Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Dual residual attention module network for single image super resolution

Xiumei Wang, Yanan Gu, Xinbo Gao*, Zheng Hui

Video & Image Processing System (VIPs) Lab, School of Electronic Engineering, Xidian University, No.2 South Taibai Road, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 29 January 2019
Revised 10 May 2019
Accepted 23 June 2019
Available online xxx

Communicated by Dr. Lu Xiaoqiang

Keywords:
Super-resolution
Dual residual attention module
Local information integration

ABSTRACT

Recent studies show that research on single image super-resolution (SISR) has achieved great success by using deep convolutional neural network (CNN). Different types of features obtained in deep CNN have different contribution. However, most of the previous models ignore the distinction between different features and deal with them in the same way, which affects the representational capacity of the models. On the other hand, receptive fields with different size capture diverse features from the input. Based on the above considerations, we propose a dual residual attention module (DRAM) network which concentrates on recovering the high-frequency details and sharing the information between two receptive fields of different sizes. We construct local information integration (LFI) module as the basic module to make full use of the local information. The LFI module is a cascade of several dual residual attention fusion (DRAF) blocks with a dense connection structure. The feature modulation can focus on important features and suppress unimportant ones. The evaluation results on five benchmark datasets demonstrate the superiority of our DRAM network against the state-of-the-art algorithms.

1. Introduction

Single image Super-Resolution (SISR), which reconstructs a high-resolution (HR) image from one low-resolution (LR) image, has gained increasing research attention for decades. It is hard to accomplish this task because the high-frequency information is lost in the process of down-sampling. While SISR is an ill-posed inverse procedure, it still is a very popular area on account of its wide applications such as medical imaging, satellite imaging, security and surveillance.

Over the past few decades, various SISR methods have been proposed and achieved amazing results, which can be roughly divided into the following categories: interpolation-based [1], reconstruction-based [2], and learning-based methods [3–15]. Interpolation-based solutions have high speed and low computational complexity, but the performance of those solutions is poor. Reconstruction-based methods try to create a correspondence between HR and LR image, but those methods are limited by the magnification factor.

Recently, due to the development of computational ability, deep learning is widely used to address the SISR problem and achieves miraculous performance. Dong et al. presented a three-layer convolutional neural network (SRCNN) [12,16] to construct an end-to-end mapping between LR and HR image. This method achieved high peak signal-to-noise ratio (PSNR) value which is widely applied to measure the quality of the reconstructed image. SRCNN is the pioneering work introducing deep learning into image super-resolution. Subsequently, various deep learning-based SR methods emerge continuously. Kim et al. built a very deep network (VDSR) [14] by using the global residual connection and gradient clipping, which demonstrated the importance of depth in the network. Besides, the receptive field of VDSR was enhanced compared with SRCNN. Most proposed SR methods can only reconstruct HR image in one up-sampling step, which would present blurry image at large upscaling factors, e.g., $4\times$, $8\times$. To address this issue, Lai et al. proposed a deep Laplacian pyramid network (LapSRN) [17] to gradually restore the sub-band residuals of HR image at different pyramid levels. On the other hand, LapSRN shares parameters across pyramid levels, meanwhile, within pyramid levels. Inspired by the fact that the human thought has persistency, Tai et al. [18] adopted memory block to aggregate features from previous blocks to accomplish persistent memory. Ledig et al. constructed a Generative Adversarial Network (SRGAN) [19] to achieve photo-realistic single image super-resolution. They used residual block which was first proposed in classification problems to construct a deeper network (SRResNet) [19] as the generator network. The image generated by SRResNet and HR image are sent into a discriminator network. The discriminator network will judge which one is more

* Corresponding author.
E-mail address: xbgao@mail.xidian.edu.cn (X. Gao).

https://doi.org/10.1016/j.neucom.2019.06.078
0925-2312/

similar to the ground truth image. Then the generator and discriminator network will be trained alternately. The generator will produce a more deceptive image and the discriminator network will have better distinguishing ability by this form of training. SRGAN can produce images that are more in line with human visual characteristics. Based on SRResNet, enhanced deep residual network (EDSR) [15] won the first prize of the NTIRE 2017 challenge [20]. The method strengthened the network by removing batch normalization (BN) layer as well as utilizing a wider and deeper network. The practice of removing the BN layer can improve the performance of the model but also reduce the number of parameters in the model. Based on this fact, the subsequent model rarely uses the BN layer. To make full use of the features, dense convolution network (Densenet) [21] was built by Huang et al. to aggregate the previous information of the network. Inspired by Densenet, Tong et al. [22] introduced dense connection into SR field. Furthermore, Zhang et al. presented residual dense network (RDN) [13] which takes full advantage of hierarchical features. They utilized deep supervision and larger growth rate to improve the representational ability of the network. To fully explore the relativity between LR and HR image, Haris et al. [23] constructed iteratively up and down projection units. The information feedback from the up and down units adjust the network to produce more reliable results. Hui et al. [24] exploited a compact network by using the information distillation blocks, which achieves a competitive result and saves inference time. They also proposed a two-stage convolutional network [25] for fast and accurate single image super-resolution.

The difficulty of SISR is the recovery of high-frequency details. Although most of the networks achieve great performance, they were not concentrated on the positions of high-frequency information. To tackle this problem, we introduce the attention mechanism into the network structure. Focusing on important features and suppressing unimportant ones, we propose a network named dual residual attention modulation (DRAM) network, which can not only restore important high-frequency details by using the channel and spatial attention mechanism but also adaptively detect the features and achieve features fusion between receptive fields with different size. In the DRAM network, we construct a local feature integration (LFI) module as the building module and stack several LFI modules with share-source connection. Inspired by Park et al. [26], we add the output of each LFI module together at the end. In each LFI module, three dual residual attention fusion (DRAF) blocks are stacked with a dense connected structure. At the end of each LFI module, a 1×1 convolution layer is used to fuse the local features and maintain persistent memory. In our DRAF block, we utilize two paths to obtain the features of LR image. The first path is utilized to detect the attention information and the second one is used to obtain features in different receptive fields. We divide the process of feature extraction into two stages. Specifically, features processed by the channel and spatial attention mechanism are fused with the features extracted at different receptive fields, respectively. Based on the structure, our model achieves competitive results.

In summary, the major contributions of our proposed algorithm are as follows:

- We develop a DRAF block which can not only obtain important high-frequency details by using the channel and spatial attention mechanism, but also can adaptively detect the image features and achieve features fusion at different receptive fields.
- We construct a LFI module via stacking three DRAF blocks with a dense connected structure and using a 1×1 convolution layer to fuse the local features and keep persistent memory.
- We propose a DRAM network for SISR and the model achieves state-of-the-art performance.

2. Related works

SISR has been extensively studied in the literature. Here, we will focus on the algorithms associated with our approach.

2.1. Single image super resolution

Single image Super-Resolution is a popular field in recent years. Benefited from the development of deep learning, plenty of amazing approaches have been proposed. Dong et al. [12,16] first introduced deep learning into SISR task, they up-sampled LR image with bicubic interpolation and then trained a CNN to learn a nonlinear mapping from the input image to high-resolution output. Kim et al. first employed the residual architecture [14] in this field to train a 20 layers network. They also presented a deeply-recursive convolutional network (DRCN) [27], which recursively broaden the receptive field with keeping constant model capacity. Similarly, to maintain persistent memory, Tai et al. constructed a deep recursive residual network (DRRN) [28] to control the number of model parameters. However, the methods mentioned above need to interpolate the LR image to the target size before entering the model. The pre-processing of the LR image increased the computational complexity of the training stage. To solve this problem, deconvolution layer [29] and sub-pixel convolution layer [30] are widely used in the later approaches. Recently, some very deep models have been proposed to achieve more competitive results. For example, Lim et al. exploited a deep and wide network (EDSR) [15] to reconstruct HR image. In addition, they also set up a multi-scale model in which most of the parameters are shared between different scales. Such a model can deal with super-resolution in every scale effectively. To adaptive detect features and achieve features fusion at different scales, Li et al. built a multi-scale residual network (MSRN) [31], which can make fully use the hierarchical features and the local multi-scale features to achieve accurate image super-resolution.

