



MFFN: image super-resolution via multi-level features fusion network

Yuantao Chen¹ · Runlong Xia^{2,3} · Kai Yang⁴ · Ke Zou⁵

Accepted: 2 February 2023 / Published online: 15 February 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Deep convolutional neural networks can effectively improve the performance of single-image super-resolution reconstruction. Deep networks tend to achieve better performance than others. However, the deep CNNs will lead to a dramatic increase in the size of parameters, limiting its application on embedding and resource-constrained devices, such as smart phone. To address the common problems of blurred image edges, inflexible convolution kernel size selection and slow convergence during training procedure due to redundant network structure in image super-resolution algorithms, this paper proposes a lightweight single-image super-resolution network that fuses multi-level features. The components are mainly two-level nested residual blocks. To better extract features and reduce the number of parameters, each residual block adopts an asymmetric structure. Firstly, it expands twice and then compresses the number of channels twice. Secondly, in the residual block, the feature information of different channels is weighted and fused by adding an autocorrelation weight unit. The quality of the reconstructed image of the proposed method is superior to the existing image super-resolution reconstruction methods in both subjective perception and objective evaluation indicators, and the reconstruction performance is better when the factor is large.

Keywords Residual learning · Multi-level features · Super-resolution · Convolutional neural network · Lightweight

1 Introduction

Image super-resolution methods are the research focus of computer vision tasks and have received much attention for

many years. Given a low-resolution image, super-resolution (SR) techniques aim to recover its corresponding high-resolution (HR) image [1]. Since one low-resolution (LR) image may correspond to multiple high-resolution images, the super-resolution problem is an ill-conditioned inverse problem [2][3]. Therefore, how to efficiently restore the lost texture details in the process of super-resolution reconstruction, maintain the integrity of the image structure, and effectively suppress the generation of distortion is a challenging problem [4][5]. Super-resolution imaging is one hot topic in the field of computer vision, and the technology of deep learning dominates the current research on single-image super-resolution (SISR) methods [6].

The deep learning-based SISR method directly learns the end-to-end mapping relationship between LR images and HR images. Dong et al. [7] had proposed the SISR method based on the convolutional neural network [8], called super-resolution convolutional neural network (SRCNN). SRCNN used three convolutional layers to directly learn nonlinear mappings of LR images and HR images in an end-to-end model. Kim et al. [9] had proposed a deeper network named very deep convolutional networks (VDSR) based on residual learning [10], which effectively improved

✉ Yuantao Chen
chenyt@hnuit.edu.cn

Runlong Xia
xiarunlong@vip.qq.com

Kai Yang
yangkai@zoomlion.com

Ke Zou
zouk@hnwjgk.cn

¹ School of Computer Science and Engineering, Hunan University of Information Technology, Changsha, Hunan, China

² Mountain Yuelu Breeding Innovation Center Limited, Changsha, China

³ Hunan Provincial Science and Technology Affairs Center, Changsha, Hunan, China

⁴ Hunan ZOOMLION Intelligent Technology Corporation Limited, Changsha, Hunan, China

⁵ Hunan WUJO High-Tech Material Corporation Limited, Loudi, China

the performance. To increase the network depth and limit the increase in network parameters, Kim et al. [11] adopted the recursive structure of shared parameters and proposed the deeply recursive convolutional network (DRCN). Tai et al. [12] had proposed deep recursive residual network (DRRN), which simultaneously utilized local residual structure, global residual structure, and recursive structure. The residual units are shared and a small number of parameters are added, which improves the performance of VDSR and DRCN. The super-resolution feedback network (SRFBN-S) [13] had used a recurrent neural network structure to share hidden layer parameters, reducing the number of parameters and improving the performance of reconstructed images. Hui et al. [14] had proposed the information multi-distillation network (IMDN) method to gradually extract feature information within the residual block and use the channel attention mechanism for feature selection, which further improves the quality of the reconstructed image. Ahn et al. [15] had proposed a SISR method based on a cascaded residual network, called cascading residual network (CARN), which combined cascaded structure and residual learning and achieved a better balance between the number of parameters and performance. Zhu et al. [16] had proposed the compact back-projection network (CBPN) method, which enhanced the reconstruction ability by extracting feature information in LR and HR spaces by cascading up/down sampling layers. The multi-scale residual network (MSRN) [17] had used convolutional layers with different receptive fields in the residual block to extract feature information of different scales to further improve the performance. Lai et al. [18] had proposed the Laplacian pyramid network structure of the SISR method Laplacian pyramid super-resolution network (LapSRN), which gradually upsamples and predicts residuals, which can simultaneously complete HR image reconstruction of multiple sizes.

Most of the existing algorithms achieve the goal of improving network performance by deepening or widening the network. However, the large size of the network model will bring two problems: (1) a large-sized model will consume too much storage space, which is not conducive to deployment in practical applications; (2) a large-sized model will introduce a heavy computational burden. It is not suitable for applications with limited computing power or high real-time requirements. Therefore, the trade-off between model size and SR reconstruction performance needs to be specially considered when designing the network.

LapSRN uses a progressive method to increase the image resolution, which will generate large-sized feature maps in the middle of the network, resulting in a significant increase in network computation. MFRN introduces a recursive learning strategy, although a larger receptive field can be obtained with a small number of parameters, the computational burden is high. CARN fuses information from multiple levels

by applying a cascade mechanism but also introduces more parameters and calculations due to the use of dense connections. In CARN's mobile version model named CARN-M, group convolution is used to reduce the number of network parameters, but the use of group convolution will significantly reduce network performance. IDN uses an information distillation mechanism to transmit some features through skip connections to reduce the number of network parameters, but it cannot effectively screen out important features that need to be further refined, and there is still room for further improvement in model performance.

The above methods use a lightweight network, however, the network depth and the number of parameters is important factors that affect the performance of SISR. Lim et al. [19] had proposed a heavyweight EDSR method that removed the normalization module, superimposed residual blocks, and more than 65 convolutional layers [20]. The MM-RealSR method [38] combined residual structure and dense structure and made full use of the hierarchical feature information of LR images and can recover high-quality HR images. Liu et al. [22] had proposed the RFANet method, which uses a spatial attention module with a larger receptive field and a smaller number of parameters in the residual block to filter feature information, and then analyze the features extracted by the residual branch of each residual block. Fusion improves the quality of image reconstruction. EDSR, MM-RealSR, and RFANet methods are currently the most representative methods that use heavyweight network SISR, which have good performance and large parameters. In the case of limited resources, it is difficult for the heavyweight SISR model to meet the application requirements. This paper considers the lightweight SISR model to provide an effective solution.

