



# Image super-resolution reconstruction based on improved Dirac residual network

Xin Yang<sup>1</sup> · Tangxin Xie<sup>1</sup> · Li Liu<sup>1</sup> · Dake Zhou<sup>1</sup>

Received: 11 September 2020 / Revised: 24 January 2021 / Accepted: 22 February 2021 /  
Published online: 19 March 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

In recent years, to improve the nonlinear feature mapping ability of the image super-resolution network, the depth of the convolutional neural network is getting deeper and deeper. In the existing residual network, the the residual block's output and input are added directly through the skip connection to deepen the nonlinear mapping layer. However, it can not be proved that every addition is useful to improve the network's performance. In this paper, based on Dirac convolution, an improved Dirac residual block is proposed, which uses the trainable parameters to adaptively control the balance of the convolution and the skip connection to increase the nonlinear mapping ability of the model. The main body network uses multiple Dirac residual blocks to learn the nonlinear mapping of high-frequency information between LR and HR images. In addition, the global skip connection is realized by sub-pixel convolution, which can learn to use linear mapping of low-frequency features of input LR image. In the training stage, the model uses Adam optimizer for network training and L1 as the loss function. The experiments compare our algorithm with some other state-of-the-art models in PSNR, SSIM, IFC, and visual effect on five different benchmark datasets. The results show that the proposed model has excellent performance both in subjective and objective evaluation.

**Keywords** Super-resolution · Deep learning · Convolutional neural network · Dirac residual network

## 1 Introduction

Super-resolution (SR) technology refers to the reconstruction of high resolution (HR) images or videos from one or more low resolution (LR) images of the same scene. SR can be generally classified into three types according to input and output, such as single input single output (SISO), multiple input single output (MISO), and multiple input multiple output (MIMO). It can also be directly divided into two categories, such as single image SR

---

✉ Xin Yang  
yangxin@nuaa.edu.cn

<sup>1</sup> College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

reconstruction (SISR) and multi-frame image SR reconstruction. SISR refers to estimating an HR image with a given single LR image if the original image cannot be acquired.

SISR is widely used in tasks of image processing, such as security and surveillance imaging (Zou & Yuen, 2011), medical imaging (Shi et al., 2013), remote sensing image processing (Yang et al., 2011), and so on. There are many methods to implement the image SR. For the same LR image, different ways often lead to different HR images. The simplest SR method is analytical interpolation methods, such as linear interpolation, bicubic interpolation, etc., which takes the average of the pixels in the known LR image as the missing pixel of the HR image. Analytic interpolation works well in the smooth region of the image. However, it has a weak effect on the image edge area, which results in ringing and blurring. In addition to the analytical interpolation method, the learning-based and reconstruction-based SR methods, such as sparse coding (Yang et al., 2010), neighborhood embedded regression (Chang et al., 2004; Timofte et al., 2013), random forest (Schulter et al., 2015), have better reconstruction effects. Most of the newly proposed algorithms take advantage of deep learning. The latest SR method based on deep learning has achieved amazing reconstruction results, which has attracted extensive attention from researchers.

The SR convolutional neural network (SRCNN) (Dong et al., 2014) is the first SR algorithm based on convolutional neural network (CNN). SRCNN directly learns the pixel-to-pixel mapping between LR image blocks and HR image blocks. The interpolated LR image is used as input and mapped to the feature map through a convolutional layer. The entire network has three convolution layers for nonlinear mapping. The reconstruction performance of the network is superior to the most advanced existing algorithms. Subsequently, Dong et al. continue to propose a fast SR convolutional neural network (FSRCNN) (Dong et al., 2016), which can directly use the input LR image. Also, FSRCNN uses more convolution kernels for nonlinear mapping and introduces a deconvolution layer at the end to reconstruct the HR image.

Kim et al. construct a very deep network for SR (VDSR) (Kim et al., 2016a) and a deeply-recursive convolutional network (DRCN) (Kim et al., 2016b). In VDSR, the author believes that the LR image is similar to the HR image in low-frequency information so that it is more efficient only to learn the high-frequency residual between them during training. DRCN recurs the same convolution layer 16 times. When the depth increases, it avoids introducing additional parameters and increasing the intermediate loss function.

Shi et al. (2016) propose an efficient sub-pixel convolutional neural network (ESPCN). In this model, several convolution layers are used to learn the image features for input LR images. Then HR images are reconstructed using a novel sub-pixel convolution layer according to the convolution features learned from the deep convolution network.

Ledig et al. (2017) design an SR generative adversarial network (SRGAN). Instead of using the usual L2 norm, the network design a loss function which accords with the characteristics of human visual perception. In addition, a residual network (RESNET) is introduced into the whole system to learn image features more effectively. Experiments show that SRGAN can restore realistic textures like photographs from larger down-sampled LR images.

Lim et al. (2017) construct an enhanced deep SR network (EDSR), which is based on RESNET (Ledig et al., 2017). EDSR modifies the residual structure, which includes removing the batch normalization layer (BN layer), increasing the dimension of each convolution feature, scaling the residual after each residual block, and reconstructing with sub-pixel convolution layer.

Tong et al. (2017) present a SR dense network (SRDenseNet), which introduces the skip connection into a very deep neural network. The network propagates the feature map of

each convolutional layer to subsequent layers and upsamples by deconvolution at the end, which alleviates the vanishing gradient problem.

The residual dense network (RDN) (Zhang, Li, et al., 2018) fully utilizes the features from all the convolution layers. The network adaptively learns more significant features based on local feature fusion technology. In addition, the global feature fusion is used to determine global hierarchical features holistically and adaptively. Zhang et al. (Zhang, Tian, et al., 2018) also propose the residual channel attention network (RCAN), which adaptively adjusts channel-wise features through channel attention mechanism, making the network focus on learning high-frequency information.

Yu et al. (1808) prove that models with the Relu activation function and more features have better performance when the parameters and computational load are the same. On this basis, Wide activation SR(WDSR) (Yu et al. 1808) network is proposed, in which there is a wider channel before the activation function of each residual block. Also, the weight normalization (WN) layer is designed to improve the accuracy of the network. Wang et al. (1904) further propose an adaptive weighted SR network (AWSRN), which devises a local fusion block for more efficient residual learning. In addition, an adaptive weighted multi-scale module is developed to reconstruct features.

