

# Adaptive deep residual network for single image super-resolution

Shuai Liu<sup>1,\*</sup>, Ruipeng Gang<sup>1,\*</sup>, Chenghua Li<sup>2</sup> (✉), and Ruixia Song<sup>1</sup>

© The Author(s) 2019.

**Abstract** In recent years, deep learning has achieved great success in the field of image processing. In the single image super-resolution (SISR) task, the convolutional neural network (CNN) extracts the features of the image through deeper layers, and has achieved impressive results. In this paper, we propose a single image super-resolution model based on Adaptive Deep Residual named as ADR-SR, which uses the Input Output Same Size (IOSS) structure, and releases the dependence of upsampling layers compared with the existing SR methods. Specifically, the key element of our model is the Adaptive Residual Block (ARB), which replaces the commonly used constant factor with an adaptive residual factor. The experiments prove the effectiveness of our ADR-SR model, which can not only reconstruct images with better visual effects, but also get better objective performances.

**Keywords** single image super-resolution (SISR); adaptive deep residual network; deep learning

## 1 Introduction

Single Image Super Resolution (SISR) is a very classic and important task in the field of computer vision. Its main purpose is to reconstruct High Resolution (HR) image from Low Resolution (LR) image through Super Resolution (SR) technology. SISR can be widely used in safety monitoring, medical treatment, and automatic driving, etc.

1 North China University of Technology, Beijing, 100043, China. E-mail: S. Liu, 18601200232@163.com; R. Gang, gang\_rp909@sina.com; R. Song, songrx@ncut.edu.cn.

2 Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China. E-mail: lichenghua2014@ia.ac.cn (✉).

\* Shuai Liu and Ruipeng Gang contributed equally to this work.

Manuscript received: 2019-12-15; accepted: 2019-12-24

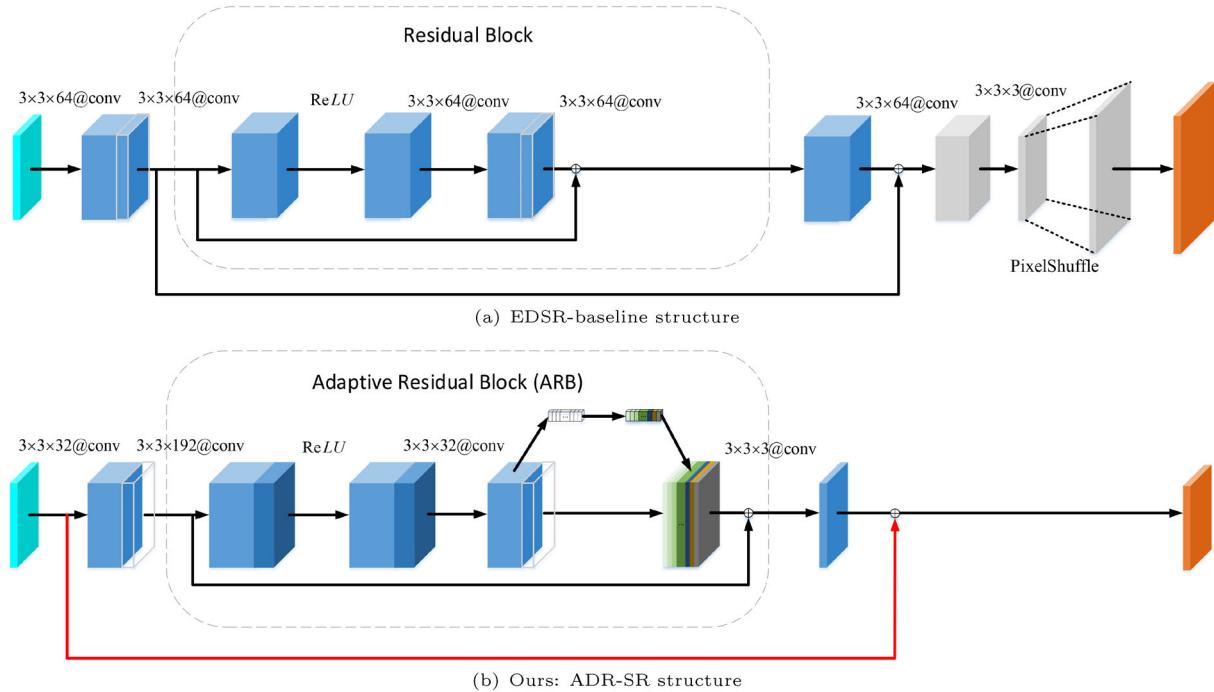
In essence, SISR is an irreversible process. At present, the simple and fast super-resolution methods mostly use light field, patch-based, and interpolation methods [1–6], all of which rely on the smooth transition assumption of adjacent pixels. However, the interpolation methods will cause aliasing and ringing effects because of image discontinuities [7].

With the development of deep learning in recent years, Convolutional Neural Network (CNN) has made breakthroughs in computer vision tasks such as classification [8], detection [9], and semantic segmentation [10]. In the field of Super-Resolution, the main feature of CNN-based methods can fit the complex mapping more directly between LR image and HR image, it enables better recovery of missing high-frequency information (such as edges, textures), so its performance goes beyond many classic methods.

Based on the EDSR [11] model, we propose a single image super-resolution model named ADR-SR, as shown in Fig. 1(b), which is a new SR model with the same size of input and output. ADR-SR releases the dependence of upsampling layers compared with the existing deep learning SR methods, and constructs a one-to-one mapping from LR pixel to HR pixel. The Adaptive Residual Block (ARB) is embedded in the ADR-SR to enhance the adaptive ability and improve the objective performance.

