{DF}(\cdot)$ represents the 1 × 1 convolution operation. Each RDG applies multiple Dense Attention Blocks (DABs) with Channel Attention to suppress redundant information (see Figures 2 and 3), which will be discussed in the following subsection.

Electronics **2021**, *10*, 1234 5 of 14

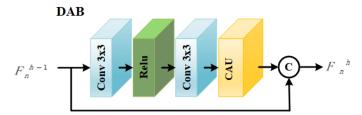


Figure 2. The architecture of the dense attention block (DAB) in RGB. C represents the operation of Channel Concatenate.

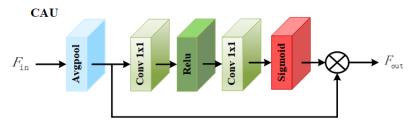


Figure 3. Channe attention unit (CAU). ⊗ denotes the element-wise product.

3.1.2. Upscale Module

The upscale module was placed after the feature extraction module to upscale the feature maps from small size to the size of ground truth. Instead of merely performing deconvolution or subpixel convolution [26] for upscaling, as in the existing methods, we used channel attention units (shown in Figure 3) and deconvolution operations cross-wise to better capture high-frequency information. Specifically, we used one deconvolution operation in the $\times 2$ experiment, two deconvolution operations in the $\times 4$ experiment, and three in the $\times 8$ experiment.

The low-level feature information includes much original image information, a lot of which will be lost during the process of forward propagation. Thus, it is important to combine the low-level features gained by bicubic-interpolation with the high-level information to obtain the final results. Considering fusing extra original image information, the upscaled features were added to the bicubic-interpolated LR image to acquire the final output of the upscale layer F_{up} .

The process is expressed as:

$$F_{up} = H_{up}(F_{DF}) + Bicubic(I_{LR}), \tag{3}$$

where the upscale operation is marked as $H_{up}(\cdot)$.

3.1.3. Reconstruction Module

The final SR image I_{SR} was obtained by the reconstruction module, which contains only one convolution layer of size 3 \times 3. This layer aims to recover the images from the feature space to the color space. The reconstruction progress can be expressed as:

$$I_{SR} = H_R(F_{up}) = H_{LDCASR}(I_{LR}), \tag{4}$$

where $H_R(\cdot)$ denotes the reconstruction layer and $H_{LDCASR}(\cdot)$ denotes the whole process of LDCASR. I_{SR} was optimized by the absolute difference between the I_{LR} and I_{HR} .

3.2. Dense Attention Block (DAB) with Channel Attention Unit (CAU)

As mentioned before, the recursive dense groups (RDGs) are an essential component of our model. Each RDG consists of staked Dense Attention Blocks (DABs) connected by dense connections. It is verified that a large number of dense blocks are beneficial to form a deep CNN in [24]. However, the stacked dense blocks will introduce redundant and conflicting information, causing a longer training time and unsatisfied reconstruction

results. Inspired by the methods based on attention, we employed the channel-dimension attention mechanism to learn the high-frequency features and propose the Dense Attention Block (DAB) (see Figure 2), which contains two 3×3 convolution operations and a Channel Attention Unit (CAU). As a result, with the aids of DABs, our model is able to focus on acquiring more important and useful information. The progress of DAB can be expressed as:

$$F_n^h = f_{cat}(f_{cau}\{f_{conv}(f_{relu}[f_{conv}(F_n^{h-1})])\}, F_n^{h-1}),$$
(5)

where F_n^h , F_n^{h-1} denote output and input of the h-th DAB in the n-th RDG separately, the operation of convolution, Relu, concatenation and CAU as $f_{conv}(\cdot)$, $f_{relu}(\cdot)$, $f_{cat}(\cdot)$, $f_{cau}(\cdot)$, respectively.

We also denote the input of CAU as F_{in} and the output as F_{out} . The specific formula is as follows:

$$F_{in} = f_{conv}(f_{relu}[f_{conv}(F_n^{h-1})]), \tag{6}$$

$$F_{in} = f_{cau}(F_{in}) = f_{avgpool}(F_{in}) \otimes f_{sigmoid}(f_{conv}\{f_{relu}[f_{avgpool}(F_{in})]\}), \tag{7}$$

where $f_{avgpool}(\cdot)$ represents the operation of average pooling, $f_{sigmoid}(\cdot)$ represents the function of sigmoid and \otimes denotes the element-wise product.