Cao et al. (2019) propose an improved deep residual network (IDRN), which can modify the residual structure and skip connection easily and effectively. Besides, the model uses a new energy-aware (EA) training loss function and lightweight network architecture to obtain fast and accurate results. Zhang et al. propose a deep plug-and-play SR network (DPSR) (Zhang et al., 1903), which can process LR images with arbitrary blur kernels. Zhang et al. (2019) also use the optical zoom to obtain real sensor data for model training. Xu et al. (2019) generate training data by simulating the imaging process of a digital camera. Experiments demonstrate that SR with raw data helps recover fine details and clear structures. The deep back-projection network (DBPN) (Haris et al., 1904) exploits the iterative up-sampling and down-sampling layers to represent different types of image degradation and image reconstruction components to solve the interdependence between LR and HR images. The SR feedback network (SRFBN) (Li & Yang, 1903) proposed by Li et al. adopt the recurrent neural network (RNN) with the constraints to process feedback information and perform feature reuse. Dai et al. (2019) propose a second-order attention network (SAN). A new second-order channel attention module (SOCA) designed by the network uses second-order feature statistics to adjust channel characteristics adaptively. Furthermore, the model also constructs a non-locally enhanced residual group structure to learn more abstract feature representation.

To deepen the nonlinear mapping layer of the network, the output and input of the residual block are directly added by the skip connection in the existing RESNET. However, it can't be proved that every addition in the network is useful. It will undoubtedly affect the network's fitting ability to SR task, and then affect the reconstruction effect. To make the network adaptively adjust the proportion of the convolution feature and the skip connection in each level of the residual block output, we propose a new residual block (Res-block) for image SR based on Dirac convolution. It can use the trained parameter adaptive control the weights of the convolution feature and the skip connection, so as to increase the nonlinear mapping ability of the network.

In summary, we construct a novel Dirac Residual SR(DRSR) network for the SISR task in this paper. The model uses the Dirac residual layer to learn the high-frequency features of the input LR image, uses the global skip connection to utilize the low-frequency feature of the input LR image directly, and reconstructs the image by sub-pixel convolution. Then, DRSR improves the residual layer of the traditional SR algorithm by weight

parameterization. Finally, the convolution feature of the input image and the learning feature of the RESNET are combined to reconstruct the output HR image. Our network does not only add hyperparameters to the branches of the two networks, but also we design a new SR network which is derived from ResNet. It is also an attempt to non-skip connection to find another way to implement residuals.

## 2 Proposed method

### 2.1 The original Dirac block

In deep learning feild, the network with large depth means that it has a strong nonlinear fitting ability. However, the depth of the network can not be increased unlimitedly. We need to train the depth neural network model through backpropagation. The gradient of each layer in the network is trained on the basis of the previous layer. Multilayer neural networks often need to face the problem of gradient disappearing, which shows that the more layers the network has, the greater the model error. RESNET provides a new way to solve the gradient disappearance (He et al., 2016). By adding the skip connections to the standard feedforward neural network, the RESNET can bypass some layers. In this way, a neural network with high depth can be built to pursue better performance. The advantage of the residual block is that it can make the network deeper. However, it can not be proved that it is useful to connect the feature map of each layer to the next layer. So RESNET has limits. When the network reaches a certain depth, deepening the network can not improve the accuracy. The structure of the residual block is shown in Fig. 1.

As shown in the figure,  $x$  and  $y$  are the input and output, respectively. In addition to the convolution layer, there are activation functions Relu and BN (Batch normalization) layers in the residual block. The function of the BN layer is to reduce the difficulty of model training. The input  $x$  is convoluted by two layers to get  $F(x)$ . Then by a skip connection, the summing of  $x$  and  $F(x)$  are linked to the activation function *Relu* to obtain the final output  $y$ . The residual structure can be expressed as

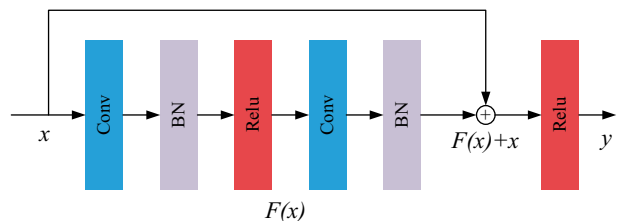
$$y = F(x) + x \tag{1}$$

$F(x) = f_{BN}(relu(w_1 * x + b_1) * w_2 + b_2)$ , where  $w_i$  is the convolution kernel of the  $i$ th convolution,  $b_i$  is the corresponding bias term, and *relu* is the activation function ReLU,  $f_{BN}$  is BN layer function. Then  $y$  can be written as

$$y = relu[f_{BN}(relu(w_1 * x + b_1) * w_2 + b_2) + x] \tag{2}$$

Our DRSR attempts to integrate residual connection into convolution operation through parameterization (Zagoruyko & Diracnets, 1706). In Eq. (1), the residual connection

Fig. 1 Residual block structure in Resnet



$y = F(x) + x$  is a linear operation, and in Eq. (2), the convolution operation is also a linear operation. We assume that  $F(x)$  in the RESNET is a single convolution layer. In addition, in order to express concisely, we omit the bias term, then the residual can be expressed as:

$$y = W * x + x \quad (3)$$

where  $*$  represents convolution operation,  $x$  is the input feature map,  $y$  is the output feature, and  $W$  is the convolution parameter matrix.

## 2.2 DRSR Res-block

In order to increase the adaptability of the network, we use the method of Dirac parameterization to combine the skip connection into the convolution parameter matrix and add the control parameters  $\alpha$  and  $\beta$ . Then we have

$$\begin{aligned} y &= \hat{W} * x \\ \hat{W} &= \alpha \times I + \beta \times W_{norm} \end{aligned} \quad (4)$$

where  $\hat{W}$  represents the combined convolution parameter matrix,  $I$  is the unit matrix, which represents the skip connection in Resnet.  $W_{norm}$  represents the normal convolution parameter matrix.  $\alpha$  and  $\beta$  are trainable parameters, which control the weight of convolution operation and the connection, respectively. If  $\alpha$  approaches 0, the convolution is dominant. On the contrary, if  $\beta$  approaches to 0, it means that the skip connection is dominant.

Because  $\alpha$  and  $\beta$  are trainable, Dirac residual can adaptively change the weight of convolution and skip connection output in the training process, so as to achieve the purpose of adaptive learning. According to this characteristic, we propose an improved Dirac Res-block.