This paper proposes a SISR method for lightweight multi-level feature fusion networks. When the magnification factor is 4 times, the amount of multi-level features fusion network (MFFN) parameters is only 1.47 M, which is 1/29 of EDSR, 1/14 of MM-RealSR, and 1/7 of RFANet. Compared with similar lightweight SISR models, our method achieves a better balance in performance and model scale. Taking the MSRN method as an example, the parameters of the proposed method are reduced by 3/4. On the test dataset, $\times 2$, $\times 3$, and $\times 4$ are enlarged, and the objective performance is comparable, while $\times 8$ is enlarged, which is consistently better than the MSRN method. For $\times 4$ and $\times 8$ upscaling, the subjective performance is also consistently better than the MSRN method. The experimental results can show that the ability of the proposed method to reconstruct the fringes is significantly better than other lightweight methods, and for the 8 times large-scale magnification factor, the reconstructed image results have more obvious advantages.

The contributions of this paper include: (1) We propose the dual residual block (DRB) with asymmetric structure. The residual block is first expanded twice, then compressed

twice, and the two layers of residual connection are used to effectively extract feature information; (2) An autocorrelation weight unit (ACW) has proposed, which can calculate the weight according to the feature information, adaptively weighting different feature channels to effectively transfer feature information; (3) Designing the shallow feature mapping unit (SFMU), SFMU extracts different levels of shallow features through the convolution layers of different receptive fields on each branch; (4) Designing the multi-path reconstruction unit (MPRU), MPRU can obtain the feature information of multiple branches, to fully utilize the feature information of different levels to reconstruct different aspects of the image.

2 Proposed algorithm

Currently, most super-resolution models utilize residual networks [19, 23]. The residual block structure generally adopts a Conv-*ReLU*-Conv layer. The problem is that the model performance is heavily dependent on the network size, which is mainly based on the number of trainable network layers and channels. How to reduce the network size while improving or not reducing the model performance is a very challenging problem. In this paper, a lightweight multi-level feature fusion network is designed. The feature channel is first expanded and then compressed with two-layer nested residual blocks, which can significantly reduce the number of parameters, and the autocorrelation weight unit can adaptively fuse feature information, which also improves the feature utilization effect. The network structure of this paper is shown in Fig. 1a, which mainly includes four parts: shallow feature extraction unit (SFEU), shallow feature mapping unit (SFMU), deep feature mapping unit (DFMU), and multi-path reconstruction unit (MPRU).

Let I_{LR} and I_{SR} be the input and output images, the shallow feature extraction unit only contains a 3×3 convolution layer, which can realize the transformation function of extracting shallow feature information and feature dimension.

$$F_0 = H_{SFEU}(I_{LR}) \quad (1)$$

H_{SFEU} is the shallow feature extraction unit, which generates the shallow feature information F_0 that meets the dimension requirements of the shallow/deep feature mapping unit from the input image I_{LR} . The shallow feature mapping unit further extracts shallow feature information from F_0 and passes the shallow feature information to the multi-path reconstruction unit.

$$(F_{S_1}, F_{S_2}, F_{S_3}) = H_{SFMU}(F_0) \quad (2)$$

$(F_{S_1}, F_{S_2}, F_{S_3})$ is the shallow feature information extracted by the shallow feature mapping unit H_{SFMU} . The deep feature mapping unit also extracts deep feature information from F_0 .

$$(F_{D_1}, F_{D_2}, F_{D_3}) = H_{DFMU}(F_0) \quad (3)$$

$(F_{D_1}, F_{D_2}, F_{D_3})$ is the deep layer feature information generated by the deep feature mapping unit H_{DFMU} , and then the multi-path reconstruction unit receives the shallow layer feature information and the deep layer feature information to reconstruct the final result I_{SR} .

$$I_{SR} = H_{MPRU}(F_{S_1}, F_{S_2}, F_{S_3}, F_{D_1}, F_{D_2}, F_{D_3}) \quad (4)$$

H_{MPRU} is the multi-path reconstruction unit, which reconstructs the image with all the feature information to generate the final result I_{SR} .

A. Shallow Features Mapping Unit (SFMU)

The SISR models usually used one 3×3 convolution to extract shallow features. The shallow features mapping unit (SFMU) module used convolution kernels with different receptive fields to extract shallow feature information hierarchically with multi-scale and multi-level shallow feature information. The rich shallow feature information can help the reconstruction module to reconstruct higher-quality SR images.

The SFMU module first transforms the input information through one 1×1 convolution to reduce the number of parameters in subsequent operations. Then, three convolutional layers with different receptive fields are used to achieve multi-scale shallow feature information extraction. The three convolution layers are divided into three branches, and the shallow feature information is gradually weighted and superimposed. It realizes the extraction of multi-scale and multi-level shallow feature information.

The three branches of the shallow feature mapping unit extract different levels of shallow feature information $(F_{S_1}, F_{S_2}, F_{S_3})$, respectively.

$$F_{S_0} = H_T(F_0) \quad (5)$$

$$F_{S_1} = H_{C_1}(F_{S_0}) \quad (6)$$

$$F_{S_2} = H_{C_2}(F_{S_0} + \alpha_1 F_{S_1}) \quad (7)$$

$$F_{S_3} = H_{C_3}(F_{S_0} + \alpha_2 F_{S_2}) \quad (8)$$

H_T is the 1×1 convolutional layer used to adjust the number of feature information channels, H_{C_1} , H_{C_2} , H_{C_3} are