We adopt the sub-pixel convolution layer at the end of our network to upscale the LR features to the desired size for the consideration of reducing computation complexity. To highlight the important features of the input and fuse the features extracted from different receptive fields, we develop the dual residual attention mechanism.

2.2. Attention mechanism

Attention mechanism was first proposed for image processing in the field of computer vision. When the neural networks analysis the image or language, it focuses on partial features each time. In this way, the network is more likely to grasp the point. The goal of the attention mechanism is to help the network mark the important features. The approach of measuring the importance of features is giving weight to the features. Recently, attention mechanism based methods have shown its superiority on a range of tasks, such as machine translation, image classification, and image restoration. Hori et al. [32] proposed a modality-dependent attention mechanism, which selects attends not just to times, but to modalities. Ashish et al. [33] built a novel network based on attention mechanism, which achieves outstanding results in machine translation task. Li et al. [34] proposed a multi-level attention model, which uses frame-level attention layer and region-level attention layer to encode the most correlated feature as the input to the RNN. Yao et al. [35] proposed a temporal attention module to choose relevant segments for video description. Hu et al. [36] exploited a compact model to focus on the inter-channel relationship. They constructed a squeeze-and-excitation (SE) block which utilizes global average-pooled features to compute channel attention. Zhao et al. [37] utilized channel-wise attention model to

Please cite this article as: X. Wang, Y. Gu and X. Gao et al., Dual residual attention module network for single image super resolution, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.06.078

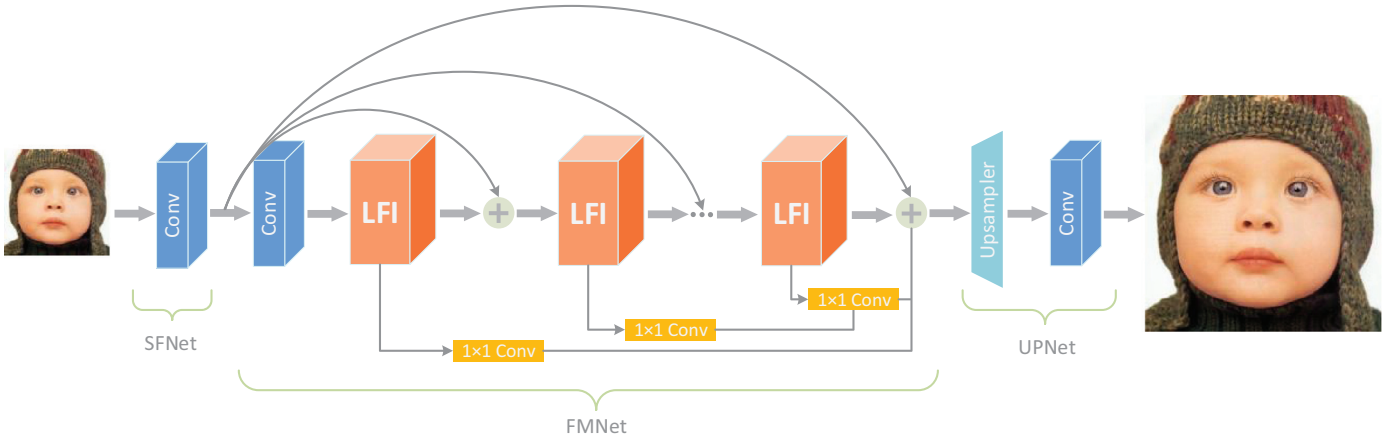


Fig. 1. The network architecture of our dual residual attention modulation (DRAM) network.

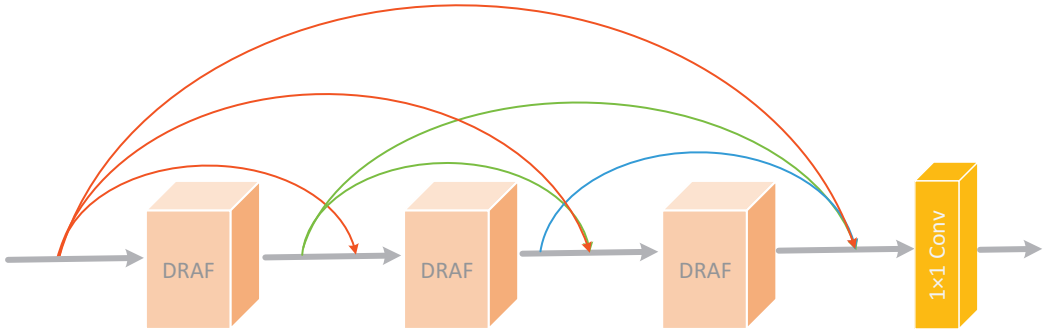


Fig. 2. The structure of local information integration in DRAM network.

discover more discriminative features for the weather recognition task. Woo et al. [38] developed convolutional block attention module (CBAM) to focus on channel and spatial attention. Inspired by CBAM, we adopt two kinds of pooling ways: max-pooling and avg-pooling. These two pooling ways gather different important clues about distinctive object features, which can help the network to infer finer channel and spatial attention.

3. Proposed method

The proposed DRAM network is designed to reconstruct the HR image I_{HR} from the LR image I_{LR}, the architecture of DRAM network is outlined in Fig. 1. In this section, we will describe our network from the whole to the part.

3.1. Network architecture

The network consists of a shallow feature sub-network (SFNet), a feature mapping sub-network (FMNet) and an up-sampling sub-network (UPNet). We use SFNet to extract shallow features from the LR input.

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

where $H_{SFNet}(\cdot)$ denotes the shallow feature sub-network. It is the first convolution layer in the network. F_0 is then used for feature mapping with FMNet. In the FMNet, D local feature integration (LFI) modules are stacked, so we can further have

$$\begin{aligned} F_1 &= H_{FMNet}(F_0) \\ &= H_{LFI,d}(H_{LFI,d\text{-}1}(\ldots(H_{LFI,1}(F_0))\ldots)), \end{aligned} \tag{2}$$

where $H_{FMNet}(\cdot)$ indicates the feature mapping sub-network. It consists of the second convolution layer and several LFI modules. $H_{LFI,d}(\cdot)$ is a composite function of operations of dth LFI module, we will give more details about the LFI module in Section 3.2. We adopt share-source skip connection to make full use of the original information. For the output of every LFI module, we use a 1×1 convolution layer to make information interaction crosses the channel and add the result at the end of the FMNet. Such a practice can facilitate the flow of information throughout the network. Then we put F_1 into the up-sample sub-network.

$$I_{SR} = H_{UPNet}(F_1) = H_{DRAM}(I_{LR}), \tag{3}$$

where $H_{UPNet}(\cdot)$ represents the up-sampling sub-network. It is composed of upscale module and one reconstruction layer. H_{DRAM} denotes the function of the whole network (DRAM). We choose L_1 loss function which is used by most of the previous methods to prove the effectiveness of our network. Given a training dataset $\{I^i_{LR}, I^i_{HR}\}_{i=1}^{i=K}$, where $\{I^i_{LR}, I^i_{HR}\}$ is the ith LR and HR patch pair. K is the number of training patch pairs, the loss function is:

$$L(\Theta) = \frac{1}{K}\sum_{i=1}^{i=K} \|H_{DRAM}(I^i_{LR}) - I^i_{HR}\|_1 \tag{4}$$

where Θ represents the parameter set of the whole network.