Figure 2 shows the structure comparison between the EDSR residual block (Lim et al., 2017) and the residual block used by our DRSR model. In each residual block of EDSR, the skip connection is realized by directly connecting the input to the output. In addition, each residual block is scaled to one-tenth of its original size by the residual scaling layer (Mult), which makes the training more stable. In the proposed DRSR Res-block, the skip connection is realized by the parameterization method of Eq. (4), which is also given in Fig. 2. In summary, DRSR Res-block is equivalent to adding control parameters  $\alpha$  and  $\beta$  to the convolution layer and the skip connection in the residual block of single-layer convolution.

Because  $\alpha$  and  $\beta$  are parameters that can be trained, the model will adjust the value of  $\alpha$  and  $\beta$  adaptively in the actual training process. It can control the weight of each layer of the model and avoid connecting the convolution output features directly to the next layer.

## 2.3 Model

DRSR is divided into two parts: the deep DRSR residual feature reconstruction network and the global skip connection reconstruction network, as shown in Fig. 2. In this paper, the deep DRSR residual feature reconstruction network is referred to by the upper-part network, which is divided into the feature extraction layer, DRSR Res-body, and sub-pixel convolution layer. The global skip connection network is referred to by the lower-part network, which is divided into feature extraction layer and sub-pixel convolution layer.

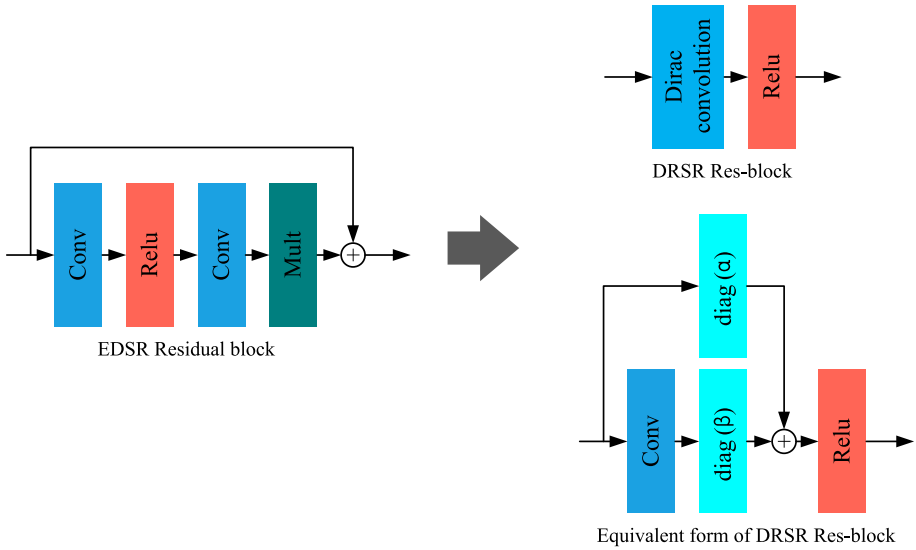


Fig. 2 EDSR residual block and DRSR Res-block

DRSR uses sub-pixel convolution as the reconstruction layer, as shown in Fig. 3. The skip connection of the network reconstructs the low-frequency part of HR the image by using the low-frequency features of the LR image so that the DRSR Res-body concentrates on learning the high-frequency residual of the HR image.

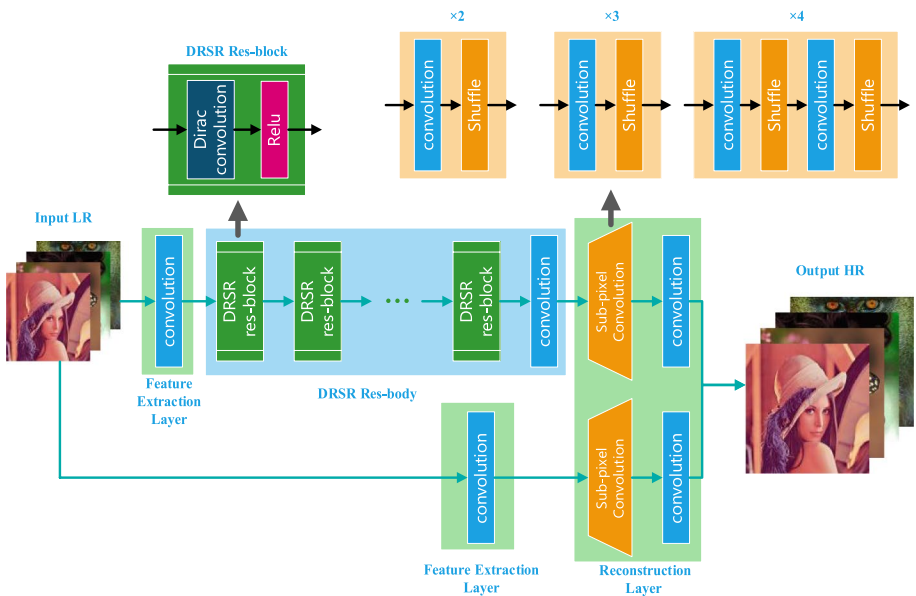


Fig. 3 the Network model structure of DRSR

DRSR consists of two network branches including the deep Dirac residual feature reconstruction branch and the global skip connection reconstruction branch. As shown in Fig. 3, the input of the entire model is an LR image, and the output is the sum of the reconstructed images of the two branches.

For the deep Dirac residual feature reconstruction branch, its input is LR image. Define  $I^{LR}$  as the input LR image,  $I^{HR}$  as the original HR image, and  $I^{SR}$  as the reconstructed HR image. The Dirac residual feature reconstruction branch first extracts the shallow features through a convolutional layer, which is expressed as

$$F_1 = E_{SFE}(I^{LR}). \tag{5}$$

where  $E_{SFE}(\cdot)$  is the shallow feature extraction operation,  $F_1$  is the extracted shallow feature. By inputting the extracted shallow features into the Dirac residual block, we have

$$F_2 = E_{Dirac}(F_1) \tag{6}$$

where  $E_{Dirac}(\cdot)$  represents the feature extraction operation of the Dirac residual block. In order to obtain more features of the LR image and focus on learning the high-frequency information in the LR image, we cascade 80 Dirac residual blocks to obtain a deep network. By inputting the features extracted from the Dirac residual block  $F_2$  into the sub-pixel convolution amplification module, we can obtain

$$F_{Dirac}^{SR} = E_{PS}(F_2), \tag{7}$$

where  $E_{PS}(\cdot)$  and  $F_{Dirac}^{SR}$  are the images reconstructed by the sub-pixel convolution amplification module and the Dirac residual branch, respectively.