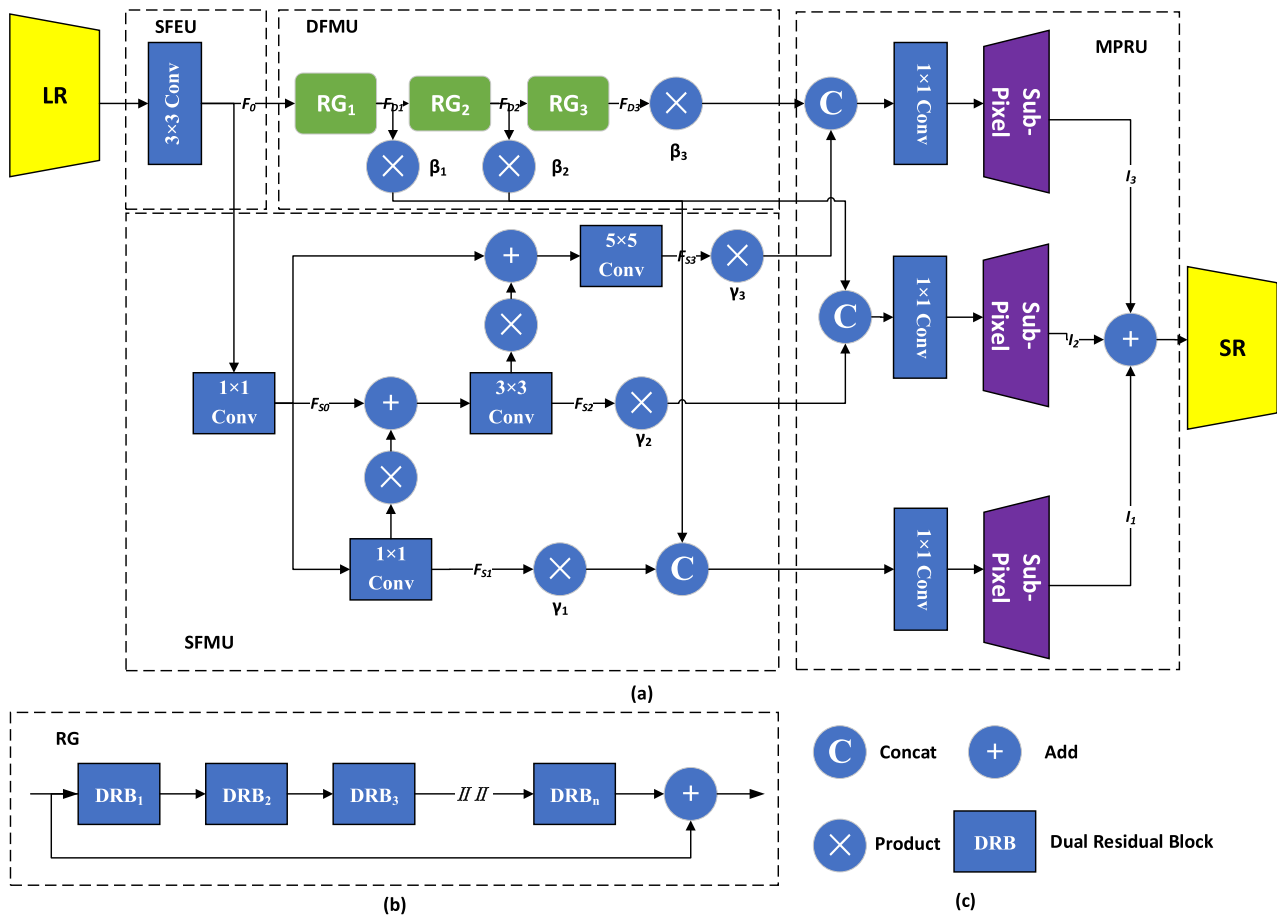


Fig. 1 a Architecture of multi-level features fusion network. b Structure of the residual group. c Symbol description

the convolutional layers on each branch, and α_1, α_2 are the adaptive weights that can be learned.

B. Deep Features Mapping Unit (DFMU)

To obtain deep feature information, this paper designs a DFMU module, as shown in Fig. 1(a). The deep features mapping unit (DFMU) module contains three residual groups (RG), and each RG contains multiple DRB modules. Because the simple stacking residual block method is not conducive to the transfer of feature information, this paper adds local skip connections in each RG to promote the effective transfer of feature information and obtain deep-level feature information through RG.

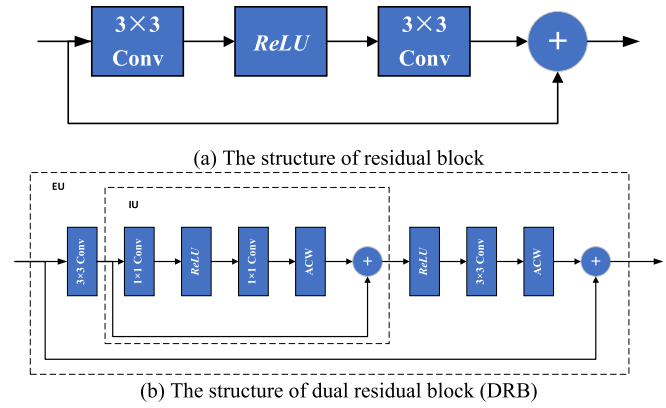
$$F_{D_i} = H_{RG_i}(F_{D_{i-1}}) \tag{9}$$

The i^{th} residual group H_{RG_i} takes the hierarchical feature information $F_{D_{i-1}}$ generated by the $(i - 1)^{th}$ residual group as input and generates hierarchical feature information F_{D_i} .

2.1 Dual residual block (DRB)

The usually used residual structure is shown in Fig. 2a, and each convolutional layer has the same number of channels. A major problem with this structure is that increasing the number of feature channels leads to a rapid increase in the number of parameters. This paper proposes a DRB module as shown in Fig. 2b, which consists of inner unit (IU) and external unit (EU), and uses a dilation-then-compression strategy [24]. This strategy can reduce the number of channels and thus reduce the number of parameters. The first convolutional layer of EU extracts feature information and expands the feature channel to obtain richer image feature information. The second convolutional layer compresses feature channels, filters feature information, and promotes effective feature information transfer. In this paper, an IU containing two 1×1 convolutions was added inside the EU to increase the number of channels without causing a sharp increase in the number of parameters.