3.3. Group Convolution

Group convolution first appeared in AlexNet [37] architecture, which is shown in Figure 4. Unlike the standard convolution operation, group convolution can divide all the R channel inputs into G groups, so each group responds to R/G channels. Then, the output is cascaded to the final output of the entire set of convolutional layers. In order to reduce the parameters of the network, we introduced group convolution in each dense attention block (DAB) to achieve a lightweight network.

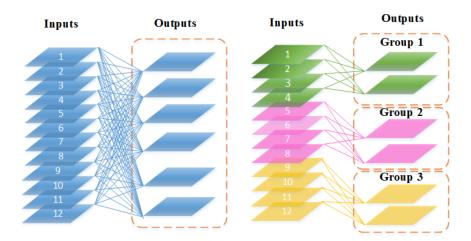


Figure 4. Standard convolution (left) and group convolution (right).

3.4. Loss Functions

In SISR tasks, loss functions are used to measure reconstruction error and lead the model optimization direction.

In the previous stages, many methods [18–20,22,23,25] employ the L2 loss, which is also named the mean square error (MSE) loss. However, researchers found that L2 loss cannot measure the reconstruction quality precisely. Moreover, the result of the reconstruction is not satisfying using L2 loss. Afterwards, more and more researchers tend to use the L1 loss (mean absolute error), which can achieve a better effect of reconstruction. Furthermore, some researchers use the L1 Charbonnier loss function to train the models,

which was first proposed in LapSRN. We mark the original high-resolution image as I_{HR} and the loss functions can be expressed as:

$$L_{I1} = (I_{SR}, I_{HR}) = \frac{1}{hwc} \sum_{n=1}^{N} ||I_{SR} - I_{HR}||,$$
 (8)

$$L_{I2} = (I_{SR}, I_{HR}) = \frac{1}{hwc} \sum_{n=1}^{N} (I_{SR} - I_{HR})^2,$$
 (9)

$$L_{I3} = (I_{SR}, I_{HR}) = \frac{1}{hwc} \sum_{n=1}^{N} \sqrt{(I_{SR} - I_{HR})^2 + \epsilon^2}.$$
 (10)

where h, w and c represent the height, width and number of channels of the feature maps, respectively. $N = h \times w \times c$. ϵ is a constant for numerical stability.

To compare different loss functions, we employed the three loss functions mentioned above to train the LDCASR. The comparison results are displayed in Section 4.

4. Experiments and Analysis

4.1. Training and Testing Datasets

Similar to previous works, our training dataset uses the DIV2K [38]. It consists of 800 training RGB images, 100 validation RGB images, and 100 test images. Following the previous works, we employed 800 training images as our training set, which was then augmented with a 90° rotation and horizontal flip. Furthermore, we used the bicubic kernel function to down-sample the ground truth images to generate LR image counterparts. The training data were generated with Matlab Bicubic Interpolation; training files were created with https://github.com/wxywhu/SRDenseNet-pytorch/tree/master/data, accessed on 11 April 2021.

For testing, we employed four publicly standard benchmark datasets: Set5 [39], Set14 [40], BSD100 [41], and Urban100 [42]. These public datasets provide different kinds of images, which are adequate to validate the models. The test image set was generated with Matlab Bicubic Interpolation; test imageset was created with https://github.com/wxywhu/SRDenseNet-pytorch/tree/master/TestSet, accessed on 11 April 2021.

We followed the previous works by converting the test images from RGB color space to YCbCr color space for evaluation. After that, we employed the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) to evaluate the performance of our structure only on the Y channel.

4.2. Implement Details

The number of RDB and DAB was set to 8, respectively, identical to the dense blocks and single blocks in SRDensenet. Our model was trained by ADAM optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, the batch size was set to 32, the initial learning rate was 0.0001, and we set the learning rate to the initial LR decayed by momentum every 30 epochs. We conducted our experiment with the scaling factors of $\times 2$, $\times 4$, $\times 8$ between the HR and the LR images. We used PyTorch to implement our models with one Titan Xp GPU.