For the global skip connection reconstruction branch network, the input image is still the LR image. The shallow feature is extracted through a convolutional layer and then directly amplified by a sub-pixel convolutional layer. It can allow the reconstruction network to pay more attention to the reconstruction of low-frequency features in the LR image. The whole process can be expressed as

$$F_{GS}^{SR} = H_{PS}(H_{SFE}(I^{LR})), \tag{8}$$

where  $F_{GS}^{SR}$  is the image reconstructed by the global skip connection branch,  $H_{SFE}(\cdot)$  and  $H_{PS}(\cdot)$  are the shallow feature extraction operation and the sub-pixel convolution amplification operation.

The output of the entire model is the sum of the image reconstructed by the global skip connection branch and the image reconstructed by the Dirac residual branch, which is expressed as

$$I^{SR} = F_{GS}^{SR} + F_{Dirac}^{SR}. \tag{9}$$

## 2.4 Training

We use the public DIV2K and Flickr2K data sets as the training sets of the network. DIV2K includes 800 training images, 100 validation images, and Flickr2K includes 2650 training images. So there are 3450 2 K images in our training set. During the training, the 801th–810th images in DIV2K are selected as the validation set, and the model with the best PSNR is saved.

After a lot of training experiments, it is shown that if the residual of Dirac is greater than 64 layers and  $\alpha, \beta$  is set to 1, the loss value of the model is very large at the beginning of training, which is not conducive to convergence. When  $\alpha=\beta=0.1$ , the gradient disappears and the model cannot converge during the training. When  $\alpha=1, \beta=0.1$  or  $\alpha=0.1, \beta=1$ , the model can be trained normally. In this paper, we set  $\alpha=1, \beta=0.1$ .

Before training, the HR image is reduced to the LR image by bicubic interpolation. The LR image is input directly into the network, and the corresponding HR' image is reconstructed. For the SR task,  $L_1$  loss function,  $L_2$  loss function and perceptual loss function are common. The  $L_1$  and  $L_2$  loss are expressed as

$$L_1 = \frac{1}{N} \sum_{i=1}^N |I_i^{SR} - I_i^{HR}|, \quad (10)$$

$$L_2 = \frac{1}{N} \sum_{i=1}^N (I_i^{SR} - I_i^{HR})^2, \quad (11)$$

where  $I_i^{SR}$  and  $I_i^{HR}$  are the pixel in  $I^{SR}$  and  $I^{HR}$ , respectively.

The perceptual loss function is often used to evaluate the visual perception quality of an image, which is usually set according to specific SR model. For  $L_2$  loss function, the previous researches have proven that it is not as effective as  $L_1$  loss function. Therefore, we choose  $L_1$  to optimize our model. ADAM is set to the optimizer, and the two parameters  $r_1$  and  $r_2$  in ADAM are set to 0.9 and 0.999, respectively. The learning rate is initially set to  $1 \times 10^{-4}$ , and then it is halved every  $2 \times 10^5$  iterations. The total number of iteration is  $6 \times 10^5$ . 16 RGB image blocks of size  $48 \times 48$  are input for each iteration. So the input size is  $[16, 48, 48, 3]$ . A single RTX2080Ti graphics card (11 GB memory) is used in training. On Ubuntu 18.4 system, Pytorch 1.1.0, CUDA 10.0, and cuDNN 7.5.0 are exploited as the deep learning frameworks. It takes about four days to complete the model training.

### 3 Experimental results

To verify the validity of the model, we take the Set5, Set14, B100, Urban100, and DIV2K data sets as the test sets, and compares with bicubic interpolation, Aplus(Timofte et al., 2014), self-exemplars SR(SelfExSR)(Huang et al., 2015), SRCNN(Dong et al., 2014), laplacian pyramid SR network (LapSRN) (Lai et al., 2017), DRCN (Kim et al., 2016b), deep recursive residual network (DRRN) (Tai et al., 2017a), VDSR (Kim et al., 2016a), MemNet(Tai et al., 2017b), Two-stage convolutional network (TSCN)(Hui et al., 2018), and EDSR (Lim et al., 2017) (without image self-integration) algorithm in terms of Peak Signal to Noise Ratio (PSNR), Structural SIMilarity (SSIM), IFC (Information fidelity criterion) and visual effects.

Table 1 provides the comparison of PSNR and SSIM on the Y channel with the magnification factors of  $\times 2, \times 3$ , and  $\times 4$  on Set5, Set14, B100, Urban100, and DIV2K test sets. The experiments are obtained from the MATLAB program. Red and blue indicate the best and second-best performance, respectively. From the table, it can be seen that the performance of DRSR is slightly better than that of EDSR, and has a certain improvement compared with that of other algorithms.

Table 2 shows the comparison of IFC (Sheikh et al., 2005) on the Y channel with magnification factors of  $\times 2, \times 3$ , and  $\times 4$  on Set5, Set14, B100, and Urban100 test sets. Red



**Table 1** Comparison of the reconstruction effects of DRSR and several state-of-the-art SISR methods on Set5, Set14, B100, Urban100 and DIV2K