Fig. 2 Structure of the different residual blocks



Assume that the input and output feature information of the i^{th} DRB module is and F_i , respectively.

$$F_i = F_{ACW}(W_{EU_2\sigma}(H_{IU}(W_{EU_1}F_{i-1}))) + F_{i-1} \quad (10)$$

The kernel size of the two convolutional layers in EU is 3×3 , and the weights are W_{EU_1} and W_{EU_2} , respectively (ignoring the bias term). H_{IU} is the inner unit operation, $\sigma(\cdot)$ is the **ReLU** activation function, F_{ACW} is the autocorrelation weight unit, and the first convolutional layer with weight W_{EU_1} processes F_{i-1} to generate feature information F_{input} , which is input to the inner unit for processing. After the inner unit is processed, the feature information F_{output} is generated. After processing through the convolution layer, the autocorrelation weight unit, and the skip connection, the output result of the EU is the feature information F_i extracted by the i^{th} DRB module.

IU adopts the structure of Conv-**ReLU**-Conv and adds skip connections.

$$F_{output} = F_{ACW}(W_{IU_2\sigma}(W_{IU_1}F_{input})) + F_{input} \quad (11)$$

Among them, W_{IU_1} and W_{IU_2} are the weights of the two convolutional layers of the inner unit (ignoring the bias term), and $\sigma(\cdot)$ is the **ReLU** activation function. The input of the inner unit comes from the output F_{input} of the first convolutional layer of the outer unit, and after processing through the convolutional layer, the autocorrelation weight unit and the skip connection, the extracted feature information bit F_{output} .

2.2 Autocorrelation weight unit (ACW)

The SISR model based on the deep residual structure still has the problem of vanishing or exploding gradient. To stabilize the training, the residual scale parameter is usually introduced [25]. This hyperparameter is usually set empirically and is difficult to optimize. The ACW module adopts the learnable optimal residual scale parameter.

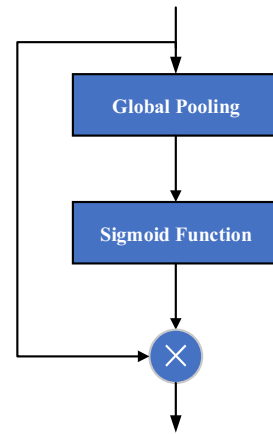


Fig. 3 Structure of autocorrelation weight unit (ACW)

The structure of ACW module is shown in Fig. 3. The ACW module consists of two parts, the global pooling layer, and the sigmoid function, with no additional parameters. The global pooling layer encodes all input feature information as initialized weights, which are then adjusted to $[0, 1]$ using the sigmoid function. Due to the difference between the feature information, different weights are generated, which enhances the feature information effective for the reconstructed image.

Let $X=[x_1, x_2, \dots, x_C]$ be the input feature information and the size are $H \times W \times C$. The initialization weight $Z = [z_1, z_2, \dots, z_C]$ is calculated from the input feature information X through the global average pooling layer H_{GAP} . The initial weights of the c^{th} input feature information are as follows:

$$z_c = H_{GAP}(x_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (12)$$

Use the sigmoid activation function $f(\cdot)$ to adjust the initialization weight Z to generate the final weight parameter W .

$$W = f(Z) \quad (13)$$

Weighting the input feature information:

$$\hat{X} = X \cdot W \quad (14)$$

C. Multi-Path Reconstruction Unit (MPRU)

At the end of the current SISR model network, most of them use transposed convolution or sub-pixel convolution for upsampling operations. Compared with transposed convolution, the image quality of sub-pixel convolution reconstruction is better [26], but it needs to use multiple 3×3 convolution layers [17, 19, 21], the magnification factor increases, and the number of parameters will increase. To reduce the number of parameters without reducing the image quality, the multi-path reconstruction unit (MPRU) module has designed. The MPRU module is shown in Fig. 1a. MPRU has three reconstruction branches, each of which consists of a 1×1 convolutional layer and a sub-pixel convolutional layer. The reconstruction result of each branch is the same size as the HR image, and the final SR image is the sum of the reconstruction results of the three branches. The MPRU module uses a 1×1 convolutional layer, which can greatly reduce the parameters and increase the amplification factor without significantly increasing the number of parameters. At the same time, the MPRU module obtains the feature information of each branch, which can also improve the reconstruction effect.

$F_S = [F_{S_1}, F_{S_2}, F_{S_3}]$ is the shallow feature information extracted by the shallow feature mapping unit, $F_D = [F_{D_1}, F_{D_2}, F_{D_3}]$ is the hierarchical feature information extracted by the deep feature mapping unit, and the i^{th} branch in the multi-path reconstruction unit is generated as follows:

$$I_i = H_{UP_i}(H_{Conv_i}([\gamma_i F_{S_i}, \beta_i F_{D_i}])) \quad (15)$$

H_{UP_i} and H_{Conv_i} represent the sub-pixel convolutional layer and 1×1 convolutional layer of the i^{th} branch, respectively, and γ_i and β_i are the adaptive weights of F_{S_i} and F_{D_i} , respectively. I_i is the reconstructed image of the i^{th} branch. $[\cdot]$ is a concatenate operation on F_{S_i} and F_{D_i} . The MPRU unit has three branches, so $i = [1, 2, 3]$.

$$I_{SR} = I_1 + I_2 + I_3 \quad (16)$$

The multi-path reconstruction unit adds the images (I_1, I_2, I_3) generated by all the branches to generate the final result I_{SR} .

D. Loss Function

In this paper, the L_1 loss function is used to optimize the MFFN. The L_1 loss function is mainly used to calculate the

average absolute value of the difference between each pixel of the input image and the target image, and is a widely used loss function in the field of image super-resolution.

For a given training dataset $\{I_{LR}^i, I_{HR}^i\}_{i=1}^N$ containing N low- and high-resolution image pairs, the objective of the network in this paper is to train the images and to optimize them using the L_1 loss function, axiomatically shown as follows:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \|H_{MFFN}(I_{LR}^i) - I_{HR}^i\|_1, \quad (17)$$

wherein $H_{MFFN}()$ denotes the network reconstruction result, $\|\cdot\|_1$ denotes the L_1 parametrization, and θ denotes the parameters in the network.