3.2. Local feature integration (LFI) module

To take full advantage of local features, we design the LFI module to integrate local information. As depicted in Fig. 2, the LFI module consists of three dual residual attention fusion (DRAF) blocks and one 1×1 convolution layer. More details about DRAF block will be given in Section 3.3. We put three DRAF blocks in dense connection and employ the 1×1 convolution layer to fuse the local features. Experimental results show that the difficulty of network training without LFI module increases with the raising of

Please cite this article as: X. Wang, Y. Gu and X. Gao et al., Dual residual attention module network for single image super resolution, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.06.078

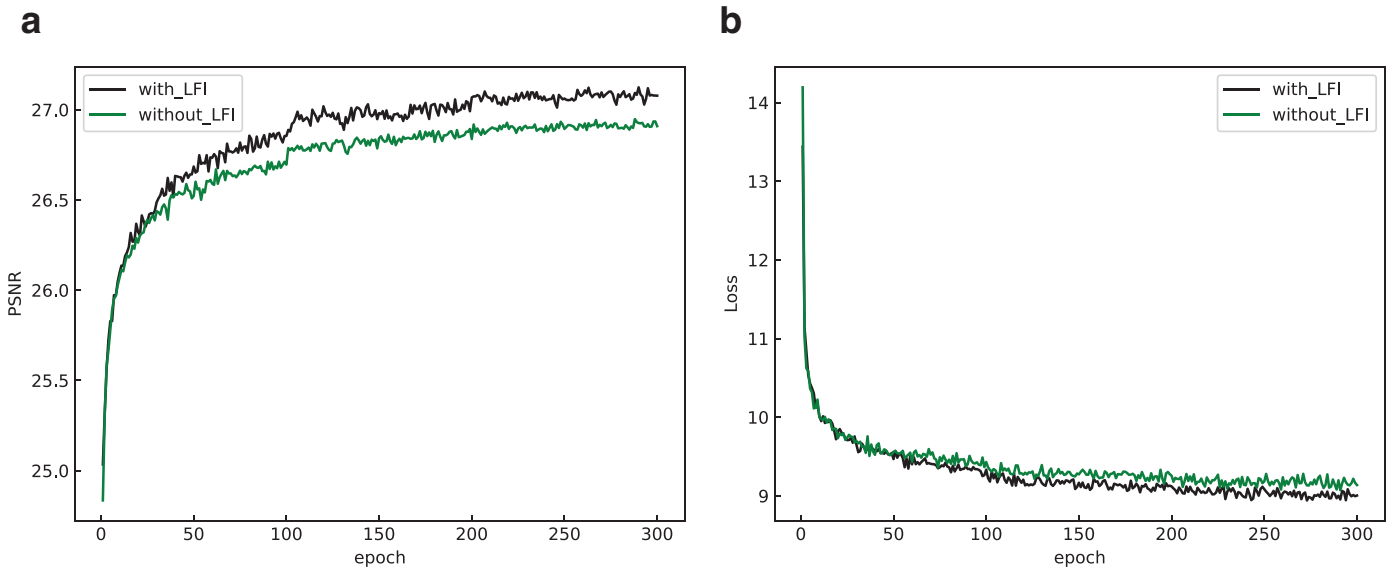


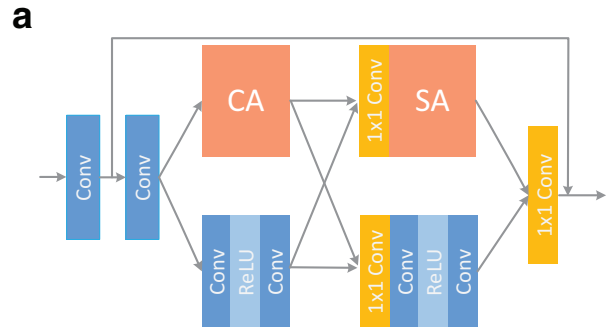
Fig. 3. (a) The converge curve of network about LFI module (b)The loss curve of network about LFI module.

the layer number. As we all know, a very deep network would suffer from a severe gradient vanishing/exploding problem, which is a cause of training difficulties. The building of the LFI module can facilitate the flow of gradients and features across the network to alleviate this problem. We can see the convergence curve and loss curve in Fig. 3. As shown in Fig. 3, the network with LFI module is more easily to converge and to achieve higher PSNR value compared with the network without LFI module. The reduction rate of loss of network with LFI module is faster than the network without LFI module.

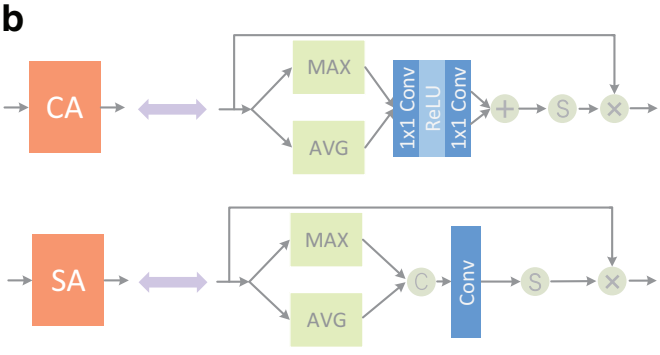
3.3. Dual residual attention fusion (DRAF) block

When a deep network reconstructs a HR image from a LR image, the network should focus on the important high-frequency details. Channel attention (CA) and spatial attention (SA) can recover diverse high-frequency information, so we construct two kinds of attention modules to highlight significant channel and spatial information. On the other hand, the bigger the size of the convolution kernel, the larger the receptive field. A wide receptive field can help network to capture more details of the input. So in another path, we utilize the convolution layer with larger kernel size to capture more global information. In consideration of the computational complexity of the network, two cascading 3×3 convolution layers are used to replace a 5×5 convolution layer. While maintaining the range of receptive fields, it reduces the number of parameters. After finishing the construction of the two paths, we carry out the features fusion between two receptive fields with different size. We divide the information fusion process into two stages. In the first stage, the features extracted by the channel attention mechanism and the features extracted by two 3×3 convolution layers are concatenated so that the information between two paths can be shared with each other. In the second stage, we fuse the features extracted by spatial attention mechanism and the features extracted by another path. By means of double features fusion, we guarantee the block can grasp the significant high-frequency components. Finally, we combine both paths into the residual block to modulate model representations adaptively (Fig. 4).

CA Mechanism:. The goal of CA mechanism is to recalibrate the important relationship between channels of input features. We denote $I = \{I_1, I_2, \ldots, I_C\}$ as a feature set consists of C feature maps



(a) Dual residual attention fusion (DRAF) block



(b) The operations of channel attention (CA) mechanism and spatial attention (SA) mechanism

Fig. 4. The structure of dual residual attention fusion (**DRAF**) block. (a) The **DRAF** block. (b) The operations of channel attention (**CA**) and spatial attention (**SA**), where + , × represent element-add operation, channel-wise multiplication operation, respectively. *C*, *S* denote concat operation and sigmoid operation respectively.

with a size of $H \times W$. Take I as the input of the CA mechanism, we get a one-dimensional vector $\rho = \{\rho_1, \rho_2, \ldots, \rho_C\}$, ρ_c represents the weight of cth channel. We recalibrate the channel importance of input by making the input I do channel-wise multiplication with the one-dimensional vector ρ. Firstly, we adopt a pooling operation across spatial dimensions $H \times W$. Inspired by CBAM [38], both max-pooling and avg-pooling are used in our model, which can fully

Please cite this article as: X. Wang, Y. Gu and X. Gao et al., Dual residual attention module network for single image super resolution, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.06.078

detect the weight relationship of the input feature maps. There are two one-dimensional vectors generated after pooling process: $\rho_{max} \in \mathbb{R}^{C\times1\times1}$, $\rho_{avg} \in \mathbb{R}^{C\times1\times1}$. Then we put ρ_{max} and ρ_{avg} into a shared network (SN), which consists of two 1×1 convolution layers and a ReLU layer. The first convolution layer followed by ReLU layer reduces the number of channels by reduction radio γ, and the following convolution layer increases the number of channels to the original amount. This way can not only enhance the nonlinearity of the network, but also can reduce the model parameters. Then we add the two one-dimensional vectors together and put the result into a sigmoid activation to generate ρ, the process can be computed as:

$$\rho(I) = \delta(SN(avgpool(I)) + SN(maxpool(I))) \tag{5}$$

where $\delta(\cdot)$, $SN(\cdot)$ represent the sigmoid function and the shared network, respectively. Then we multiply ρ and I together to generate more channel discrimination feature maps.