In summary, the main contributions of this paper are as follows:

- We propose an Input Output Same Size (IOSS) structure for the same size super-resolution task, which releases the dependence of upsampling layers compared with the existing deep learning SR methods. IOSS can solve SR task with the same input and output size as the actual needs.
- We propose an Adaptive Residual Block (ARB) based on adaptive residual factor, which solves the problem of poor adaptability caused by



**Fig. 1** Comparison of (a) EDSR-baseline structure with our (b) ADR-SR structure. Note that our ADR-SR does not have any upsampling layers and uses Adaptive Residual Block (ARB). The position of the global residuals is modified, and the depth and width of the network are also modified.

constant residual factor. Each channel in ARB has a different adaptive residual factor, and both adaptive ability and learning ability improve a lot.

- We propose a new idea for Super-Resolution network design. In some cases, adding width of the network has a significant performance improvement, and the convergence speed is faster.

## 2 Related works

### 2.1 Super-resolution model

According to whether the input size and output size are the same, the Super-Resolution model based on deep learning is divided into two types: model with different input and output size and model with the same input and output size.

The first type task: model with different input and output size, such as SRResNet [12], LapSRN [13], EDSR [11], etc, which reconstructs large image from small image. The key operation is mainly to increase the image size by the upsampling layer, in order to obtain a high-resolution output image. Currently, the commonly used upsampling layers include pixel-shuffle, transposed convolution, etc. The essence of

the first type task is to build a one-to-many mapping from LR pixel to HR pixel. The upsampling layer of EDSR is set at the end of the entire network, the feature after upsampling layer is the output image, so the EDSR increases the dependence on the upsampling layer. It is very unstable for one-to-many mapping and it cannot be better adapted to the second type task.

The second type task: model with the same input and output size, such as SRCNN [14], DRRN [15], VDSR [16], etc. They are more suitable for practical applications, such as mobile phone, camera, and other mobile devices. Due to the camera quality is low, the photos we take are not clearly, which means that Super-Resolution processing is needed. It is more in line with the needs of camera equipment that importing the captured photo directly into the network to reconstruct high-resolution photo of the same size. The second type task is the focus and difficulty of Super-Resolution research and application in the future, but there are few studies at present, and it has just begun to attract attention in recent years. When constructing a dataset, the high-resolution images are down-sampled and then up-sampled using bicubic interpolation to obtain low-resolution image of the same size. Since the input and

output are the same size, no additional upsampling layer is needed in the network, and thus we can construct a one-to-one mapping from LR pixel to corresponding HR pixel, which is more stable compared to one-to-many mapping.

The comparison between the first type and the second type shown in Fig. 2 can clearly express that the first type of model reconstructs 4 output pixels from 1 input pixel, when the scale is 2. The pixel ratio of the input and output is 1:4 (the ratio is 1:16 when the scale is 4), the information of input is seriously insufficient, the spatial position information of the output pixel also needs to be trained, and the network pressure is large and unstable. The second type of model reconstructs 1 output pixel from 1 input pixel, ensures the spatial position, and reduces the pressure of the network. The overall performance of the network is greatly improved.

## 2.2 Residual block and residual scale factor

The residual block proposed by He et al. [17] adds the learned features to the residuals, further weakens the gradient disappearance and gradient explosion in deep networks, allows us to train deeper networks successfully, and has a good performance. SRResNet [12] uses the residual block in the SR task first and deletes the ReLU activation function layer between the connected residual blocks; EDSR [11] modifies the residual block based on SRResNet, deletes the batch normalization (BN) layer, multiplies the learned features by a constant residual scale factor (default is 0.1), and then adds it to the residual. They suppress the features to reduce the change of the residual, which is more conducive to the fast convergence in the early stage of training. However, multiplying all features by a constant residual scale factor forms a simple linear mapping, and the lack of nonlinear

factor makes the network unable to handle more complex situations and reduces the learning ability.

## 2.3 Squeeze and excitation module

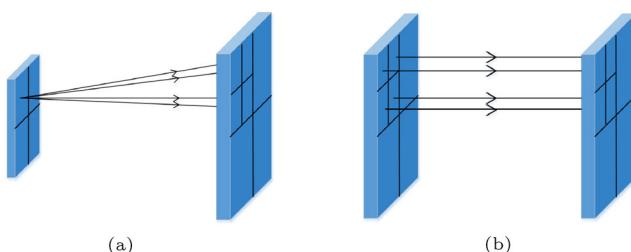
CNN is characterized by a series of convolution layers, nonlinear layers, and down-sampling layers. This structural feature enables CNN to extract features with global receptive fields. Moreover, the performance of CNN can be greatly enhanced by adding multi-scale (Inception [18]), attention [19], context (Inside–Outside [20]), and other spatial feature enhancement mechanisms.

The Squeeze and Excitation Network (SENet [21]) enhances feature extraction by building a Squeeze and Excitation (SE) module, which can clearly construct the relationship between different feature channels in the convolution layer. The SE module consists of two operations: Squeeze and Excitation. The squeeze operation compresses all 2-dimensional feature channels into 1-dimensional values by a global average pooling, in order to obtain an output vector with global corresponding features (dimensions are the same as the number of channels, assuming  $C$ ). The excitation operation learns the relationship between each channels by learning a weight vector (the dimensions are still  $C$ ). Afterwards, the SE module uses weight vectors to enhance or suppress individual feature channels.

Since the different feature maps have different image feature coding characteristics [8] (such as contour, color, region, etc.), different features have different importance to the Super-Resolution task. Therefore, the characteristic of recalibration operation in feature map of the SE module is bound to improve the performance of the Super-Resolution model. This is one of the main motivations of this paper.