Scale	Method	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Average PSNR/SSIM
×2	Bicubic	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403	30.09/0.8705
	Aplus (Timofte et al., 2014)	36.54/0.9544	32.28/0.9056	31.21/0.8863	29.20/0.8938	32.31/0.9100
	SelfExSR (Huang et al., 2015)	36.50/0.9536	32.22/0.9034	31.17/0.8853	29.52/0.8965	32.35/0.9097
	SRCNN (Dong et al., 2014)	36.66/0.9542	32.45/0.9067	31.36/0.8879	29.51/0.8946	32.50/0.9109
	LapSRN(Lai et al. 2017)	37.44/0.9581	32.96/0.9117	31.78/0.8941	30.39/0.9093	33.14/0.9183
	DRCN (Kim et al., 2016b)	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	33.32/0.9195
	DRRN(Wang et al. 1904)	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	33.56/0.9222
	VDSR (Kim et al., 2016a)	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	33.31/0.9203
	MemNet(Tai et al. 2017b)	37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195	33.61/0.9228
	TSCN (Hui et al., 2018)	37.88/0.9602	33.28/0.9147	32.09/0.8985	31.29/0.9198	33.64/0.9233
	EDSR (Lim et al., 2017)	38.11/0.9601	<b>33.92/0.9195</b>	<b>32.32/0.9013</b>	<b>32.93/0.9351</b>	<b>34.32/0.9290</b>
		<b>38.21/0.9612</b>	<b>33.92/0.9204</b>	<b>32.33/0.9016</b>	<b>32.82/0.9345</b>	<b>34.32/0.9294</b>
	DRSR (Ours)	(3.21/0.0024)	(2.34/0.0034)	(3.27/0.0009)	(4.20/0.0031)	(3.28/0.0015)
	×3	Bicubic	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349
Aplus (Timofte et al., 2014)		32.58/0.9088	29.13/0.8188	28.29/0.7835	26.03/0.7973	29.01/0.8271
SelfExSR(Huang et al., 2015)		32.64/0.9097	29.15/0.8196	28.29/0.7840	26.46/0.8090	29.14/0.8306
SRCNN (Dong et al., 2014)		32.75/0.9090	29.29/0.8215	28.41/0.7863	26.24/0.7991	29.17/0.8290
LapSRN (Lai et al., 2017)		—/—	—/—	—/—	—/—	—/—
ESPCN (Shi et al., 2016)		33.13/—	29.45/—	28.54/—	—/—	—/—
DRCN (Kim et al., 2016b)		33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	29.88/0.8444
DRRN (Wang et al., 1904)		34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	30.12/0.8494
VDSR(Kim et al. 2016a)		33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	29.85/0.8446
MemNet (Tai et al., 2017b)		34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376	30.15/0.8494
TSCN (Hui et al., 2018)		34.18/0.9256	29.99/0.8351	28.95/0.8012	27.46/0.8362	30.15/0.8495
EDSR (Lim et al. 2017)		34.65/0.9282	30.52/0.8462	<b>29.25/0.8093</b>	<b>28.80/0.8653</b>	<b>30.81/0.8623</b>
DRSR (Ours)		<b>34.68/0.9290</b>	<b>30.55/0.8464</b>	<b>29.25/0.8088</b>	<b>28.72/0.8636</b>	<b>30.80/0.8620</b>
		(8.21/0.0027)	(7.83/0.0053)	(5.28/0.0006)	(6.91/0.0037)	(2.37/0.0064)

**Table 1** (continued)

Scale	Method	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Average PSNR/SSIM
×4	Bicubic	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	25.88/0.7096
	Aplus (Timofte et al., 2014)	30.28/0.8603	27.32/0.7491	26.82/0.7087	24.32/0.7183	27.19/0.7591
	SelfExSR(Huang et al. 2015)	30.30/0.8620	27.38/0.7516	26.84/0.7106	24.80/0.7377	27.33/0.7655
	SRCNN (Dong et al., 2014)	30.48/0.8628	27.50/0.7513	26.90/0.7103	24.52/0.7226	27.35/0.7618
	LapSRN (Lai et al., 2017)	31.52/0.8854	28.08/0.7687	27.31/0.7255	25.21/0.7545	28.03/0.7835
	ESPCN (Shi et al., 2016)	30.90/–	27.73/–	27.06/–	–/–	–/–
	DRCN (Kim et al., 2016b)	31.53/0.8854	28.03/0.7673	27.24/0.7233	25.14/0.7511	27.99/0.7818
	DRRN (Wang et al., 1904)	31.68/0.8888	28.21/0.7721	27.38/0.7284	25.44/0.7638	28.18/0.7883
	VDSR (Kim et al., 2016a)	31.35/0.8838	28.02/0.7678	27.29/0.7252	25.18/0.7525	27.96/0.7823
	MemNet (Tai et al., 2017b)	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630	28.23/0.7882
	TSCN(Hui et al. 2018)	31.82/0.8907	28.28/0.7734	27.42/0.7301	25.44/0.7644	28.24/0.7897
	EDSR (Lim et al., 2017)	<b>32.46/0.8968</b>	28.80/0.7876	27.71/0.7420	26.64/0.8033	28.90/0.8074
	DRSR (Ours)	<b>32.43/0.8984</b> (6.25/0.0019)	<b>28.86/0.7879</b> (4.52/0.0006)	<b>27.74/0.7415</b> (5.33/0.0054)	<b>26.69/0.8039</b> (6.25/0.0036)	<b>28.93/0.8079</b> (3.27/0.0080)

Bold and italic indicate the best and second-best performance. The data in brackets is the standard deviation

**Table 2** Comparison of the IFC in the reconstruction effects of the algorithm and advanced SR network on Set5, Set14, B100, Urban100 and DIV2K

Dataset	Scale	Bicubic	SRCNN (Dong et al., 2014)	VDSR (Kim et al., 2016a)	DRCN (Kim et al., 2016b)	LapSRN (Lai et al., 2017)	EDSR (Lim et al., 2017)	DRSR
Set5	×2	6.083	8.081	8.580	8.367	8.401	<i>8.903</i>	<b>8.905</b>
	×3	3.580	4.658	5.203	5.180	5.178	<i>5.529</i>	<b>5.540</b>
	×4	2.329	3.017	3.542	3.532	3.515	<i>3.952</i>	<b>3.971</b>
Set14	×2	6.105	7.795	8.159	8.082	8.042	<i>8.544</i>	<b>8.552</b>
	×3	3.473	4.337	4.691	4.699	4.674	<i>4.996</i>	<b>5.014</b>
	×4	2.237	2.758	3.106	3.094	3.089	<i>3.434</i>	<b>3.475</b>
B100	×2	5.619	7.175	7.494	7.236	7.295	<i>7.669</i>	<b>7.675</b>
	×3	3.138	3.843	4.141	4.071	4.046	<b>4.286</b>	4.284
	×4	1.978	2.401	2.679	2.583	2.618	<i>2.825</i>	<b>2.846</b>
Urban100	×2	6.230	7.983	8.629	8.461	8.441	<b>9.628</b>	9.607
	×3	3.522	4.578	5.159	5.153	5.112	<i>5.976</i>	<b>5.981</b>
	×4	2.283	2.972	3.462	3.409	3.448	<i>4.213</i>	<b>4.267</b>

Bold and italic indicate the best and second-best performance

and blue indicate the best and second-best performance, respectively. It can be seen from the table that DRSR achieves better performance compared with other algorithms, which proves that the image information restored by DRSR is more accurate than that by other networks.

Table 3 is a detailed comparison of DRSR and EDSR (Lim et al., 2017) with magnification factor  $\times 4$ . In training, the training set of DRSR is more than that of EDSR. Also, the depth of DRSR is relatively deeper. From the table, the PSNR value and IFC of DRSR on the test set are higher than those of EDSR.