4.3. Different Loss Function Analysis

We show the training results of 60 epochs to compare the training convergence curves of L1 loss, L2 loss, and L1 Charbonnier loss. During the process of gradient descent, the L1 loss was more robust than the L2 loss. Moreover, L1 loss was not easily affected by extreme feature points. This is because the L2 loss function is based on the squared error, which increases twice as much as the L1 loss. As a result, L1 loss is more stable than L2 loss. Additionally, L1 Charbonnier loss adds a variable ϵ^2 compared with L1 loss, making L1 Charbonnier loss more robust than L1 loss. Thus, the learning curve of L1 Charbonnier loss yields the best result. We chose the L1 Charbonnier loss to conduct our following experiments. As shown in Figures 5 and 6, we found that the performance of the training

using L1 Charbonnier loss is better than others as its curve is more stable and the average PSNR or SSIM is higher. Therefore, we chose the L1 Charbonnier loss to train our models. The equations of loss functions are shown in Equations (8)–(10).

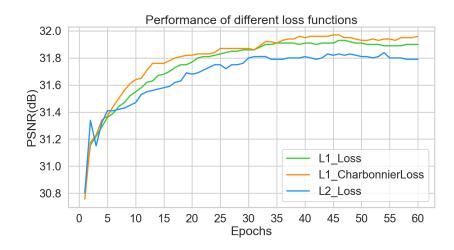


Figure 5. PSNR of networks training with L1, L1 Charbonnier and L2 loss function.

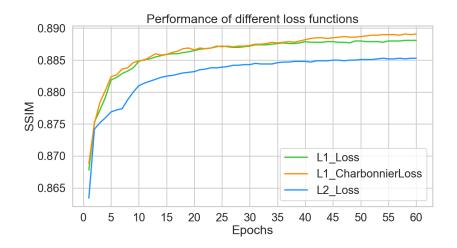


Figure 6. SSIM of networks training with L1, L1 Charbonnier and L2 loss function.

4.4. SISR Performance Comparison

In this section, we investigate the effectiveness of LDCASR. We compare the average PSNR (dB)/SSIM values with the scale factors $\times 2$, $\times 4$, and $\times 8$ on Set5, Set14, BSD100, Urban100.

In Table 1, we compare our model with the state-of-the-art methods, including SRCNN, VDSR, LapSRN, and SRDensenet; all the mentioned methods were trained under the same condition. Bold indicates the best performance. As is shown in Table 1, our model achieved was more effective than the other methods under the scale factors of $\times 2$, $\times 4$, $\times 8$ on the four benchmark datasets. Obviously, under the $\times 2$ scale factor, the value of PSNR on the Urban100 dataset increased by almost 0.47 compared to SRDensenet. In Table 2, we compare the average computational time of different methods. As can be seen, our LDCASR achieved the second fastest performance for $\times 2$, $\times 4$, and $\times 8$ SR experiments, preceded only by SRCNN, which is a simple model with only three convolution layers. Furthermore, we present the performance curve on Set5 for $\times 4$ SR in Figure 7, which intuitively reflects that our model LDCASR is superior to SRDensent. By using the bicubic interpolation, our model exhibited a good performance from the beginning. Subsequently,

Electronics **2021**, 10, 1234 9 of 14

as the training process continued, the PSNR value of LDCASR stayed at around 32, while the PSNR of SRDensenet stayed at approximately 31.54.

Table 1. Benchmark results of several state-of-the-art SISR methods.

Method	Scale	Set5	Set14	BSD100	Urban100
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	×2	33.69/0.9309	30.33/0.8702	29.56/0.8451	26.88/0.8419
SRCNN	$\times 2$	36.65/0.9542	32.45/0.9067	31.40/0.8902	29.54/0.8956
VDSR	$\times 2$	37.53/0.9590	33.05/0.9136	31.95/0.8960	30.77/0.9140
LapSRN	$\times 2$	37.52/0.9591	33.08/0.9130	31.08/0.8950	30.41/0.9103
SRDensenet	$\times 2$	37.83/0.9604	33.26/0.9160	32.08/0.8994	31.58/0.9235
LDCASR(Ours)	$\times 2$	38.12/0.9618	33.56/0.9186	32.21/0.8998	32.07/0.9288
Bicubic	$\times 4$	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6603
SRCNN	$\times 4$	30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221
VDSR	$\times 4$	31.42/0.8824	28.06/0.7689	27.26/0.7265	25.21/0.7757
LapSRN	$\times 4$	31.54/0.8850	28.19/0.7720	27.32/0.7270	25.21/0.7560
SRDensenet	$\times 4$	31.54/0.8834	28.12/0.7712	27.32/0.7296	25.36/0.7640
LDCASR(Ours)	$\times 4$	31.94/0.8898	28.36/0.7758	27.44/0.7328	25.77/0.7781
Bicubic	×8	24.39/0.6582	23.10/0.5660	23.67/0.5480	20.75/0.5160
SRCNN	$\times 8$	25.34/0.6900	23.76/0.5926	24.15/0.5665	20.73/0.5540
VDSR	$\times 8$	25.93/0.7246	24.26/0.6150	24.48/0.5836	21.72/0.5713
LapSRN	$\times 8$	26.15/0.7380	23.38/0.6208	24.55/0.5865	21.83/0.5816
SRDensenet	$\times 8$	25.92/0.7284	24.19/0.6178	24.43/0.5859	21.73/0.5843
LDCASR(Ours)	$\times 8$	26.21/0.7383	24.43/0.6238	24.57/0.5886	21.89/0.5895