SA mechanism:. SA mechanism is aimed at highlighting the high-frequency details in spatial position. As we all know, the high-frequency information gathers around the edge or texture of the picture. We can restore a more reliable image by detecting the inter-spatial relationship of features. In SA mechanism, we adopt the same pooling way as CA mechanism and obtain two vectors: $\rho_{max} \in \mathbb{R}^{1\times H\times W}$, $\rho_{avg} \in \mathbb{R}^{1\times H\times W}$. Then we concatenate these two vectors and put the result $\rho_{cat} \in \mathbb{R}^{2\times H\times W}$ into a 3×3 convolution layer to produce a vector whose shape is $1\times H\times W$. Finally, the vector goes through the sigmoid function to generate the final weight coefficient.

4. Experiments

In this section, we first present training and testing data setting, then show the setting of model hyper-parameters. Next, we show the result of the ablation experiment to demonstrate the effectiveness of different components of the proposed DRAM network. Finally, we compare our DRAM network with several state-of-the-art algorithms on five benchmark datasets.

4.1. Datasets

Following [13,15,20,39], 800 high-quality training images from DIV2K are chosen as our training set. Data augmentation is performed on the 800 training images. Specifically, we random flip the images horizontally and random rotate them by 90°. For testing, we adopt five widely used datasets: Set5 [42], Set14 [43], B100 [44], Urban100 [45], and Manga109 [46].

4.2. Implement detail

In our model, RGB color channels are used for both input and output image. The average RGB values of the DIV2K are subtracted from all the image. HR patches with a size of 192 × 192 are cropped from the HR image as the input of our model. The number D of LFI module and the mini-batch size are both set to 16. The value of reduction radio γ is set as 4. All the convolution layers of our model except the share convolution layers in CA mechanism have 64 filters. We train DRAM with Adam by setting $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$, the learning rate is initialized to 10^{-4} and halved at every 2×10^5 mini-batch updates. All experiments are conducted using Pytorch, MATLAB R2017a on NVIDIA GTX 1080Ti GPUs.

4.3. Model analysis

In this subsection, we will study the contributions of different components in our model via ablation experiments. We define the model which removes CA, SA, the information exchange structure (IES) and LFI (removing dense connection) as our baseline model. All the models utilized for comparisons are trained on DIV2K dataset with 300 epochs. Table 2 shows the ablation investigation results on the effect of CA, SA, IES and LFI. From Table 2, we can see the PSNR value of model with CA mechanism is higher 0.02 dB than the baseline model. The CA mechanism can adaptively weight large values to the feature maps which play important roles to the reconstruction of the HR image. Compared with baseline model, SA mechanism improves the PSNR value 0.03 dB. Different from CA mechanism which concentrates on the relationship between feature maps, SA mechanism pays attention to the contrast in one feature map. The SA mechanism aims to highlight the texture information to enhance the contrast of the feature map, which benefits the reconstruction of high frequency details. The IES improves 0.03 dB compared with baseline model. Inspired by Li et al. [31], we utilize two paths to obtain the features of LR image in our DRAF block. The first path is a cascade of CA and SA mechanism, and the second path is a series of convolution layers. The first path is utilized to detect the attention information, and the second one is used to obtain features in different receptive fields. Attention unit can grasp the high frequency information and wide receptive fields can offer more global information. Specifically, features processed by the channel and spatial attention mechanism are fused with the features extracted at different receptive fields, respectively. By exchanging the information of the two paths, the information in two paths can be compensated from each other. Inspired by Zhang et al. [13,18], we put three DRAF blocks in dense connection to construct a LFI module, which facilitates the flow of gradients and features across the network to alleviate gradient vanishing/exploding problem. We can demonstrate the effectiveness of the presented LFI module through the results of Table 2 (improving 0.05 dB vs. baseline model). Overall, the performance of the complete network is much better than the network with only one component and the baseline model achieves pool performance. As shown in Fig. 7, we then add one of CA, SA, IES, or LFI to the baseline model, which proves the effectiveness of each component of our model. Each component indeed improves the performance of our network. Although the effectiveness of one component may be not particularly noticeable, the combination of them presents a significant performance.

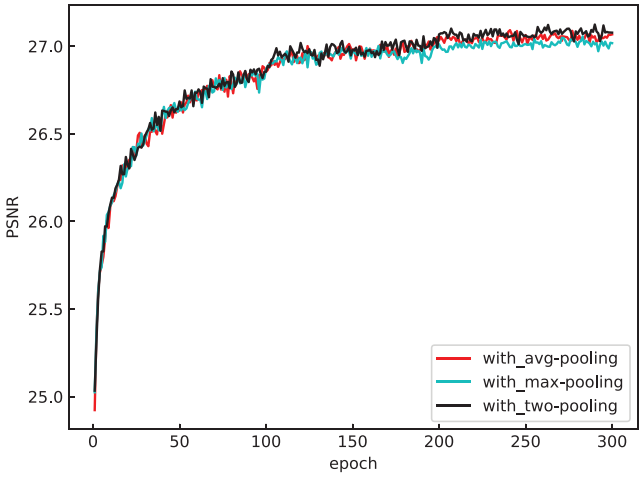


Fig. 5. Convergence analysis on different pooling way.

We also explore the effect of the two pooling ways in attention mechanism. Fig. 5 shows the comparative experiment curves. We can observe that the model with both pooling ways achieves the best performance compared with the model with the only one pooling way. The reason is that the avg-pooling way can grasp the

Please cite this article as: X. Wang, Y. Gu and X. Gao et al., Dual residual attention module network for single image super resolution, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.06.078

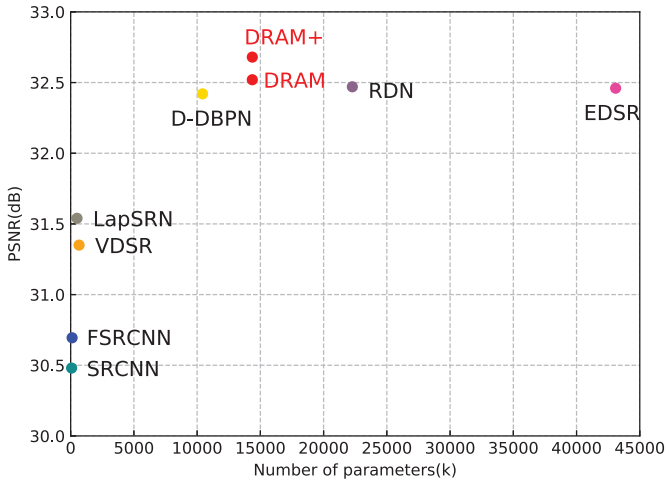


Fig. 6. PSNR performance versus the number of parameters. The results are evaluated on Set5 dataset for a scale factor of 4 ×. Our **DRAM** and **DRAM+** network has a better trade off between performance and model size.