3 Experimental analysis and results

E. The Datasets and Metrics

In this paper, we choose DIV2K [20] dataset as the training dataset of the network, which consists of 800 images of the training dataset and 100 images of the validation dataset. To test the effect of the proposed model, we had used five benchmark datasets, namely Set5 [30], Set14 [30], BSD100 [31], Urban100 [32] and Manga109 [33]. Among them, the test dataset BSD100 contains images of various style types, Urban100 is images of various types of buildings, and Manga109 is images of various types of cartoons.

These five test datasets have rich and diverse information, which can well-verify the effectiveness of super-resolution methods. To evaluate the super-resolution performance, this paper uses two commonly used full-reference image quality assessment criteria to evaluate the differences: peak signal-to-noise ratio (PSNR) [34] and structural similarity index (SSIM) [34]. Following the super-resolution convention, the luminance channel is chosen for full-reference image quality assessment because the intensity of an image is more sensitive to human vision than chromaticity.

F. The experimental details

In each training round, this paper cuts the low-resolution RGB images and the corresponding high-resolution RGB images into blocks of 48×48 . The training data are increased by randomly rotating 90 degrees, 180 degrees, 270 degrees, and flipping horizontally. In this paper, the number of dense blocks in the stacked pooled attentional dense blocks is set to, in each pooled attentional dense block, there are three residual dense blocks and three pooled attentional blocks in this paper. The growth rate of the residual dense blocks is 32,

Table 1 Average PSNR and number of parameters of different DRB in residual group on Set5 and DIV2K-10 with scaling factor $\times 4$ in 200 Epochs

Number	Parameter size	Set5	DIV2K-10
5	1.23 M	32.23	29.51
6	1.47 M	32.26	29.55
7	1.71 M	32.25	29.55

the number of channels not specified in the paper is 64, and the final output of the network is 3.

The Adam optimizer [27] can adaptively adjust the learning rate for different parameters, which can effectively improve the convergence speed of the model and thus reach the optimal point faster, so the Adam optimizer is used for the experiments. The initial learning rate of the network is set to 2×10^{-4} , and the learning rate is halved for every 2×10^5 iterations. The proposed method is implemented in a hardware environment with an Intel i9-9900 K (3.6 GHz), 8 GB of RAM, and NVIDIA GeForce RTX 2080Ti GPU. The software environment is 64-bit Ubuntu operating system and *PyTorch* platform.

G. Experiment results and analysis

1. Residual group analysis

The proposed model contains three residual groups, and each group contains same number of double-nested residual blocks. To verify the influence of different numbers of double-nested residual blocks in residual group on the model, we performed four times super-resolution experiments on the Set5 standard test dataset and the DIV2K-10 dataset, with the number of groups in each group being 5, 6, 7 of the residual group for comparative experiments, as shown in Table 1. When the number of comparison groups is 5, although the number of parameters is increased by 0.24 M, the PSNR index is increased by 0.03 dB on the data of Set5 and DIV2K-10, respectively. When the number of groups is 7, the improvement of the PSNR index is not obvious, but the mount of parameters is increased. 0.24 M.

2. SFMU analysis

To verify the influence of the convolution kernel size on the different paths of the shallow feature mapping unit and the non-use of the shallow feature mapping unit on the model, we performed a fourfold superimposition on the Set5 standard test dataset and the DIV2K-10 dataset. Resolution comparison experiment. As shown in Table 2, the shallow feature

Table 2 Average PSNR of different SFMU convolution kernel on Set5 and DIV2K-10 with scaling factor $\times 4$ in 200 epochs

Convolution Kernel	Set5	DIV2K-10
- \times - \times -	32.22	29.52
$1 \times 1 \times 1$	32.18	29.50
$3 \times 3 \times 3$	32.24	29.53
$5 \times 5 \times 5$	32.25	29.53
$1 \times 3 \times 5$	32.26	29.55

Table 3 Influence of residual blocks with different structures on model performance. We report the average PSNR on Set5 and DIV2K-10 with scaling factor $\times 4$ in 200 epochs

Model	Residual block size	Set5	DIV2K-10
I	73.8 K	32.11	29.42
II	53.5 K	32.12	29.47

mapping unit is not used, so the model cannot use the shallow feature information, resulting in a poor reconstruction effect. When each branch uses the same size of the convolution kernel and the size of the convolution kernel increases continuously, the reconstruction effect is improved, but the amount of parameters also increases. When the branch convolution kernels are all 1, the PSNR index is lower than the model that does not use the shallow feature mapping unit, which is due to the less feature information extracted and redundant information, resulting in a poorer effect. We find that the best results are obtained when the three branch convolution kernels are set to 1, 3, and 5, respectively. This is because each branch can extract different levels of shallow feature information, which can be effectively combined with deep-level feature information. Therefore, we use shallow feature mapping units with branch convolution kernel sizes of 1, 3, and 5, respectively.

3. DRB analysis

Compared with the popular residual block structure, as shown in Fig. 2(a), the DRB module in this paper has advantages in performance and parameter quantity. This paper conducts a fourfold super-resolution comparison experiment on the Set5 and DIV2K-10 datasets. We construct two test models, called Model I and Model II, respectively. Model I is an EDSR structure, but the number of input and output channels of the convolutional layer is reduced from 256 to 64, and the number of residual blocks is reduced from 32 to 18. Model II replaces the residual block in the model I with a DRB module. In the DRB expansion stage, the number of EU input and output feature channels is set to 32 and 64,

Table 4 Effects of ACW. We report the average PSNR on Set5 and DIV2K-10 with scaling factor $\times 4$ in 200 Epochs

ACW	Set5	DIV2K-10
\times	32.11	29.42
\checkmark	32.23	29.55

Table 5 Effects of MPRU. We report the average PSNR on Set5 and DIV2K-10 with scaling factor $\times 4$ in 200 Epochs

Reconstruction model	Parameter size	Set5	DIV2K-10
EDSR	297.16 K	32.11	29.42
MPRU	99.36 K	32.23	29.57

respectively, and the number of IU input and output feature channels is set to 64 and 128, respectively; in the compression stage, the number of EU input and output channels is set to 64 and 32, respectively, and the number of IU input and output channels is set to 64 and 32, respectively. The number of feature channels is set to 128 and 64, respectively. The other parameters in Model I and Model II are the same: The residual scale parameter is 0.1, running 200 iterations, the PSNR of Model II is 0.01 dB and 0.05 dB higher than that of Model I, respectively, and the parameter amount of the DRB module is higher than that of the ERSR residual block. 20.3 K down.