Table 2. The average computational time (sec) of different methods on the four benchmark datasets for $\times 2$, $\times 4$, $\times 8$ SR. The fastest results are highlighted and our model achieved the second-fastest results.

Dataset	Scale	SRCNN	VDSR	LapSRN	SRDensenet	LDCASR (Ours)
Set5	×2	0.15036	0.33625	0.27346	1.47526	0.22986
	imes 4	0.16908	0.31396	0.91862	1.09084	0.26533
	$\times 8$	0.10912	0.87221	0.56385	0.90353	0.33260
Set14	$\times 2$	0.23497	0.65412	0.47673	2.75815	0.43869
	$\times 4$	0.25417	0.62675	1.03996	2.16803	0.46524
	$\times 8$	0.21497	1.75756	0.54369	1.13913	0.48765
BSD100	$\times 2$	0.04776	0.43824	0.32861	1.49976	0.26857
	$\times 4$	0.05159	0.44911	0.71945	1.46824	0.28653
	$\times 8$	0.15037	1.16250	0.35662	0.76287	0.28375
Urban100	×2	0.56531	2.54659	2.88541	3.57819	1.98290
	$\times 4$	0.54750	2.05804	3.67770	3.98965	1.65246
	$\times 8$	0.71415	3.70353	1.79873	3.80090	1.40827

Electronics 2021, 10, 1234 10 of 14

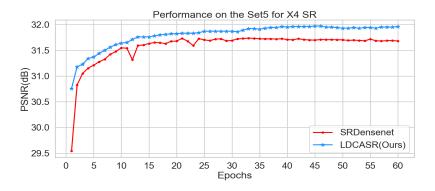


Figure 7. Comparison of the performance between SRDensenet and our model.

In addition, we provide the comparisons of evaluations in terms of visual quality, which are displayed in Figures 8 and 9.

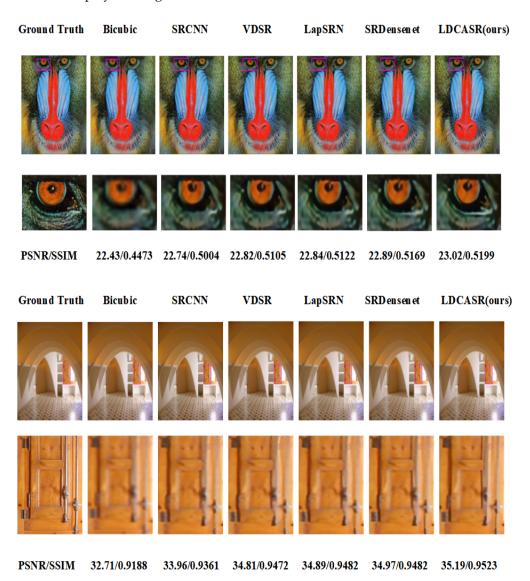


Figure 8. Comparison of our models with other works on $\times 4$ SR. The images are from Set14 and Urban100.

Electronics 2021, 10, 1234 11 of 14

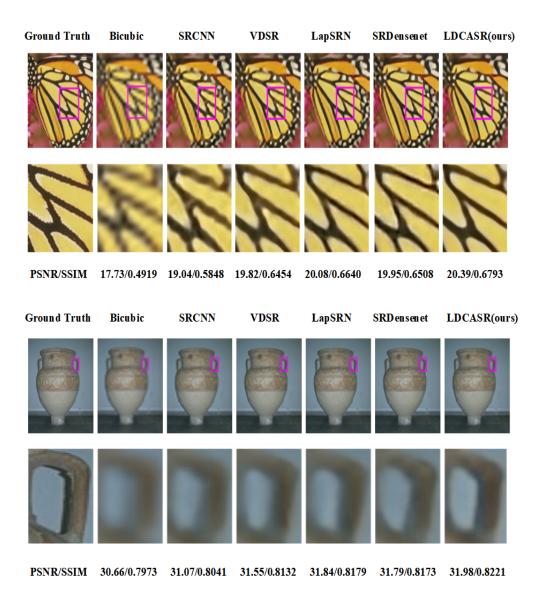


Figure 9. Comparison of our models with other works on $\times 8$ SR. The images are from Set5 and BSD100.