## 2.4 Deeper and wider model

For the classification task, the residual network (ResNet [17]) won the championship in ILSVRC [22], and the accuracy of the model has been greatly improved. The number of layers in ResNet is 152. A deeper layer means a deeper semantic feature which has a strong effect on the network's understanding. In the Super-Resolution task, SRCNN [14] is the first network to use CNN, and it only has about 3 convolutional layers; SRResNet [12] embeds the residual block in the network, and it has 15 residual



**Fig. 2** Comparison of input and output between the first type and the second type of network. (a) The first type of network, which has different input and output sizes, reconstructs 4 output pixels from 1 input pixel. (b) The second type of network, which has the same input and output size, reconstructs 1 output pixel from 1 input pixel.

blocks; VDSR [16] uses the global residual structure to perform residual learning on the high-frequency information of the image, and uses gradient clipping to enhance the gradient transmission; meanwhile, they propose a theory of “the deeper, the better”, so the VDSR has 20 convolution layers; EDSR [11] modifies SRResNet and has 32 residual blocks, but the training time also increases.

### 3 Proposed method

We choose EDSR-baseline [11] as the base model (As shown in Fig. 1(a)). EDSR is Enhanced Deep Residual Networks, and it has been modified on the basis of SRResNet [12]; not only the number of parameters is reduced, but also the performance is significantly improved. EDSR won the first place in the internationally renowned NTIRE2017 Super Resolution Challenge, representing the highest level of the current Super-Resolution field. However, EDSR cannot solve the same size super-resolution task and has poor adaptability. In order to make up the shortcomings of EDSR, we propose a new super resolution network named ADR-SR, which uses Input Output Same Size (IOSS) structure to ensure the same size of input and output (see more details in Section 3.1), embeds Adaptive Residual Block (ARB) into the network to enhance adaptive ability (see more details in Section 3.2), follows the new design idea and increases the width of the network (see more details in Section 3.3).

#### 3.1 Network structure

In this paper, we propose an Input Output Same Size structure named IOSS for the second type task (Section 2.1). The upsampling layer in the base model is redundant because upsampling operation is not required. The convolution layer before the upsampling layer is used to expand the number of feature maps so that it can be better upsampled, but it is redundant after deleting the upsampling layer. The IOSS deletes the redundant layer, and it can not only reduce the complexity of the network, but also reduce the number of parameters. In addition, IOSS also modifies the global residuals from the first layer to the network input, in order to accommodate the second type task better. The gray layer of the base model in Fig. 1(a) is the redundant layer which is to be deleted. The IOSS structure can be applied

not only to Super-Resolution task, but also to other image processing tasks.

#### 3.2 Adaptive residual block

In order to increase the nonlinear mapping missing in the base model because of using a constant residual scale factor, we propose an Adaptive Residual Block named ARB. As shown in Fig. 1(b), ARB uses the SE module to obtain the importance of different feature channels (adaptive residual scale factors), which are used to replace the constant scale factor, so that each channel has different adaptive residual scale factor to enhance adaptive and nonlinearity. Due to the feature suppression, the advantages of rapid convergence at the beginning of training are preserved.

Therefore, the ARB can be expressed as

$$P_1 = \sigma(K^*(B_{i-1})), \quad i = 1, 2 \dots n \quad (1)$$

where  $B_{i-1}$  is the  $(i-1)$ -th output of the residual block,  $K^*$  is a convolution operation whose channel width is 192,  $\sigma$  means the activation function of *ReLU*.

$$B_i = SE(K(P_1)) + B_{i-1}, \quad i = 1, 2 \dots n \quad (2)$$

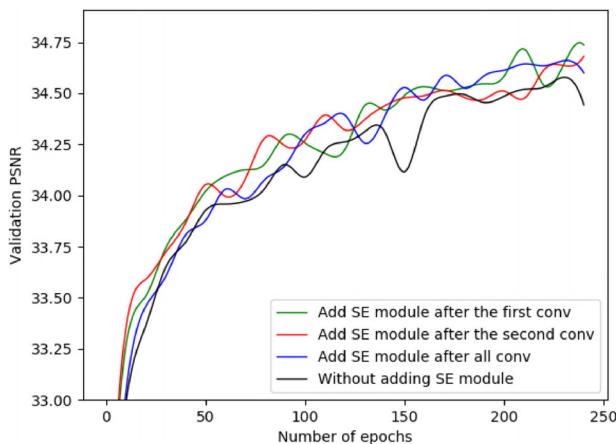
We have a  $3 \times 3$  convolution operation  $K$  for the  $P_1$  inside the local residual block, and then mix and compress feature maps into 32 channels. The outputs above continue to enter the SE module which express as *SE*. Finally, its output is added to the output of the first residual block to obtain a local residual.

The global residual of this paper can be expressed as

$$\begin{aligned} B_0 &= K(x) \\ y &= K^l(B_n \dots (B_{i-1} \dots (B_1(B_0)))) + x \end{aligned} \quad (3)$$

where  $B_0$  is the input of the local residuals,  $K^l$  is a convolution operation with a feature channel number of 3. We add the output to the LR image to form a global residual to get the final super-resolution image  $y$ .

Adding the SE module after both the first and second convolution layer in the residual block will have an effect of feature suppression, but for the former, the feature after suppression will also pass the activation function and the second convolution layer resulting in weakening the suppression effect again. For the latter, the SE module is added after the second convolution layer to suppress the feature, then we do the addition because of the residuals structure, and the suppression effect remains unchanged. The comparison of different situations is shown in Fig. 3.