4. ACW analysis

To verify the effectiveness of the ACW module, two cases of the model I including and not including the ACW module are considered, and the results are shown in Table 4. After using the ACW module, the PSNR is improved by 0.12 dB and 0.13 dB on Set5 and DIV2K-10 datasets, respectively. The experimental results show that the ACW module is effective in automatically learning the optimal residual scale parameters.

5. MPRU Analysis

To verify the reconstruction performance of MPRU module, after replacing the EDSR reconstruction unit of the model I with the MPRU module, it is compared with the model I. The results of the two models correspond to Table 5, respectively. The parameter quantity of the MPRU module is only 99.36 K, which is only about 1/3 of the parameter size of EDSR reconstruction unit. In the Set5 and DIV2K-10 datasets, the MPRU module improves PSNR by 0.12 dB and 0.15 dB, respectively.

H. Model analysis

The model in this paper verifies $\times 2$, $\times 3$, $\times 4$ and $\times 8$ super-resolution factor on Set14, B100, Urban100, and Manga109 datasets. We had chosen SRCNN [7], FSRCNN [35], VDSR [9], DRCN [11], LapSRN [18], DRRN [12], CBPN [16], IMDN [14], BFDN [42], BSRN [41], Swin-IR [39], MM-RealSR [38], and LDL [36] for performance comparison with MFFN.

The objective indicators are shown in Table 6. In the case of the $\times 2$ factor on the Set14 dataset, the PSNR of the model in this paper is 0.27 dB higher than that of the BFDN model, which is similar in the case of other amplification factors, and the number of parameters of the model in this paper is less than that of BFDN (about 120 K less). On all test datasets, when the amplification factor is four times, the PSNR of our model is 0.06 dB and 0.16 dB higher than that of MM-RealSR model and LDL on average. Although the performance of BSRN is slightly better on some datasets, the comprehensive performance of the model in this paper is better, and the four times the number of parameters of the BSRN model is four times that of the model in this paper.

In order to compare the reconstruction performance of different super-resolution methods in terms of visual quality, Fig. 4 and Fig. 5 show the super-resolution reconstruction results of “Img048” and “Img092” images in Urban100 at $\times 4$ factor, respectively. Figure 6 and Fig. 7 show the super-resolution reconstructions of “223,061” and “253,027” from dataset B100 at $\times 4$ factor, respectively. The ground truth represents the original HR image. In order to highlight the contrast effect, a local area of the image is selected for magnification using the double triple interpolation method. By observing Fig. 5 and Fig. 7, it can be seen that although the MM-RealSR [38] can clearly recover the significant texture information in the image, texture information has obvious orientation problems, while Swin-IR [39] and BSRN [41] can recover correct texture information to some extent, but it is difficult to suppress the wrong texture, and the textures of these two methods are more blurred.

In contrast, the method in this paper is able to produce correctly oriented textures and sharper edges on locally zoomed-in regions of the graph, and is more consistent with human vision. This is due to the strong feature alignment capability of the deformable convolution in SFMU module, which enables the network model to recover more correctly the complex texture structures in different images. It can be clearly observed in most of local zooms that the details of the images reconstructed by other methods are blurred, and even the edge information of the images cannot be reconstructed, while the details reconstructed by the method in this paper are much clearer and have better recognition. These results in Fig. 6 also indicate that the proposed method achieves better results in terms of subjective performance.

Table 6 Average PSNR/SSIM of various SISR methods

Scale	Model	Parameter size	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM	
× 2	SRCNN [7]	57 K	32.42/0.9063	31.36/0.8879	29.50/0.8946	35.74/0.9661	
	FSRCNN [35]	12 K	32.63/0.9088	31.53/0.8920	29.88/0.9020	36.67/0.9694	
	VDSR [9]	665 K	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9729	
	DRCN [11]	1774 K	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.63/0.9723	
	LapSRN [18]	813 K	33.08/0.9130	31.80/0.8950	30.41/0.9100	37.27/0.9740	
	DRRN [12]	297 K	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.92/0.9760	
	CBPN [16]	677 K	33.28/0.9142	32.08/0.8978	31.31/0.9195	37.72/0.9740	
	IMDN [14]	694 K	33.32/0.9150	32.05/0.8980	31.33/0.9200	38.07/0.9761	
	BFDN [42]	1021 K	33.52/0.9166	32.09/0.8978	31.92/0.9256	38.36/0.9765	
	BSRN [41]	326 K	33.70/0.9186	32.23/0.9002	32.29/0.9303	38.69/0.9772	
	Swin-IR [39]	878 K	33.35/0.9156	32.00/0.8970	31.41/0.9207	38.06/0.9757	
	LDL [36]	694 K	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774	
	Proposed	447 K	33.79/0.9196	32.20/0.8998	32.40/0.9301	38.88/0.9774	
	× 3	SRCNN [7]	57 K	29.28/0.8209	28.41/0.7863	26.24/0.7989	30.59/0.9107
FSRCNN [35]		12 K	29.43/0.8242	28.53/0.7910	26.43/0.8080	30.98/0.9212	
VDSR [9]		665 K	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9310	
DRCN [11]		1774 K	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.31/0.9328	
DRRN [12]		297 K	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.74/0.9390	
CBPN [16]		677 K	30.00/0.8350	28.96/0.8001	27.56/0.8376	32.51/0.9369	
IMDN [14]		703 K	30.04/0.8370	28.97/0.8030	27.57/0.8400	33.00/0.9403	
BFDN [42]		1102 K	30.29/0.8407	29.06/0.8034	27.38/0.8404	33.50/0.9440	
BSRN [41]		340 K	30.41/0.8437	29.15/0.8064	28.33/0.8561	33.67/0.9456	
Swin-IR [39]		886 K	30.10/0.8372	28.96/0.8010	27.66/0.8415	33.02/0.9404	
Proposed		456 K	30.40/0.8428	29.13/0.8056	28.35/0.8557	33.85/0.9460	
× 4		SRCNN [7]	57 K	27.49/0.7503	26.90/0.7101	24.52/0.7221	27.66/0.8505
		FSRCNN [35]	12 K	27.59/0.7535	26.98/0.7150	24.62/0.7280	27.90/0.8517
		VDSR [9]	665 K	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8809