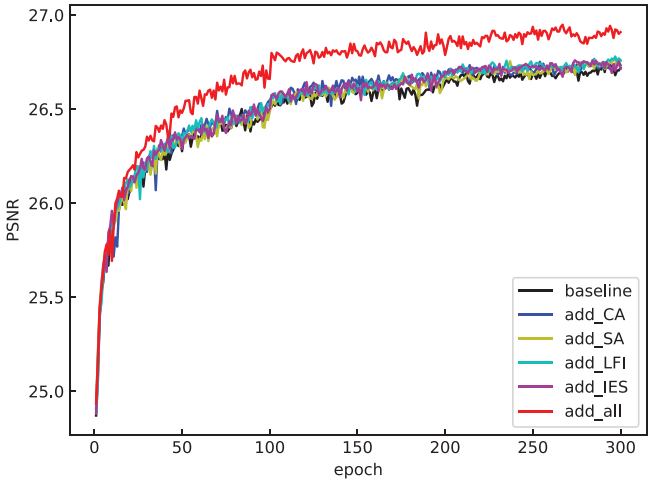


Fig. 7. Convergence analysis on **CA, SA, IES** and **LFI**.

average information of the feature maps and the max-pooling can discover the special point of the feature maps. The combination of the two pooling ways can represent attention information better compared with one pooling way.

To prove our algorithm has a better tradeoff between the performance and model size, we make a comparison with other methods in Fig. 6. The results are evaluated on Set5 dataset for 4 × enlargement. We can see that our DRAM model outperforms the state-of-the-art models. For EDSR and RDN, our approach achieves higher PSNR value but with 67% and 36% fewer parameters respectively. In addition, we also present the mean inference time of our model in Table 3. From the Table 3, it can be seen that as the scale factor increases, the mean inference time decreases. For example, the time complexity of scale 2 × is five times that of scale 8 ×. The reason is that the size of the input image of scale 2 × is larger than that of the scale 8 ×. Therefore, the inference process of scale 2 × is going to be more computationally intensive than the scale 8 ×. In addition, since the difference of image size, mean inference time is variant for different datasets.

To explore the function of information exchange structure (IES), we visualize the feature maps in the DRAF block. As illustrated in Fig. 8(a), most of the feature maps have high pixel values but fuzzy contour. The pixel values of the outline are close to the pixel values of the smooth area, which makes high-frequency information not prominent. As shown in Fig. 8(b), the pixel values of the feature maps are reduced and the outline information is highlighted. This is caused by the combined effect of the two paths.

4.4. Compare with the state-of-the-arts

To demonstrate the effectiveness of our DRAM model, we compare the results produced by the proposed model with several state-of-the-art SISR algorithms, including SRCNN [12,16], FSRCNN [29], VDSR [14], LapSRN [17], MsLapSRN [41], DRRN [28], SCN [40], Memnet [18], EDSR [15], SRMDNF [39], D-DBPN [23], RDN [13]. Similar to [13,15], we introduce self-ensemble strategy to improve our DRAM network and set the self-ensembled one as DRAM + . The details of self-ensemble strategy are as follows. Firstly, we rotate and flip the LR image I^{LR} to generate augmented LR images

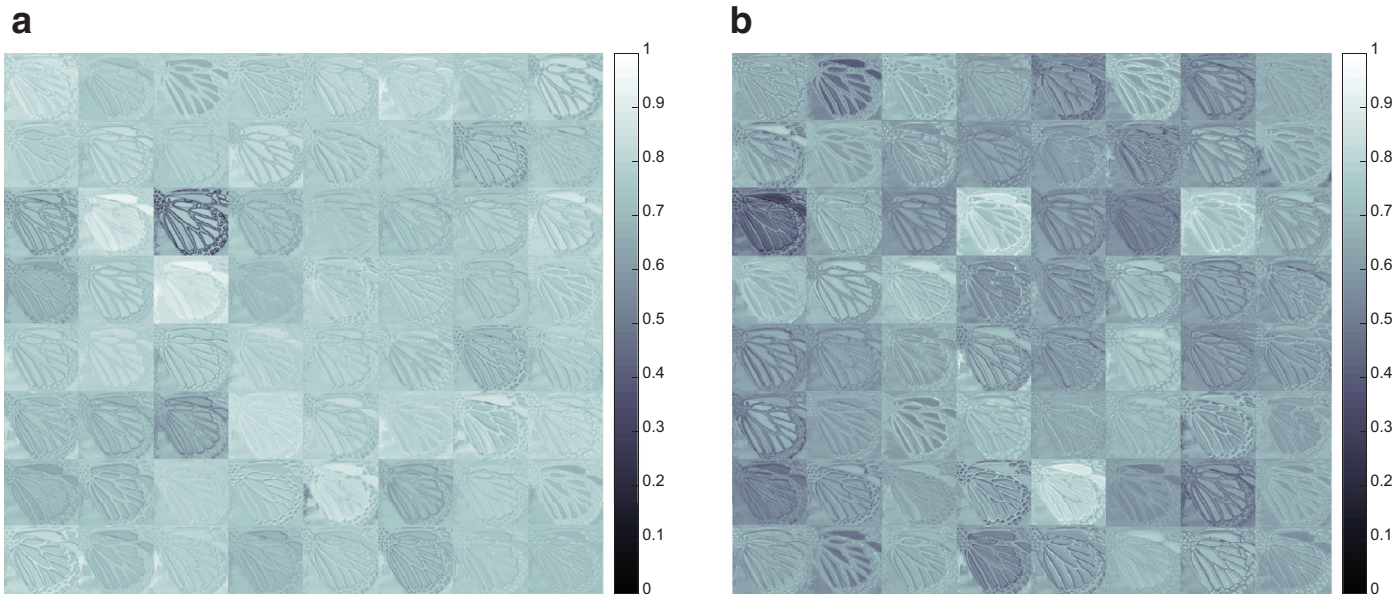


Fig. 8. Visualization of the average feature maps in the DRAF block. The **residual** image of the "butterfly" image is from Set5 and the enlargement scale is 4 ×. (a) The feature maps before two paths. (b) The feature maps after information exchange structure.

Please cite this article as: X. Wang, Y. Gu and X. Gao et al., Dual residual attention module network for single image super resolution, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.06.078

Fig. 9. Visual comparison for 2× SR on Urban100, Set14 and Manga109 dataset. The best results are **highlighted.**

Please cite this article as: X. Wang, Y. Gu and X. Gao et al., Dual residual attention module network for single image super resolution, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.06.078