**Fig. 3** Effects of adding SE module in different positions on the model.

The performance of the SE module after the second convolution layer is a little different from other cases; however, the PSNR of the validation set is small at the initial stage of training, and the model converges are faster and more stably. Therefore, the SE module is set after the second convolution layer in our ARB. It is worth noting that the PSNR of the model without the SE module is relatively low, and the additional complexity brought by the addition of the SE module is minimal (2%–10% additional parameters, <1% additional computation [21]), which also verifies the effectiveness of our ARB.

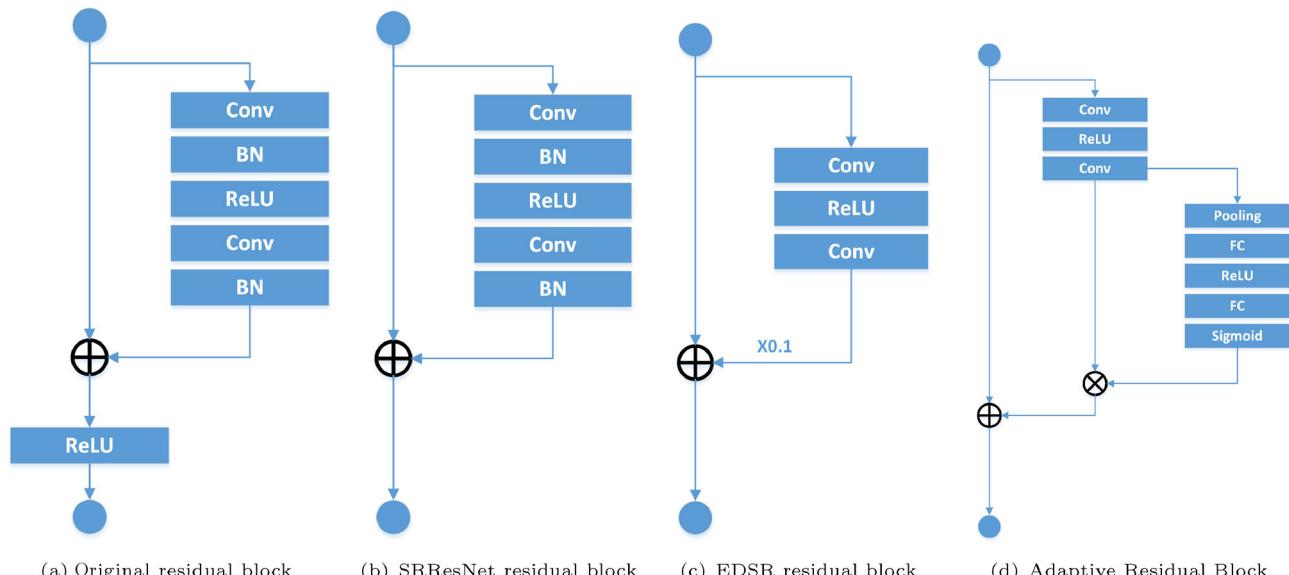
As shown in Fig. 4, we compare the residual block

structure of different models including the original residual block, the SRResNet residual block, the EDSR residual block, and our ARB.

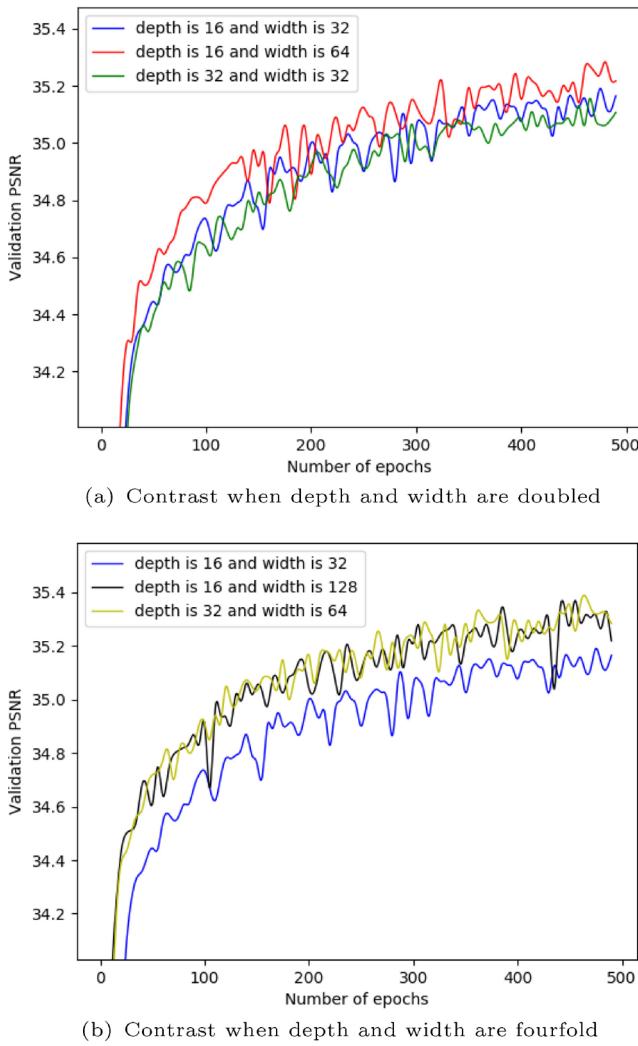
### 3.3 The increase of channel width

For Super-Resolution task, a wider network can achieve similar or even better results than a deeper network in some cases, when we construct a mapping from LR pixel to HR pixel. Excessive network layers will not bring huge upgrades, but increase training costs.