Table 4 is a detailed ablation study with magnification factor  $\times 4$ . All networks in the table have 64 channels. The models of EDSR and EDSR + skip have 16 residual blocks. There are 32 dirac residual blocks in DRSR. From the table, DRSR has fewer parameters and better results. When  $\alpha = 1$ ,  $\beta = 1$ , the model has the best performance. However, when the network deepens, it is difficult to converge. In this paper, we set  $\alpha = 1$ ,  $\beta = 0.1$ .

Figures 4 and 5 show the loss convergence curve and PSNR value change curve of the model during the training process. As can be seen from the figure, our model works best when  $\alpha = 1$ ,  $\beta = 1$ . In addition, our model is basically superior to EDSR regardless of the value of  $\alpha$  and  $\beta$ . It is because that compared with ordinary residual network, Dirac residual network can train deeper network model and enhance the ability of feature extraction by adaptively selecting parameters to control the weights of convolution operation and skip operation. In EDSR network, ordinary residual network is used to design feature extraction network, while DRSR network uses Dirac residual network to design feature extraction network. Therefore, we can conclude that DRSR has better performance than EDSR because Dirac residual network has better feature expression ability.

Figures 6, 7, 8, 9, and 10 show the visual comparisons of the reconstruction effects of DRSR and other state-of-the-art networks with magnification factors  $\times 2$ ,  $\times 3$ , and  $\times 4$  on Set2, Set14, B100, and Urban100. Figure 7 shows the model reconstruction effects with the magnification factor  $\times 3$  of “Barbara” in Set14. It can be seen that the HR image reconstructed by DRSR has been dramatically improved visually compared

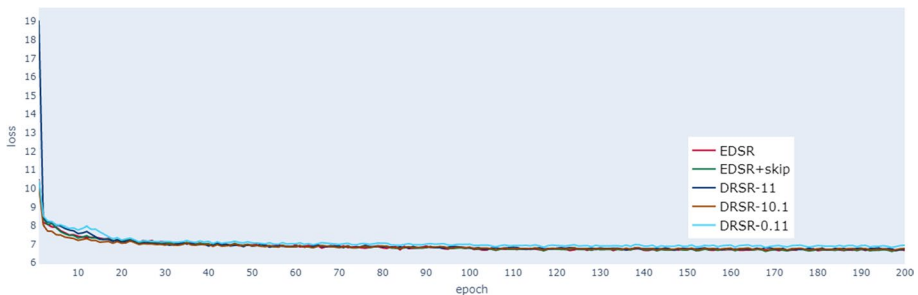
**Table 3** Detailed comparison of DRSR and EDSR

Method	EDSR(Lim et al., 2017)	DRSR
Training set	DIV2K (1–800)	DIV2K (1–800)
Convolution Layer Number	71	87
Number of convolution channels	256	256
Loss function	L1	L1
Number of iterations	$6 \times 10^6$	$6 \times 10^6$
Set5 $\times$ 4 (PSNR/IFC)	<b>32.46/3.952</b>	32.43/3.971
Set14 $\times$ 4 (PSNR/IFC)	28.80/3.434	<b>28.86/3.475</b>
B100 $\times$ 4 (PSNR/IFC)	27.71/2.825	<b>27.74/2.846</b>
Urban100 $\times$ 4 (PSNR/IFC)	26.64/4.213	<b>26.69/4.267</b>

Bold indicates the best performance

**Table 4** Comparison of parameters and ablation study (the magnification scale is  $\times 4$ )

Method	Params	$\alpha$	$\beta$	Set5	Set14	BSD100	Urban100
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
EDSR	1.45 M	$\times$	$\times$	31.97/0.8915	28.46/0.7787	27.49/0.7730	25.72/0.7743
EDSR + skip	2.23 M	$\times$	$\times$	31.98/0.8921	28.48/0.7794	27.50/0.7333	25.75/0.7753
DRSR-1 1	1.11 M	1	1	31.94/0.8924	28.50/0.7790	27.50/0.7331	25.91/0.7794
DRSR-1 0.1	1.11 M	1	0.1	31.96/0.8921	28.45/0.7782	27.47/0.7323	25.73/0.7743
DRSR-0.1 1	1.11 M	0.1	1	31.50/0.8856	28.28/0.7740	27.37/0.7280	25.54/0.7659

**Fig. 4** The loss convergence curve of our model DRSR

to that of other advanced algorithms. The texture reconstructed by this algorithm is more real and accurate. In contrast, other algorithms have more or less reconstructed the wrong texture. In Figs. 6, 8, and 9, the reconstruction effect of DRSR is similar to that of EDSR and is much better than that of other algorithms. The reconstruction HR image details of DRSR are entirely accurate. For the Urban100 dataset, as shown in Fig. 10, the edges of the camera are sharper and more visible in the HR image reconstructed by DRSR. Other textures are also accurate and precise.