4.5. Model Size Comparison

The comparisons of model size and performance are illustrated in Figure 10. It shows the model size and performance of different state-of-the-art methods. The abscissa refers to the model parameter sizes, while the ordinate denotes the image average PSNR obtained by these models, and each point represents a model. The red point represents our proposed network. As can be seen, SRCNN and VDSR exhibited lightweight but slightly lower performances. LapSRN exhibited a better performance than SRDensenet with fewer parameters. Our LDCASR had fewer parameters and exhibited a relatively better performance, which indicates our model achieves a better trade-off between parameter scale and performance.

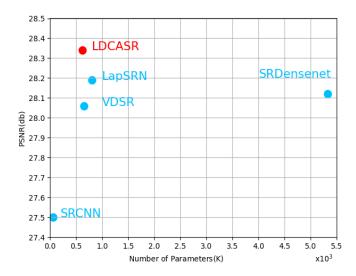


Figure 10. Performance and number of parameters. Results on Set14 (\times 4).

5. Conclusions

In this paper, in order to solve the deficiency of dense network in the SR task, a lightweight dense connected approach with attention is proposed for SISR, where dense attention blocks (DABs) capture important information in the channel dimension by a channel attention unit (CAU). The design of DABs makes the whole network focus on high-frequency details and successfully suppresses useless information in smooth areas. In addition, our model achieved a lightweight effect with fewer parameters and had a relatively superior performance. The extensive experiments on four benchmark datasets illustrate that our LDCASR achieves better results than other state-of-the-art SR methods in terms of objective evaluation and subjective visual effects. In our future work, we will explore more advanced model structures, such as the model based on improving the spatial attention mechanism or multi-attention mechanism. Moreover, we will study the application of our model in other fields.

Author Contributions: Conceptualization, J.W.; methodology, L.Z. and Z.L.; writing—original draft preparation, L.Z.; writing—review and editing, L.Z. and J.W.; validation, L.Z., Y.Y. and Z.Z.; formal analysis, Y.Y. and Z.Z.; supervision, J.W.; funding acquisition, J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 61802410).

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

References

- Huang, Y.; Shao, L.; Frangi, A.F. Simultaneous Super-Resolution and Cross-Modality Synthesis of 3D Medical Images Using Weakly-Supervised Joint Convolutional Sparse Coding. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5787–5796.
- 2. Zhang, L.; Zhang, H.; Shen, H.; Li, P. A super-resolution reconstruction algorithm for surveillance images. *Signal Process.* **2010**. [CrossRef]
- 3. Rasti, P.; Uiboupin, T.; Escalera, S.; Anbarjafari, G. Convolutional Neural Network Super Resolution for Face Recognition in Surveillance Monitoring. In Proceedings of the International Conference on Articulated Motion and Deformable Objects, Palma de Mallorca, Spain, 12–13 July 2016.
- 4. Yang, X.; Wu, W.; Liu, K.; Kim, P.W.; Sangaiah, A.K.; Jeon, G. Long-distance object recognition with image super resolution: A comparative study. *IEEE Access* **2018**, *6*, 13429–13438. [CrossRef]
- 5. Mario, H.J.; Ruben, F.B.; Paoletti, M.E.; Javier, P.; Antonio, P.; Filiberto, P. A New Deep Generative Network for Unsupervised Remote Sensing Single-Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6792–6810.
- 6. Bai, Z.; Li, Y.; Chen, X.; Yi, T.; Wei, W.; Wozniak, M.; Damasevicius, R. Real-Time Video Stitching for Mine Surveillance Using a Hybrid Image Registration Method. *Electronics* **2020**, *9*, 1336. [CrossRef]

7. Huang, S.; Yang, Y.; Jin, X.; Zhang, Y.; Jiang, Q.; Yao, S. Multi-Sensor Image Fusion Using Optimized Support Vector Machine and Multiscale Weighted Principal Component Analysis. *Electronics* **2020**, *9*, 1531. [CrossRef]