Table 6 (continued)

Scale	Model	Parameter size	Set14	B100	Urban100	Manga109
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
× 8	DRCN [11]	1774 K	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.98/0.8816
	LapSRN [18]	813 K	28.19/0.7720	27.38/0.7280	25.21/0.7560	29.09/0.8845
	DRRN [12]	297 K	28.21/0.7720	27.40/0.7284	25.44/0.7638	29.46/0.8960
	CBPN [16]	677 K	28.26/0.7723	27.49/0.7281	25.50/0.7630	29.42/0.8942
	IMDN [14]	715 K	28.35/0.7770	27.58/0.7340	25.68/0.7730	30.09/0.9024
	BFDN [42]	1180 K	28.60/0.7806	27.59/0.7349	26.07/0.7837	30.47/0.9084
	BSRN [41]	352 K	28.63/0.7836	27.61/0.7380	26.22/0.7911	30.57/0.9103
	Swin-IR [39]	897 K	28.45/0.7779	27.44/0.7313	25.71/0.7719	29.91/0.9008
	LDL [36]	710 K	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
	Proposed	470 K	28.66/0.7830	27.61/0.7371	26.27/0.7909	30.74/0.9114
	SRCNN [7]	57 K	23.86/0.5443	24.14/0.5043	21.29/0.5133	22.46/0.6606
	FSRCNN [35]	12 K	23.94/0.5482	24.21/0.5112	21.32/0.5090	22.39/0.6357
	VDSR [9]	655 K	23.20/0.5110	24.34/0.5169	21.47/0.5289	22.73/0.6688
	DRCN [11]	1774 K	24.25/0.5510	24.49/0.5168	21.71/0.5289	23.20/0.6686
	LapSRN [18]	813 K	24.45/0.5792	24.54/0.5293	21.81/0.5555	23.39/0.7068
	BSRN [41]	360 K	24.88/0.5961	24.70/0.5410	22.37/0.5977	24.28/0.7517
Proposed	520 K	25.02/0.6426	24.80/0.5968	22.46/0.6170	24.60/0.7811	

I. Ablation experiments

In order to verify the effectiveness of SFMU and DFMU, ablation experiments are conducted in this paper to verify the superiority of our model in the test dataset Set5 for the case of image magnification times.

The convergence process of these five networks is given in Fig. 8. In this paper, 18 RRDB blocks are chosen as the baseline, and these five networks have the same number of RRDBs. When the SFMU module and DFMU module are added to the baseline respectively in this paper, two curves, Baseline + SFMU and Baseline + DFMU, are obtained. Thus, it is verified that both modules can effectively improve the performance of the baseline. When the DRB is removed from the DFMU module in this paper, the curve Baseline + DFMU_no_DRB is obtained. Comparing the curve Baseline + SFMU, we can see that after losing the DRB constraint, although the network converges faster, the final PSNR decreases by 0.03 dB, but it is still 0.04 dB higher than the

baseline network, thus verifying that the effectiveness of the SFMU and DFMU modules. When two modules are added to the baseline network simultaneously in this paper, the curve Baseline + SFMU + DFMU is obtained. It can be seen that the combined performance of the two modules is better than that of only one module. These quantitative and visual analyses demonstrate the effectiveness of SFMU and DFMU.

Table 7 gives the experimental results for the case when the network contains one or both of the SFMU module and the DFMU module. From the table, it can be seen that when the network of this paper contains both SFMU and DFMU modules, the values improve by 0.07 dB and 0.05 dB, respectively, compared to the cases when only SFMU and DFMU modules are included, and the maximum value is obtained at SSIM.

To better demonstrate the effect of the MPRU module in the network, the feature maps containing only the shallow feature extraction and the feature maps with the MPRU module added are visualized in this paper, where Fig. 9(a)