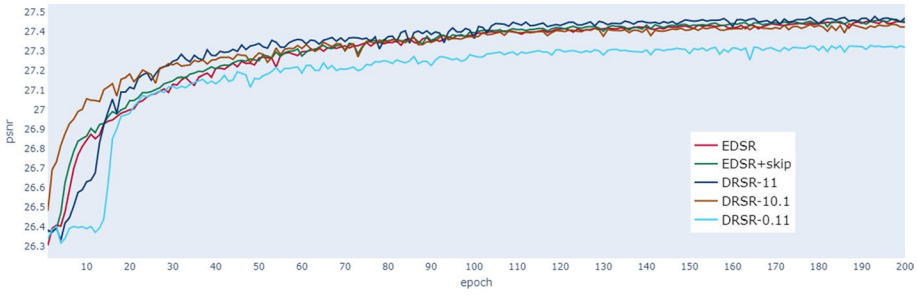


Fig. 5 The PSNR convergence curve of our model DRSR

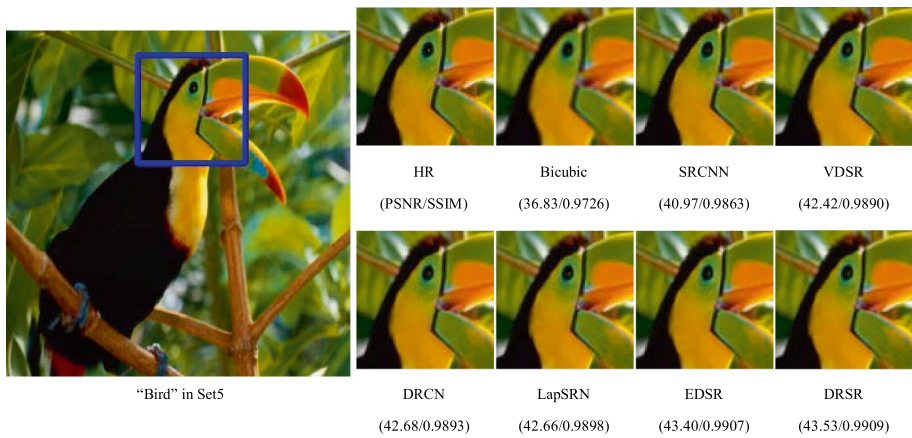
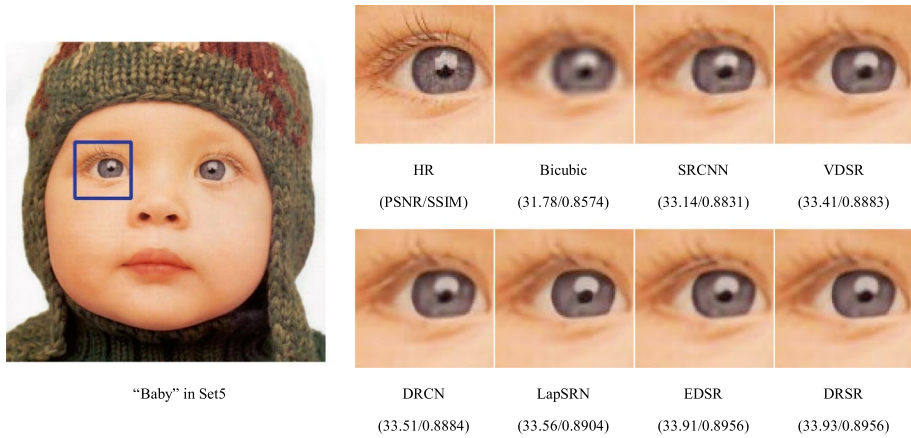


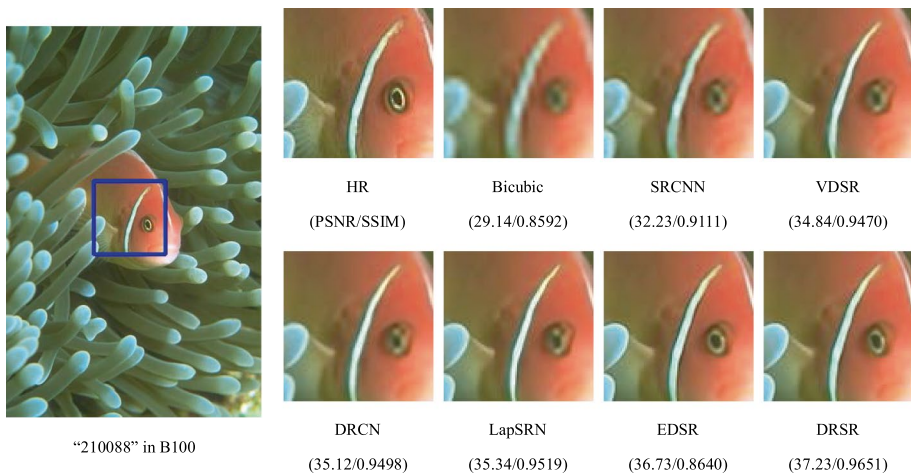
Fig. 6 Qualitative comparisons of DRSR and other models with scale  $\times 2$  using image “bird” on Set5



Fig. 7 Qualitative comparisons of DRSR and other models with scale  $\times 3$  using image “Barbara” on Set14



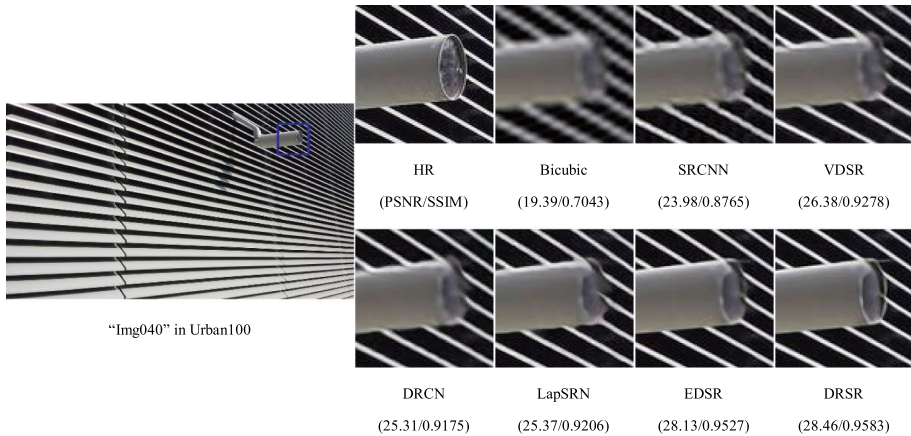
**Fig. 8** Qualitative comparisons of DRSR and other models with scale  $\times 4$  using image “Baby” on Set5



**Fig. 9** Qualitative comparisons of DRSR and other models with scale  $\times 4$  using image “210,088” on B100

## 4 Conclusion

In this paper, an SR reconstruction algorithm based on the original Dirac residual is proposed for the SISR task. The network learns the high-frequency features of the input LR image through the Dirac residual layer, uses the global skip connection to utilize the low-frequency features directly, and reconstructs the image through the sub-pixel convolution layer. In addition, the residual layer of the traditional SR algorithm is improved by weight parameterization. Finally, the reconstruction results of the input image and the RESNET learned feature are combined as the final reconstruction result. The network can adaptively adjust the proportion of the convolution feature and the skip connection in each level of the residual block output.



**Fig. 10** Qualitative comparisons of DRSR and other models with scale  $\times 4$  using image “Img040” on Urban100

Our model does not only add hyperparameters to the branches of the two networks, but we also design a new SR network reconstruction model. This network is derived from ResNet, and it is also an attempt to non-skip connection to find another way to implement residuals. Experiments show that the algorithm has achieved excellent results in both objective performance indexes such as PSNR, SSIM, IFC, and subjective visual perception.

**Acknowledgements** This research was supported by the National Natural Science Foundation of China (61573182), and by the Fundamental Research Funds for the Central Universities (NS2020025).