- 8. Xin, Y.; Yan, Z.; Dake, Z.; Ruigang, Y. An improved iterative back projection algorithm based on ringing artifacts suppression. *Neurocomputing* **2015**, *162*, 171–179. [CrossRef]
- 9. Stark, H.; Oskoui, P. High-resolution image recovery from image-plane arrays, using convex projections. *J. Opt. Soc. Am. A* **1989**, *6*, 1715–1726. [CrossRef]
- 10. Fattal, R. Image upsampling via imposed edge statistics. ACM Trans. Graph. (TOG) 2007, 26, 95. [CrossRef]
- 11. Keys, R. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* **1981**, 29, 1153–1160. [CrossRef]
- 12. Aràndiga; F. A nonlinear algorithm for monotone piecewise bicubic interpolation. Appl. Math. Comput. 2016, 100–113. [CrossRef]
- 13. Chang, H.; Yeung, D.-Y.; Xiong, Y. Super-resolution through neighbor embedding. Super-resolution Through Neighbor Embedding. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; Volume 1, p. I-275. [CrossRef]
- 14. Timofte, R.; De, V.; Gool, L.V. Anchored Neighborhood Regression for Fast Example-Based Super-Resolution. In Proceedings of the IEEE International Conference on Computer Vision, Columbus, OH, USA, 24–27 June 2014.
- 15. Jiang, J.; Yu, Y.; Wang, Z.; Tang, S.; Hu, R.; Ma, J. Ensemble Super-Resolution With a Reference Dataset. *IEEE Trans. Cybern.* **2019**, 1–15. [CrossRef]
- 16. Jiang, J.; Xiang, M.; Chen, C.; Tao, L.; Ma, J. Single Image Super-Resolution via Locally Regularized Anchored Neighborhood Regression and Nonlocal Means. *IEEE Trans. Multimed.* **2017**, *19*, 15–26. [CrossRef]
- 17. Yang, J.; Wright, J.; Huang, T.; Ma, Y. Image Super-Resolution Via Sparse Representation. *IEEE Trans. Image Process.* **2010**, 19, 2861–2873. [CrossRef] [PubMed]
- 18. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [CrossRef] [PubMed]
- 19. Dong, C.; Loy, C.C.; Tang, X. Accelerating the Super-Resolution Convolutional Neural Network. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Volume 9906, pp. 391–407. [CrossRef]
- 20. Lai, W.-S.; Huang, J.-B.; Ahuja, N.; Yang, M.-H. Deep Laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 624–632.
- 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 (CVPR), Las Vegas, NV, USA, 27–30 June 2015.
- 22. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-Recursive Convolutional Network for Image Super-Resolution. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016.
- 23. Tai, Y.; Yang, J.; Liu, X. Image Super-Resolution via Deep Recursive Residual Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2790–2798.
- 24. 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.
- 25. 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 (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- 26. 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 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
- 27. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 105–114
- Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1132–1140.
- 29. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
- 30. Mnih, V.; Heess, N.; Graves, A.; Kavukcuoglu, K. Recurrent Models of Visual Attention. arXiv 2014, arXiv:1406.6247.
- 31. Liu, C.; Liang, Y.; Xue, Y.; Qian, X.; Fu, J. Food and Ingredient Joint Learning for Fine-Grained Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, 1. [CrossRef]
- 32. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv* 2016, arXiv:1409.0473.
- 33. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual Attention Network for Image Classification. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6450–6458.
- 34. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

Electronics 2021, 10, 1234 14 of 14

35. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. *arXiv* 2018, arXiv:1807.02758.

- 36. Dai, T.; Cai, J.; Zhang, Y.; Xia, S.; Zhang, L. Second-Order Attention Network for Single Image Super-Resolution. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 11057–11066.
- 37. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, 25, 1097–1105. [CrossRef]
- 38. Timofte, R.; Agustsson, E.; Gool, L.V.; Yang, M.H.; Zhang, L.; Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M.; et al. NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1110–1121.
- 39. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Alberi Morel, M.-L. Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding. In Proceedings of the British Machine Vision Conference (BMVC), Guildford, UK, 3–7 September 2012.
- 40. Zeyde, R.; Elad, M.; Protter, M. On Single Image Scale-Up Using Sparse-Representations. *Lecture Notes Comput. Sci.* **2010**, 6920, 711–730. [CrossRef]
- 41. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour Detection and Hierarchical Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, 33, 898–916. [CrossRef] [PubMed]
- 42. Huang, J.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5197–5206.