In this section, we compare the effects of increasing the width and increasing the depth on the model, where the number of parameters is approximately the same (about 0.3M). As shown in Fig. 5(a), the horizontal axis represents the number of training epochs, and the vertical axis represents the PSNR of the validation dataset. The model curve with the depth of 16 and the width of 32 is a control group. It can be clearly seen that the effect of the model with the depth of 16 and the width of 64 (the depth remains unchanged and the width is expanded by 2 times) is significantly improved in numerical value. However, the effect of the model with the depth of 32 and the width of 32 (the depth is expanded by 2 times and the width remains unchanged) is slightly lower than the control group. In Fig. 5(b), we provide another set of experiments to verify the above points.



**Fig. 4** Comparison of different residual block structures. (a) Original residual block, which is proposed in ResNet. (b) SRResNet residual block, which removes the last activation function from the original residual block. (c) EDSR residual block, which removes the BatchNormalization from the SRResNet residual block and adds a fixed factor of 0.1. (d) Ours Adaptive Residual Block (ARB), which replaces the fixed factor with the SE modules to increase adaptability.



**Fig. 5** Effects of increasing depth and increasing width on the model.

Based on the above conclusions, we propose a new idea for Super-Resolution model design. Compared with increasing the depth of the network, increasing the width of the network can better adapt to the image restoration task. Thus, the reconstruction task has equal or better enhancements than the deep features as the increase of the shallow features. Compared with the base model, we increase the width in the residual block about 3 times which is from 64 to 192, so that the model has more shallow features. In addition, in order to balance the number of parameters and training time, we also reduce the number of input feature channels of the residual block by half which is from 64 to 32. The constant  $n$  is the number of residual blocks, and our ADR-SR and the EDSR-baseline have the same number of residual blocks ( $n = 16$ ).

## 4 Experiment

### 4.1 Datasets and evaluation performances

Following the setting in EDSR [11], we train our ADR-SR on the DIV2K [23] dataset, and evaluate it on four standard benchmark datasets (including Set5 [24], Set14 [25], B100 [26], and Urban100 [27]). DIV2K has 1000 2K HD images, including 800 training images, 100 validation images, and 100 testing images. In the process of constructing LR training images, we first use the bicubic interpolation function to reduce the original HR image to different scales, and then interpolate them to the original size. In this paper, the objective signal is evaluated by two performances: Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM).

### 4.2 Training details

In the training of this paper, the RGB image is randomly cropped into a  $96 \times 96$  pixel frame as input. Data enhancement methods include: random horizontal flip, random vertical flip, and random rotation of 90 degrees. The pre-processing operations include: meanization (minus the mean of the training set to make the mean of the input is 0) and normalized (divided the variance of the training set to make the variance of the input is 1). The optimizer is Adam [28], where the hyperparameter is set to:  $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$ . The training batch size is 16, the learning rate is initialized as 0.001, which is reduced in 200k, 400k, 600k, and 800k iterations, and the scale of the reduction is 0.5. The loss function is L1 loss. The training environment is NVIDIA Titan XP GPUs and PyTorch framework.

### 4.3 Experimental results

As shown in Table 1, we test the performance of different algorithms on standard benchmark datasets and give quantitative evaluation results. The first line gives a comparison model, including LapSRN [13], VDSR [16], DRRN [15], SRResNet [12], SRDenseNet [29], CARN [30], MCAN [31], EDSR-baseline [11], and our ADR-SR. The first and second columns represent different benchmark datasets and corresponding scales. The table gives the quantitative value (PSNR/SSIM) results of the various models in different datasets and different scale settings, where the optimal results are shown in bold.

In order to ensure the fairness of the experiment

**Table 1** Quantitative comparison with the state-of-the-art methods based on  $\times 2$ ,  $\times 4$  SR with bicubic degradation model

Scale	Method	Set5		Set14		B100		U100		DIV2K Val	
		PSNR	SSIM								
$\times 2$	LapSRN	34.80	0.936	29.60	0.869	25.92	0.790	26.83	0.868	33.34	0.925
	VDSR	34.96	0.936	30.53	0.880	29.58	0.870	28.53	0.893	33.45	0.925
	DRRN	35.01	0.936	30.61	0.882	29.90	0.877	28.88	0.898	33.62	0.927
	CARN	35.62	0.942	30.15	0.881	25.88	0.793	27.77	0.886	34.50	0.936
	MCAN	35.80	0.943	30.26	0.882	25.85	0.792	28.03	0.888	34.58	0.937
	EDSR baseline	35.78	0.943	30.19	0.881	25.89	0.793	27.82	0.887	34.62	0.937
$\times 4$	<b>ADR-SR</b>	<b>35.82</b>	<b>0.943</b>	<b>31.41</b>	<b>0.893</b>	<b>30.12</b>	<b>0.880</b>	<b>30.04</b>	<b>0.914</b>	<b>34.64</b>	<b>0.937</b>
	LapSRN	29.39	0.849	25.30	0.699	24.83	0.677	22.22	0.688	28.17	0.799
	VDSR	28.90	0.832	25.86	0.707	25.86	0.707	23.23	0.703	27.92	0.784
	SRResNet	29.76	0.855	25.36	0.704	24.83	0.679	22.21	0.696	28.10	0.806
	SRDenseNet	29.03	0.838	25.70	0.713	25.68	0.695	23.54	0.725	27.96	0.790
	DRRN	29.21	0.838	26.04	0.713	25.66	0.683	23.54	0.717	28.07	0.789
	CARN	30.23	0.864	25.66	0.712	24.98	0.684	22.72	0.711	28.96	0.818
	MCAN	<b>30.42</b>	<b>0.865</b>	25.73	0.714	24.94	0.685	22.90	0.718	<b>29.03</b>	<b>0.820</b>
	EDSR baseline	30.18	0.863	25.65	0.712	24.96	0.684	22.69	0.711	28.96	0.818
	<b>ADR-SR</b>	30.21	0.863	<b>26.71</b>	<b>0.741</b>	<b>26.08</b>	<b>0.706</b>	<b>24.48</b>	<b>0.760</b>	28.94	0.817