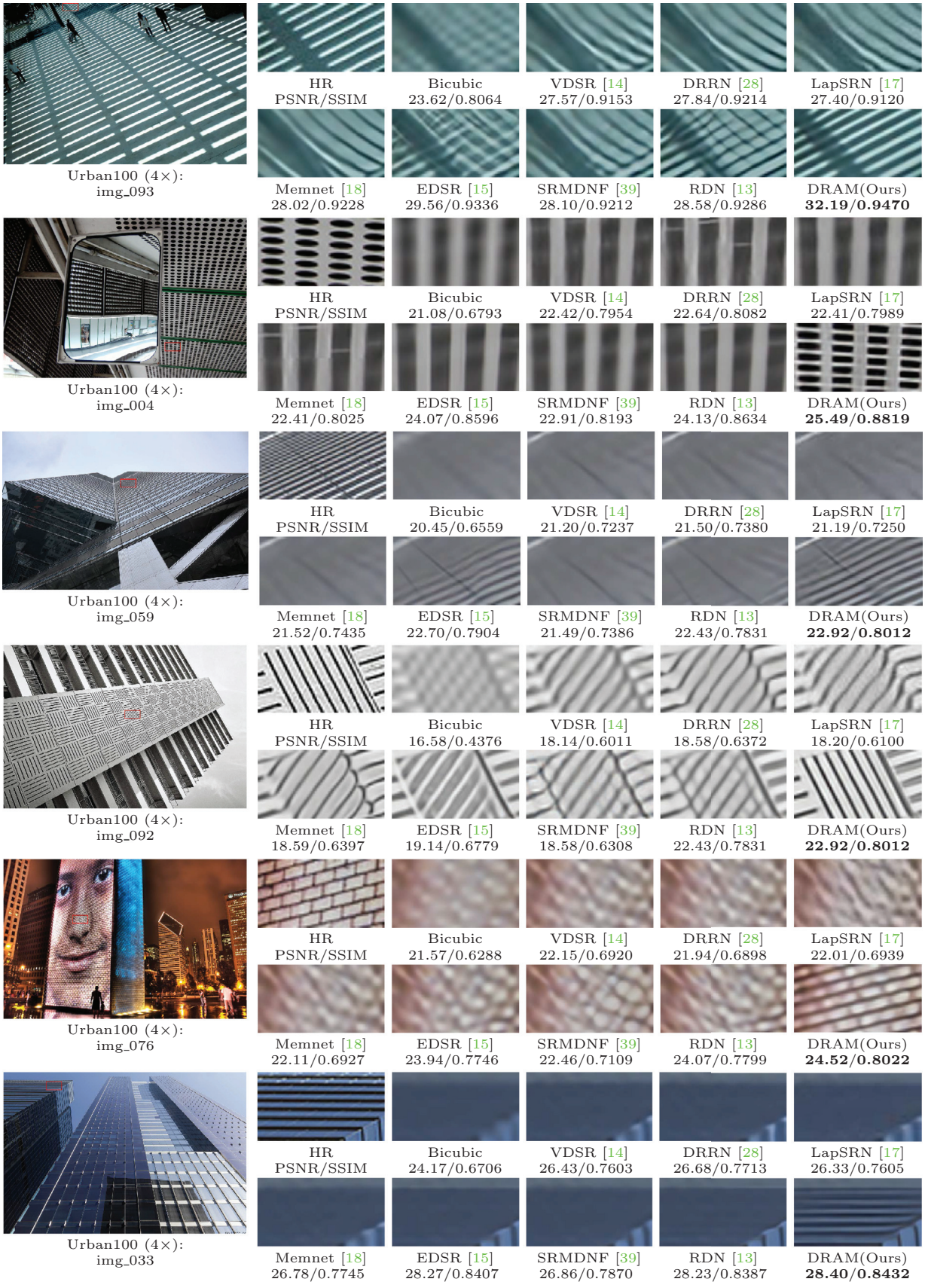


Fig. 10. Visual comparison for 4× SR on Urban100 dataset. The best results are highlighted.

Please cite this article as: X. Wang, Y. Gu and X. Gao et al., Dual residual attention module network for single image super resolution, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.06.078

Table 1
Quantitative evaluations of state-of-the-art SR methods. Bold text indicates the best performance and bold italic indicates the second best.

Method	Scale	Set5		Set14		B100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	×2	33.66	0.9299	30.24	0.8688	29.56	0.8431	26.88	0.8403	30.80	0.9339
SRCNN [12]	×2	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946	35.60	0.9663
FSRCNN [29]	×2	37.05	0.9560	32.66	0.9090	31.53	0.8920	29.88	0.9020	36.67	0.9710
VDSR [14]	×2	37.53	0.9590	33.05	0.9130	31.90	0.8960	30.77	0.9140	37.22	0.9750
LapSRN [17]	×2	37.52	0.9591	33.08	0.9130	31.08	0.8950	30.41	0.9101	37.27	0.9740
DRRN [28]	×2	37.74	0.9591	33.23	0.9136	32.05	0.8973	31.23	0.9188	37.60	0.9736
MemNet [18]	×2	37.78	0.9597	33.28	0.9142	32.08	0.8978	31.31	0.9195	37.72	0.9740
EDSR [15]	×2	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.93	0.9351	39.10	0.9773
SRMDNF [39]	×2	37.79	0.9601	33.32	0.9159	32.05	0.8985	31.33	0.9204	38.07	0.9761
MSRN [31]	×2	38.08	0.9605	33.74	0.9170	32.23	0.9013	32.22	0.9326	38.82	0.9868
D-DBPN [23]	×2	38.09	0.9600	33.85	0.9190	32.27	0.9000	32.55	0.9324	38.89	0.9775
RDN [13]	×2	38.24	0.9614	34.01	0.9212	32.34	0.9017	32.89	0.9353	39.18	***0.9780***
DRAM (Ours)	×2	***38.25***	***0.9615***	***34.18***	***0.9228***	***32.40***	***0.9023***	***33.19***	***0.9373***	***39.24***	0.9777
DRAM+ (Ours)	×2	**38.30**	**0.9617**	**34.32**	**0.9240**	**32.45**	**0.9029**	**33.42**	**0.9191**	**39.41**	**0.9781**
Bicubic	×3	30.39	0.8682	27.55	0.7742	27.21	0.7385	24.46	0.7349	26.95	0.8556
SRCNN [12]	×3	32.75	0.9090	29.30	0.8215	28.41	0.7863	26.24	0.7989	30.48	0.9117
FSRCNN [29]	×3	33.18	0.9140	29.37	0.8240	28.53	0.7910	26.43	0.8080	31.10	0.9210
VDSR [14]	×3	33.67	0.9210	29.78	0.8320	28.83	0.7990	27.14	0.8290	32.01	0.9340
LapSRN [17]	×3	33.82	0.9227	29.87	0.8320	28.82	0.7980	27.07	0.8280	32.21	0.9350
DRRN [28]	×3	34.03	0.9244	29.26	0.8349	28.95	0.8004	27.53	0.8378	32.42	0.9359
MemNet [18]	×3	34.09	0.9248	30.00	0.8350	28.96	0.8001	27.56	0.8376	32.51	0.9369
EDSR [15]	×3	34.65	0.9280	30.52	0.8462	29.25	0.8093	28.80	0.8653	34.17	0.9476
SRMDNF [39]	×3	34.12	0.9254	30.04	0.8382	28.97	0.8025	27.57	0.8398	33.00	0.9403
MSRN [31]	×3	34.38	0.9262	30.34	0.8395	29.08	0.8041	28.08	0.8554	33.44	0.9427
RDN [13]	×3	34.71	0.9296	30.57	0.8468	29.26	0.8093	28.80	0.8653	34.13	0.9484
DRAM (Ours)	×3	***34.80***	***0.9303***	***30.63***	***0.8480***	***29.31***	***0.8106***	***28.98***	***0.8684***	***34.22***	***0.9490***
DRAM+ (Ours)	×3	**34.83**	**0.9307**	**30.73**	**0.8491**	**29.37**	**0.8116**	**29.20**	**0.8715**	**34.53**	**0.9504**
Bicubic	×4	28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577	24.89	0.7866
SRCNN [12]	×4	30.48	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221	27.58	0.8555
FSRCNN [29]	×4	30.72	0.8660	27.61	0.7550	26.98	0.7150	24.62	0.7280	27.90	0.8610
VDSR [14]	×4	31.35	0.8830	28.02	0.7680	27.29	0.7252	25.18	0.7540	28.83	0.8870
LapSRN [17]	×4	31.54	0.8850	28.19	0.7720	27.32	0.7270	25.21	0.7560	29.09	0.8900
DRRN [28]	×4	31.68	0.8888	28.21	0.7721	27.38	0.7284	25.44	0.7638	29.18	0.8914
MemNet [18]	×4	31.74	0.8893	28.26	0.7723	27.40	0.7281	25.50	0.7630	29.42	0.8942
EDSR [15]	×4	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	31.02	0.9148
SRMDNF [39]	×4	31.96	0.8925	28.35	0.7787	27.49	0.7337	25.68	0.7731	30.09	0.9024
MSRN [31]	×4	32.07	0.8903	28.60	0.7751	27.52	0.7273	26.04	0.7896	30.17	0.9034
D-DBPN [23]	×4	32.47	0.8980	28.82	0.7860	27.72	0.7400	26.38	0.7946	30.91	0.9137
RDN [13]	×4	32.47	0.8990	28.81	0.7871	27.72	0.7419	26.61	0.8028	31.00	0.9151
DRAM (Ours)	×4	***32.52***	***0.8992***	***28.83***	***0.7877***	***27.74***	***0.7422***	***26.69***	***0.8044***	***31.06***	***0.9163***
DRAM+ (Ours)	×4	**32.68**	**0.9008**	**28.95**	**0.7899**	**27.81**	**0.7436**	**26.94**	**0.8095**	**31.41**	**0.9192**
Bicubic	×8	24.40	0.6580	23.10	0.5660	23.67	0.5480	20.74	0.5160	21.47	0.6500
SRCNN [12]	×8	25.33	0.6900	23.76	0.5910	24.13	0.5660	21.29	0.5440	22.46	0.6950
FSRCNN [29]	×8	20.13	0.5520	19.75	0.4820	24.21	0.5680	21.32	0.5380	22.39	0.6730
SCN [40]	×8	25.59	0.7071	24.02	0.6028	24.30	0.5698	21.52	0.5571	22.68	0.6963
VDSR [14]	×8	25.93	0.7240	24.26	0.6140	24.49	0.5830	21.70	0.5710	23.16	0.7250
LapSRN [17]	×8	26.15	0.7380	24.35	0.6200	24.54	0.5860	21.81	0.5810	23.39	0.7350
MemNet [18]	×8	26.16	0.7414	24.38	0.6199	24.58	0.5842	21.89	0.5825	23.56	0.7387
MSLapSRN [41]	×8	26.34	0.7558	24.57	0.6273	24.65	0.5895	22.06	0.5963	23.90	0.7564
EDSR [15]	×8	26.96	0.7762	24.91	0.6420	24.81	0.5985	22.51	0.6221	24.69	0.7841
MSRN [31]	×8	26.59	0.7254	24.88	0.5961	24.70	0.5410	22.37	0.5977	24.28	0.7517
D-DBPN [23]	×8	27.21	0.7840	25.13	0.6480	24.88	0.6010	22.73	0.6312	***25.14***	***0.7987***
DRAM (Ours)	×8	***27.27***	***0.7857***	***25.16***	***0.6482***	***24.94***	***0.6028***	***22.86***	***0.6372***	25.06	0.7954
DRAM+ (Ours)	×8	**27.39**	**0.7891**	**25.28**	**0.6513**	**24.99**	**0.6043**	**23.01**	**0.6421**	**25.29**	**0.8005**