## References

- Cao, Y., He, Z., Ye, Z., et al. (2019). Fast and accurate single image super-resolution via an energy-aware improved deep residual network [J]. *Signal Processing*, 162, 115–125.
- Chang, H., Yeung, D. Y., & Xiong, Y. (2004). Super-resolution through neighbor embedding[C]. In: Proceedings of the 2004 In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. IEEE, vol. 1, pp. I–I.
- Dai, T., Cai, J., Zhang, Y., et al. (2019) Second-order attention network for single image super-resolution[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 11065–11074.
- Dong, C., Loy, CC., He, K., et al. (2014). Learning a deep convolutional network for image super-resolution[C]. In: European Conference on Computer Vision. Springer, Cham pp. 184–199.
- Dong, C., Loy, C. C., & Tang X. (2016). Accelerating the super-resolution convolutional neural network[C]. In: European Conference on Computer Vision. Springer, Cham, pp. 391–407.
- Haris, M., Shakhnarovich, G., & Ukita, N. (2019). Deep back-projection networks for single image super-resolution [J]. arXiv preprint [arXiv:1904.05677](https://arxiv.org/abs/1904.05677).
- He, K., Zhang, X., Ren, S., et al. (2016). Deep residual learning for image recognition[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.
- Huang, J. B., Singh, A., & Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. In: IEEE Conference on Computer Vision Pattern Recognition.
- Hui, Z., Wang, X., & Gao X. (2018) Two-stage convolutional network for image super-resolution. In: ICPR, pp. 2670–2675.
- Kim, J., Kwon, L. J., & Mu L. K. (2016). Accurate image super-resolution using very deep convolutional networks[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1646–1654.
- Kim, J., Kwon, L.J., & Mu L. K. (2016). Deeply-recursive convolutional network for image super-resolution[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1637–1645.

- Lai, W. S., Huang, J. B., Ahuja, N., et al. (2017). Deep laplacian pyramid networks for fast and accurate super-resolution[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 624–632.
- Ledig, C., Theis, L., Huszár, F., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4681–4690.
- Lim, B., Son, S., Kim, H., et al. (2017). Enhanced deep residual networks for single image super-resolution[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 136–144.
- Schulter, S., Leistner, C., & Bischof, H. (2015). Fast and accurate image upscaling with super-resolution forests[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3791–3799.
- Sheikh, H. R., Bovik, A. C., & De Veciana, G. (2005). An information fidelity criterion for image quality assessment using natural scene statistics [J]. *IEEE Transactions on Image Processing*, 14(12), 2117–2128.
- Shi, W., Caballero, J., Huszár, F., et al. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1874–1883.
- Shi, W., Caballero, J., Ledig, C., et al. (2013). Cardiac image super-resolution with global correspondence using multi-atlas patchmatch[C]. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Berlin, Heidelberg, pp. 9–16.
- Tai, Y., Yang, J., & Liu, X. (2017). Image super-resolution via deep recursive residual network. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3147–3155
- Tai, Y., Yang, J., Liu, X., et al. (2017). Memnet: A persistent memory network for image restoration [C]. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4539–4547.
- Timofte, R., De Smet, V., & Van Gool, L. (2013). Anchored neighborhood regression for fast example-based super-resolution[C]. In: Proceedings of the IEEE international conference on computer vision. pp. 1920–1927.
- Timofte, R., Smet, V. D., & Gool, L. V. (2014). A+: adjusted anchored neighborhood regression for fast super-resolution. In: Asian Conference on Computer Vision.
- Tong, T., Li, G., Liu, X., et al. (2017). Image super-resolution using dense skip connections[C]. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4799–4807.
- Wang, C., Li, Z., & Shi, J. (2019). Lightweight image super-resolution with adaptive weighted learning network[J]. arXiv preprint [arXiv:1904.02358](https://arxiv.org/abs/1904.02358).
- Xu, X., Ma, Y., & Sun, W. (2019). Towards real scene super-resolution with raw images[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1723–1731.
- Yang, J., Wright, J., Huang, T. S., et al. (2010). Image super-resolution via sparse representation[J]. *IEEE Transactions on Image Processing*, 19(11), 2861–2873.
- Yang, S., Sun, F., Wang, M., et al. (2011). Novel super resolution restoration of remote sensing images based on compressive sensing and example patches-aided dictionary learning[C]. In: 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping. IEEE, pp. 1–6.
- Yu, J., Fan, Y., Yang, J., et al. (2018). Wide activation for efficient and accurate image super-resolution[J]. arXiv preprint [arXiv:1808.08718](https://arxiv.org/abs/1808.08718).
- Zagoruyko, S., & Komodakis N. (2017). Diraconets: Training very deep neural networks without skip-connections [J]. arXiv preprint [arXiv:1706.00388](https://arxiv.org/abs/1706.00388).
- Zhang, K., Zuo, W., & Zhang, L. (2019). Deep plug-and-play super-resolution for arbitrary blur kernels[J]. arXiv preprint [arXiv:1903.12529](https://arxiv.org/abs/1903.12529).
- Zhang, X., Chen, Q., Ng, R., et al. (2019). Zoom to learn, learn to zoom[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3762–3770.
- Zhang, Y., Li, K., Li, K., et al. (2018). Image super-resolution using very deep residual channel attention networks[C]. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 286–301.
- Zhang, Y., Tian, Y., Kong, Y., et al. (2018). Residual dense network for image super-resolution[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2472–2481.
- Zhen, Li., Jinglei Y., & Zheng L. (2019). Feedback network for image super-resolution [J]. arXiv preprint [arXiv:1903.09814](https://arxiv.org/abs/1903.09814).
- Zou, W. W. W., & Yuen, P. C. (2011). Very low resolution face recognition problem [J]. *IEEE Transactions on Image Processing*, 21(1), 327–340.





**Xin Yang** was born in 1978. He is an associate professor and master supervisor at Nanjing University of Aeronautics and Astronautics, China. He received his master and PhD degrees all from Southeast University in 2005 and 2007, respectively. His current research interests include digital image processing, pattern recognition, and etc.



**Tangxin Xie** was born in 1995. He is a postgraduate at Nanjing University of Aeronautics and Astronautics, China. He received his Master's degree from School of Automation, Nanjing University of Aeronautics and Astronautics. His current research interests include image processing, pattern recognition, and etc.



**Li Liu** was born in 1997. He is a postgraduate at Nanjing University of Aeronautics and Astronautics, China. He received his Bachelor's degree from School of information Science and Engineering, Ningbo University. His current research interests include image processing, pattern recognition, and etc.



**Dake Zhou** was born in 1974. He is an Associate Professor and master supervisor at Nanjing University of Aeronautics and Astronautics, China. He received his PhD degree in Pattern recognition and intelligent system from Shanghai Jiaotong University in 2005. His current research interests include image processing, pattern recognition, and computer vision.