data, we reproduce the test results of the comparison model and obtain the relevant experiment data, and all the pre-training comparison models are derived from the network open source. The datasets are constructed by the bicubic interpolation function, and PSNR and SSIM are calculated on the three channels of RGB space. There is a slight deviation from the original data, due to the different construction methods of the partial comparison model and some of the original paper calculate PSNR and SSIM on the Y channel of the YCbCr space.

It can be seen from Table 1 that in the tasks of different training sets with scale 2, our ADR-SR is optimal in objective performances, and the PSNR and SSIM are higher than the second method. Due to the error caused by different data construction methods, in the DIV2K validation dataset of scale 4, ADR-SR has a small lower of 0.02 and 0.001 on the PSNR and SSIM compared with the EDSR-baseline model, but on other datasets, our ADR-SR are higher obviously.

Experiment shows that our ADR-SR achieves relatively good visual effects and objective performances on different scale tasks of different standard benchmark datasets, and ADR-SR has obvious advantages in image clarity, spatial similarity, and image texture details.

In order to verify the validity of the ADR-SR, we take the Urban100 and DIV2K datasets as examples

to select some images, and compare them with LapSRN, VDSR, DRRN, CARN, MCAN, and EDSR-baseline. The bicubic interpolation method is also shown as a reference. As shown in Fig. 6 and Fig. 7, the red dotted box highlights the obvious advantages of our ADR-SR. It can be clearly seen from the experiment results that our ADR-SR has a better Super-Resolution effect than other models when dealing with the edge of the object. The edge distinction is more significant, the detail information missing from many other models is reconstructed, and the visual effect is greatly improved.

## 5 Conclusions

In summary, we propose a single image super-resolution model named ADR-SR based on adaptive deep residual, which can be used for super-resolution task with the same size of input and output image. The visual effects and objective performances of the experiment demonstrate the effectiveness of our ADR-SR. The specific innovations are: (1) Input Output Same Size structure (IOSS) for same size super-resolution task. (2) Adaptive Residual Block (ARB), the adaptive ability and convergence speed improve a lot. (3) A new idea for super-resolution network design increases the width of the network instead of the depth to obtain additional performance improvements.