$I_{n,i}^{LR} = T_i(I_n^{LR})$, where $T_i(i = 1 \ldots 8)$ denotes one of the 8 transformations including the original input. Then we put all the 8 LR images into the DRAM network to produce 8 SR images $\{I_{n,1}^{SR} \ldots I_{n,8}^{SR}\}$. The SR image $I_{n,i}^{SR}$ will be applied inverse transform to get the original geometry $\bar{I}_{n,i}^{SR} = T_i^{-1}(I_{n,i}^{SR})$. Finally, we average the results all together to get the self-ensemble SR image $I_n^{SR} = \frac{1}{8}\sum_{i=1}^{8} \bar{I}_{n,i}^{SR}$.

PSNR and structural similarity (SSIM) are utilized to measure the quality of the SR image. Higher PSNR and SSIM values usually indicate better quality. We use only the luminance channel (Y) to measure the two values as existing methods. In addition, for a fair comparison, we crop S pixels near the image boundary before measurement by the factor S. We make the quantitative evaluation in the five datasets for four scales ($2\times$, $3\times$, $4\times$, $8\times$). Since some algorithms such as SRCNN, FSRCNN, VDSR, Memnet and EDSR do not perform on $8\times$ scale, we retrained the network on $8\times$ scale by the code provided by authors with recommended parameters.

All the quantitative results are shown in Table 1. When compared with existing solutions, our DRAM+ performs the best on all the datasets with all the scales. Even though without self-ensembled, our DRAM also outperforms the others except in Manga109 dataset. For the $2\times$ enlargement, a scale which most of the state-of-the-art models have achieved pretty good results, the PSNR values of our DRAM still over RDN and EDSR are 0.3 dB and 0.26 dB on Urban100 dataset respectively. For the $8\times$ enlargement, which is challenging scale because of the significant loss of details belongs to the ground truth image, our method outperforms EDSR by a large margin.

Please cite this article as: X. Wang, Y. Gu and X. Gao et al., Dual residual attention module network for single image super resolution, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.06.078

Table 2
Ablation investigation of channel attention (**CA**), spatial attention (**SA**), information exchange structure (**IES**) and local information integration (**LFI**). We observe the best performance PSNR on **Set5** with scaling factor 8 × in 300 epochs.

	Different combinations of CA, SA, IES and LFI							
CA	×	✓	×	×	×	✓	✓	✓
SA	×	×	✓	×	×	✓	✓	✓
IES	×	×	×	✓	×	×	✓	✓
LFI	×	×	×	×	✓	×	×	✓
PSNR	26.73	26.75	26.76	26.76	26.78	26.78	26.82	26.95

Table 3
The mean inference time of proposed model.

Time(s)	Dataset				
	Set5	Set14	B100	Urban100	Manga109
Scale					
2x	0.25	0.51	0.39	1.93	2.43
3x	0.12	0.23	0.16	0.89	1.13
4x	0.08	0.14	0.09	0.51	0.65
8x	0.05	0.05	0.05	0.13	0.17

Visual comparisons on 2 × enlargement between our approach and others are shown in Fig. 9. In image “img_060”, we can observe that most of the compared models cannot produce the true texture and suffer from blurring artifacts, on the contrary, the result of our DRAM is almost the same as the ground truth image. For image “barbara”, our DRAM generates more faithful lines of the clothes compared with others. For image “ppt3”, all the other methods fail to restore the English alphabet “*n*” of the word while our DRAM produces a clear result. To further illustrate the effectiveness of our DRAM model on different enlargement scale, we compare the results of our DRAM with other solutions on factor 4 × . The result is shown in Fig. 10. In image “img_093”, we can see that other models cannot rebuild the stripe and suffer from heavy blurring artifacts. In contrast, our DRAM not only recovers the shape of the stripe but also shows the right direction. For image “img_004”, most of the compared images suffer from so serious blurring artifacts that we cannot recognize what it is. While, the recovered image by our DRAM is almost identical to the original image. For image “img_092”, most of the other methods generate some lines with wrong directions while only our DRAM produces more credible results. The above visual comparisons prove that our model has more powerful representational ability so that it can extract more complicated features from the LR image.

5. Conclusion

In this paper, we propose a dual residual attention modulation (DRAM) network for highly accurate image SR, which utilizes stacked local information integration (LFI) module with the share-source connection structure to improve the flow of information and gradient. The LFI module consists of three dual residual attention fusion (DRAF) blocks and an information integration convolution. The DRAF block contains two paths which share information with each other. One path is a cascade of channel attention (CA) and spatial attention (SA) mechanism and another is a series of convolution layers with a wide receptive field. The structure of the DRAF block helps the network to capture more high-frequency details and enhance the expression ability of the network. Extensive experiments on benchmark datasets demonstrate the effectiveness of our proposed DRAM network.

Declarations of interest

None.

Acknowledgments

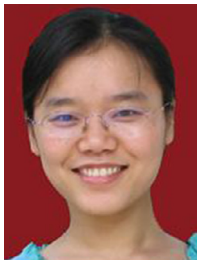
This work was supported in part by the National Natural Science Foundation of China under Grant 61871308, 61472304, 61772402 and U1605252, in part by the Fundamental Research Funds for the Central Universities, and in part by the Innovation Fund of Xidian University.