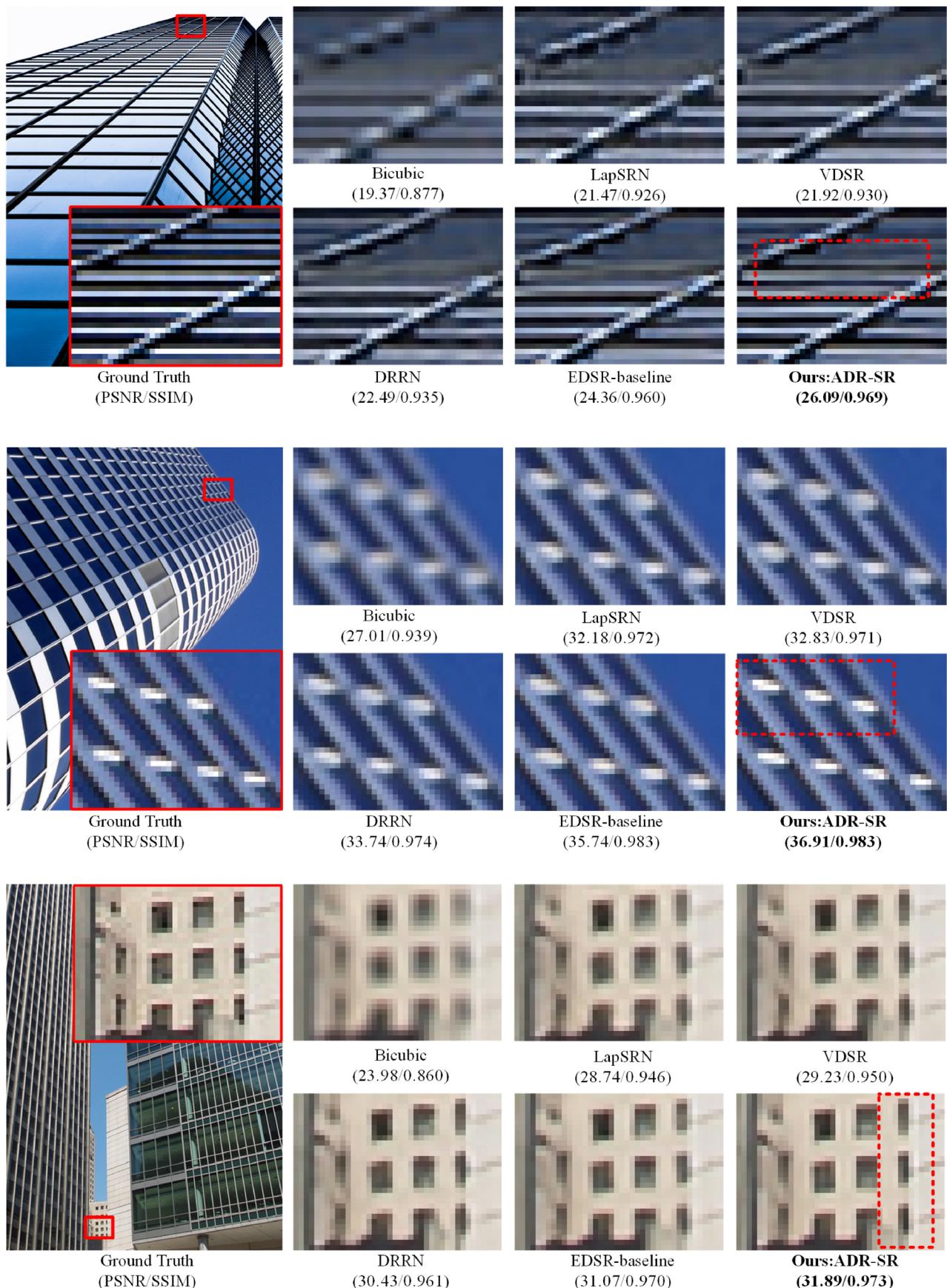
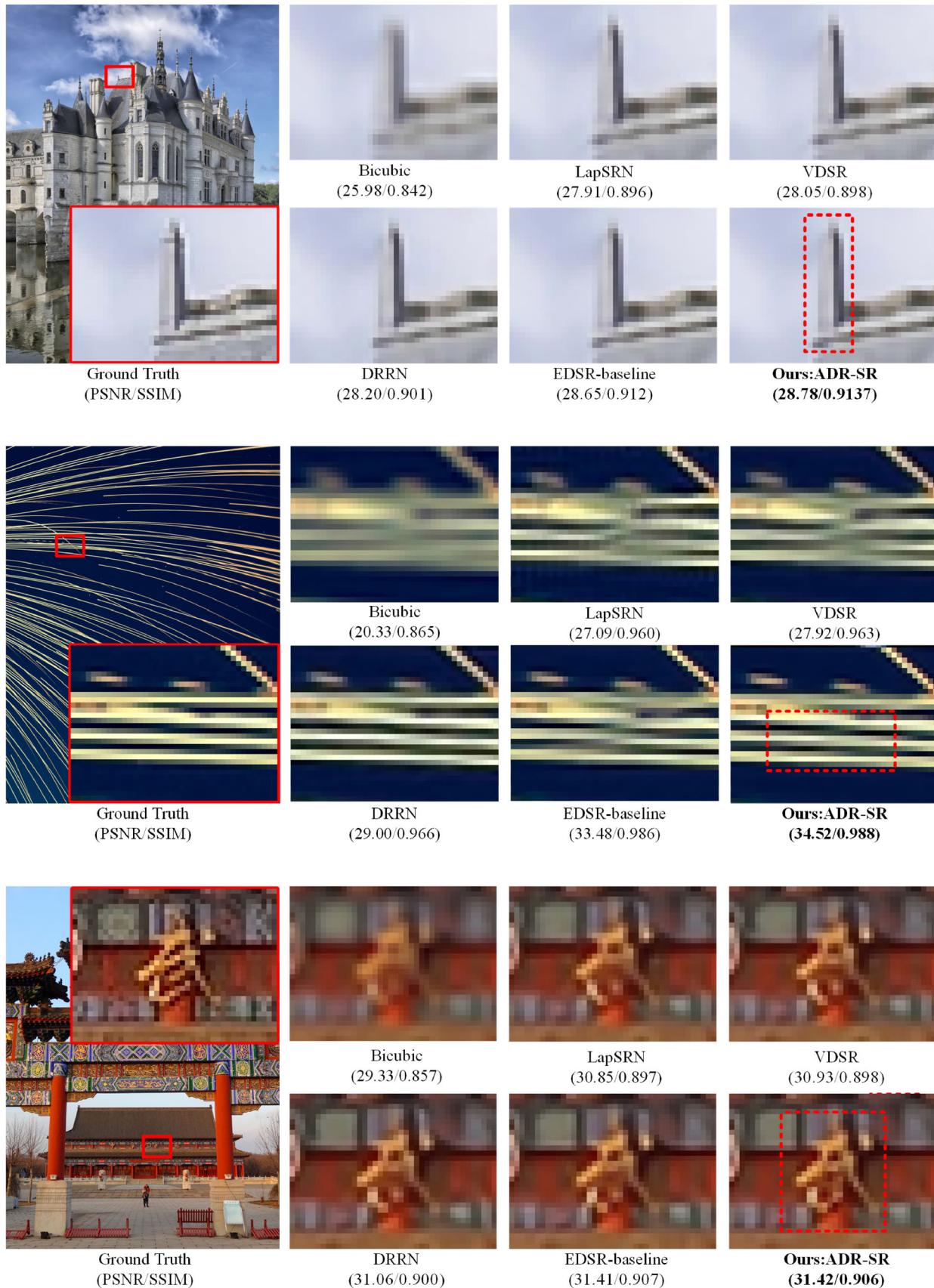


Fig. 6 Comparison of experimental effects of Urban100 dataset.



**Fig. 7** Comparison of experimental effects of DIV2K validation dataset.

## Acknowledgements

This work was supported in part by National Natural Science Foundation of China (No. 61571046) and National Key R&D Program of China (No. 2017YFF0209806).

## References

- [1] Su, B.; Jin, W. POCS-MPMAP based super-resolution image restoration. *Acta Photonica Sinica* Vol. 32, No. 4, 502–504, 2003.
- [2] Choi, J.-S.; Kim, M. Single image super-resolution using global regression based on multiple local linear mappings. *IEEE Transactions on Image Processing* Vol. 26, No. 3, 1300–1314, 2017.
- [3] Yang, J. C.; Wang, Z. W.; Lin, Z.; Cohen, S.; Huang, T. Coupled dictionary training for image super-resolution. *IEEE Transactions on Image Processing* Vol. 21, No. 8, 3467–3478, 2012.
- [4] Yang, J. C.; Wright, J.; Huang, T. S.; Ma, Y. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing* Vol. 19, No. 11, 2861–2873, 2010.
- [5] Zhang, F. L.; Wang, J.; Shechtman, E.; Zhou, Z. Y.; Shi, J. X.; Hu, S. M. PlenoPatch: Patch-based plenoptic image manipulation. *IEEE Transactions on Visualization and Computer Graphics* Vol. 23, No. 5, 1561–1573, 2017.
- [6] Barnes, C.; Zhang, F.-L. A survey of the state-of-the-art in patch-based synthesis. *Computational Visual Media* Vol. 3, No. 1, 3–20, 2017.
- [7] Liu, H.; Fu, Z. L.; Han, J. G.; Shao, L.; Hou, S. D.; Chu, Y. Z. Single image super-resolution using multi-scale deep encoder-decoder with phase congruency edge map guidance. *Information Sciences* Vol. 473, 44–58, 2019.
- [8] Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the Advances in Neural Information Processing Systems 25, 1097–1105, 2012.
- [9] Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 39, No. 4, 640–651, 2017.
- [10] Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580–587, 2014.
- [11] Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K. M. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 136–144, 2017.
- [12] Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Tora, J.; Wang, Z.; Shi, W. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4681–4690, 2017.
- [13] 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, 624–632, 2017.
- [14] Dong, C.; Loy, C. C.; He, K. M.; Tang, X. O. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 38, No. 2, 295–307, 2016.
- [15] Tai, Y.; Yang, J.; Liu, X. M. Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3147–3155, 2017.
- [16] 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, 1646–1654, 2016.
- [17] He, K. M.; Zhang, X. Y.; Ren, S. Q.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In: Proceedings of the IEEE International Conference on Computer Vision, 1026–1034, 2015.
- [18] Szegedy, C.; Liu, W.; Jia, Y. Q.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1–9, 2015.
- [19] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In: Proceedings of the Advances in Neural Information Processing Systems 30, 5998–6008, 2017.
- [20] Bell, S.; Zitnick, C. L.; Bala, K.; Girshick, R. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2874–2883, 2016.
- [21] Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7132–7141, 2018.



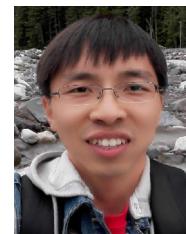
- [22] Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S. A.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A. C.; Fei-Fei, L. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* Vol. 115, No. 3, 211–252, 2015.
- [23] Agustsson, E.; Timofte, R. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 126–135, 2017.
- [24] 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 23rd British Machine Vision Conference, 135.1–135.10, 2012.
- [25] Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations. In: *Curves and Surfaces 2010. Lecture Notes in Computer Science*, Vol. 6920. Boissonnat, J. D. et al. Eds. Springer Berlin Heidelberg, 711–730, 2012.
- [26] Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the 8th IEEE International Conference on Computer Vision, Vol. 2, 416–423, 2001.
- [27] Huang, J.-B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5197–5206, 2015.
- [28] Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [29] Tong, T.; Li, G.; Liu, X. J.; Gao, Q. Q. Image super-resolution using dense skip connections. In: Proceedings of the IEEE International Conference on Computer Vision, 4799–4807, 2017.
- [30] Ahn, N.; Kang, B.; Sohn, K.-A. Fast, accurate, and lightweight super-resolution with cascading residual network. In: *Computer Vision – ECCV 2018. Lecture Notes in Computer Science*, Vol. 11214. Ferrari, V.; Hebert, M.; Sminchisescu, C.; Weiss, Y. Eds. Springer Cham, 256–272, 2018.
- [31] Ma, H.; Chu, X.; Zhang, B.; Wan, S. A matrix-in-matrix neural network for image super resolution. *arXiv preprint arXiv:1903.07949*, 2019.



**Shuai Liu** received his B.S. degree from North China University of Technology, Beijing, China, in 2018. He is now a postgraduate student in North China University of Technology. His current interests include deep learning and image processing. He win the ICCV 2019 AIM Challenge on Constrained Super-Resolution Runner-up Award.



**Ruipeng Gang** received her B.S. degree from North China University of Technology, Beijing, China, in 2017. She is now a postgraduate student in North China University of Technology. Her current interests include deep learning and image processing. She win the ICCV 2019 AIM Challenge on Constrained Super-Resolution Runner-up Award.



**Chenghua Li** is a research asisitant in Institute of Automation, Chinese Academy of Sciences (CASIA). He received his Ph.D. degree in National Laboratory of Pattern Recognition from CASIA, Beijing, in 2018. He received his B.S. degree from North China University of Technology, Beijing, China in 2014, and his M.S. degree from Anyang Normal University, Henan Province, China in 2011. His research interests include deep learning and compute vision.



**Ruixia Song** graduated from Beijing Normal University in 1980, and received her master degree in science from Hangzhou University (now Zhejiang University) in 1990. She is now a professor in North China University of Technology. Her major areas of research interest include computer graphics, image processing, and pattern recognition, etc.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Other papers from this open access journal are available free of charge from <http://www.springer.com/journal/41095>. To submit a manuscript, please go to <https://www.editorialmanager.com/cvmj>.