References

[1] L. Zhang, X. Wu, An edge-guided image interpolation algorithm via directional filtering and data fusion, IEEE Trans. Image Process. 15 (8) (2006) 2226–2238.
[2] K. Zhang, X. Gao, D. Tao, X. Li, Single image superresolution with non-local means and steering kernel regression., IEEE Trans. Image Process. 21 (11) (2012) 4544–4556.
[3] Y. Tang, L. Shao, Pairwise operator learning for patch-based single-image super-resolution, IEEE Trans. Image Process. 26 (2) (2017) 994–1003.
[4] X. Lu, H. Yuan, P. Yan, Y. Yuan, X. Li, Geometry constrained sparse coding for single image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1648–1655.
[5] X. Lu, Y. Yuan, P. Yan, Alternatively constrained dictionary learning for image superresolution, IEEE Trans. Cybern. 44 (3) (2014) 366–377.
[6] Y. Tang, Y. Yuan, Learning from errors in super-resolution, IEEE Trans. Cybern. 44 (11) (2014) 2143–2154.
[7] X. Lu, Y. Yuan, P. Yan, Image super-resolution via double sparsity regularized manifold learning, IEEE Trans. Circuits Syst. Video Technol. 23 (12) (2013) 2022–2033.
[8] F. Zhou, X. Li, Z. Li, High-frequency details enhancing densenet for super-resolution, Neurocomputing 290 (2018) 34–42.
[9] Y. Tang, H. Chen, Z. Liu, B. Song, Q. Wang, Example-based super-resolution via social images, Neurocomputing 172 (C) (2016) 38–47.
[10] Y. Tang, Y. Yuan, Image pair analysis with matrix-value operator, IEEE Trans. Cybern. 45 (10) (2015) 2042–2050.
[11] G. Lin, Q. Wu, L. Qiu, X. Huang, Image super-resolution using a dilated convolutional neural network, Neurocomputing 275 (2018) 1219–1230.
[12] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, in: Proceedings of the European Conference on Computer Vision (ECCV), 2014, pp. 184–199.
[13] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 2472–2481.
[14] J. Kim, J. Kwon Lee, K. Mu Lee, Accurate image super-resolution using very deep convolutional networks, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1646–1654.
[15] B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1132–1140.
[16] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intel. 38 (2) (2016) 295–307.
[17] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Deep laplacian pyramid networks for fast and accurate super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5835–5843.
[18] Y. Tai, J. Yang, X. Liu, C. Xu, Memnet: a persistent memory network for image restoration, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 4549–4557.
[19] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 105–114.
[20] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, Ntire 2017 challenge on single image super-resolution: methods and results, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1110–1121.
[21] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269.

Please cite this article as: X. Wang, Y. Gu and X. Gao et al., Dual residual attention module network for single image super resolution, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.06.078

[22] T. Tong, G. Li, X. Liu, Q. Gao, Image super-resolution using dense skip connections, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 4809–4817.
[23] M. Haris, G. Shakhnarovich, N. Ukita, Deep back-projection networks for super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 1664–1673.
[24] Z. Hui, X. Wang, X. Gao, Fast and accurate single image super-resolution via information distillation network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 723–731.
[25] Z. Hui, X. Wang, X. Gao, Two-stage convolutional network for image super-resolution, in: Proceedings of the 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 2670–2675.
[26] S.-J. Park, H. Son, S. Cho, K.-S. Hong, S. Lee, Srfeat: single image super-resolution with feature discrimination, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 455–471.
[27] J. Kim, J. Kwon Lee, K. Mu Lee, Deeply-recursive convolutional network for image super-resolution, in: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2016, pp. 1637–1645.
[28] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2790–2798.
[29] C. Dong, C.L. Chen, X. Tang, Accelerating the super-resolution convolutional neural network, in: Proceedings of the European Conference on Computer Vision (ECCV), 2016, pp. 391–407.
[30] W. Shi, J. Caballero, F. Huszar, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, 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 (CVPR), 2016, pp. 1874–1883.
[31] J. Li, F. Fang, K. Mei, G. Zhang, Multi-scale residual network for image super-resolution, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 527–542.
[32] C. Hori, T. Hori, T.-Y. Lee, Z. Zhang, B. Harsham, J. R. Hershey, T. K. Marks, Sumi, Kazuhiko, Attention-based multimodal fusion for video description, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 4203–4212.
[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems 30, 2017, pp. 5998–6008.
[34] X. Li, B. Zhao, X. Lu, Mam-rnn: multi-level attention model based rnn for video captioning, in: Proceedings of the International Joint Conferences on Artificial Intelligence Organization, 2018, pp. 2208–2214.
[35] L. Yao, A. Torabi, K. Cho, N. Ballas, C. Pal, H. Larochelle, A. Courville, Describing videos by exploiting temporal structure, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 4507–4515.
[36] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132–7141.
[37] B. Zhao, X. Li, X. Lu, Z. Wang, A cnnrnn architecture for multi-label weather recognition, Neurocomputing 322 (17) (2018) 47–57.
[38] S. Woo, J. Park, J.-Y. Lee, I. So Kweon, Cbam: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[39] K. Zhang, W. Zuo, L. Zhang, Learning a single convolutional super-resolution network for multiple degradations, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 3262–3271.
[40] Z. Wang, D. Liu, J. Yang, W. Han, T. Huang, Deep networks for image super-resolution with sparse prior, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 370–378.
[41] W.S. Lai, J.B. Huang, N. Ahuja, M.H. Yang, Fast and accurate image super-resolution with deep Laplacian pyramid networks, IEEE Trans. Pattern Anal. Mach. Intel. (2017), doi:10.1109/TPAMI.2018.2865304.
[42] M. Bevilacqua, A. Roumy, C. Guillemot, Alberi-Morel, Low-complexity single-image super-resolution based on nonnegative neighbor embedding, in: Proceedings of the British Machine Vision Conference (BMVC), 2012, pp. 135.1–135.10.
[43] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, in: Proceedings of the International Conference on Curves and Surfaces, 2012, pp. 711–730.
[44] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2001, pp. 416–423.
[45] J.-B. Huang, A. Singh, N. Ahuja, Single image super-resolution from transformed self-exemplars, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 5197–5206.
[46] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, K. Aizawa, Sketch-based manga retrieval using manga109 dataset, Multimedia Tools Appl. 76 (20) (2017) 21811–21838.



Xiumei Wang received the Ph.D. degree from Xidian University, Xi'an, China, in 2010. She is currently a Lecturer with the School of Electronic Engineering, Xidian University. Her current research interests include nonparametric statisti- cal models and machine learning. She has published several scientific articles, including the IEEE Trans. Cybernetics, Pattern Recognition, and Neurocomputing in the above areas.

Yanan Gu received the B.S. degree in Electronic Engineering from Shandong Agriculture University, China, in 2017. he is currently pursuing the M. S. degree with the School of Electronic Engineering, Xidian University. He research interests focus on Image Super-resolution.

Xinbo Gao received the B.Sc., M.Sc. and Ph.D. degrees in signal and information processing from Xidian University, China, in 1994, 1997 and 1999 respectively. From 1997 to 1998, he was a research fellow in the Department of Computer Science at Shizuoka University, Japan. From 2000 to 2001, he was a postdoctoral research fellow in the Department of Information Engineering at the Chinese University of Hong Kong. Since 2001, he joined the School of Electronic Engineering at Xidian University. Currently, he is a professor of pattern recognition and intelligent system, and Director of the VIPS Lab, Xidian University. His research interests include machine learning, computational intelligence, pattern recognition, and video content analysis. In these areas, he has published 4 books and around 100 technical articles in refereed journals and proceedings including IEEE TPAMI, TIP, TCSVT, TNN, TSMC etc.

Prof. Gao is on the editorial boards of international journals including EURASIP Signal Processing (Elsevier) and Neurocomputing (Elsevier). He served as General Chair/Co-Chair or Program Committee (PC) Chair/Co-Chair or PC member for around 30 major international conferences. He is a Fellow of the Institute of Engineering and Technology and a Fellow of the Chinese Institute of Electronics.

Zheng Hui received the B.S. degree in School of Electronic Engineering from Xidian University, China, in 2016. he is currently pursuing the Phd. degree with the School of Electronic Engineering, Xidian University. He research interests focus on Image processing, such as image super-resolution, image enhancement and image inpainting.

Please cite this article as: X. Wang, Y. Gu and X. Gao et al., Dual residual attention module network for single image super resolution, Neurocomputing, https://doi.org/10.1016/j.neucom.2019.06.078