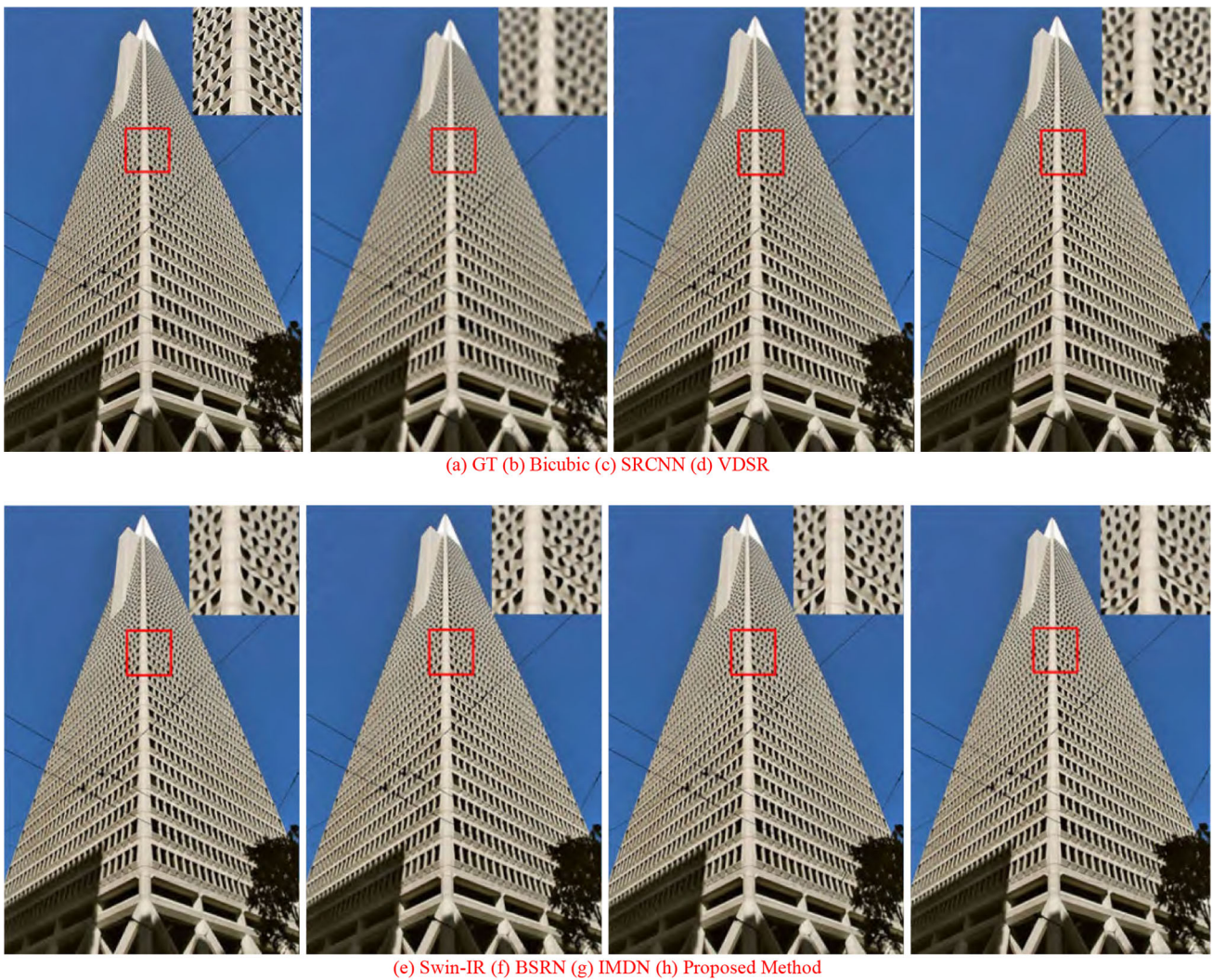


Fig. 4 Super-resolution results of “Img048” in Urban100 dataset for $\times 4$ factor

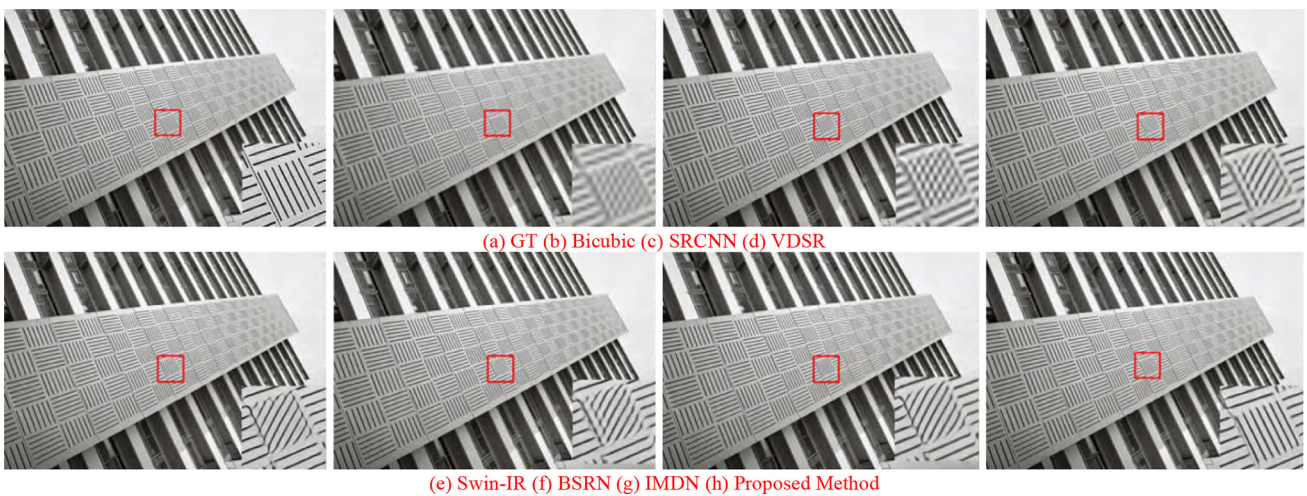


Fig. 5 Super-resolution results of “Img092” in Urban100 dataset for $\times 4$ factor

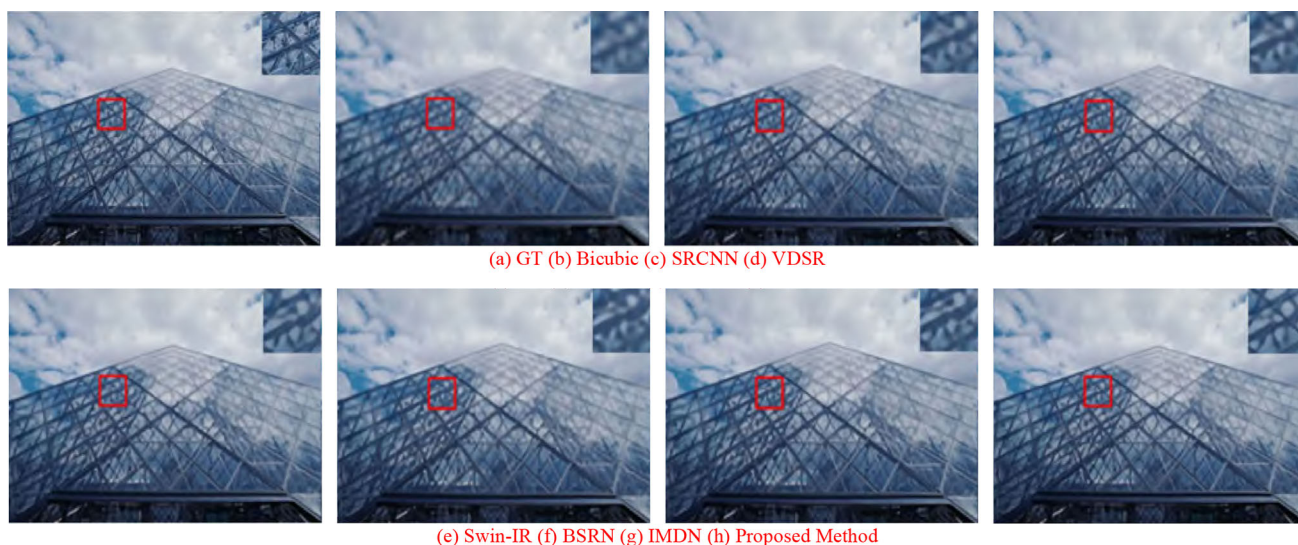


Fig. 6 Super-resolution results of “223,061” in BSD100 dataset for $\times 4$ factor

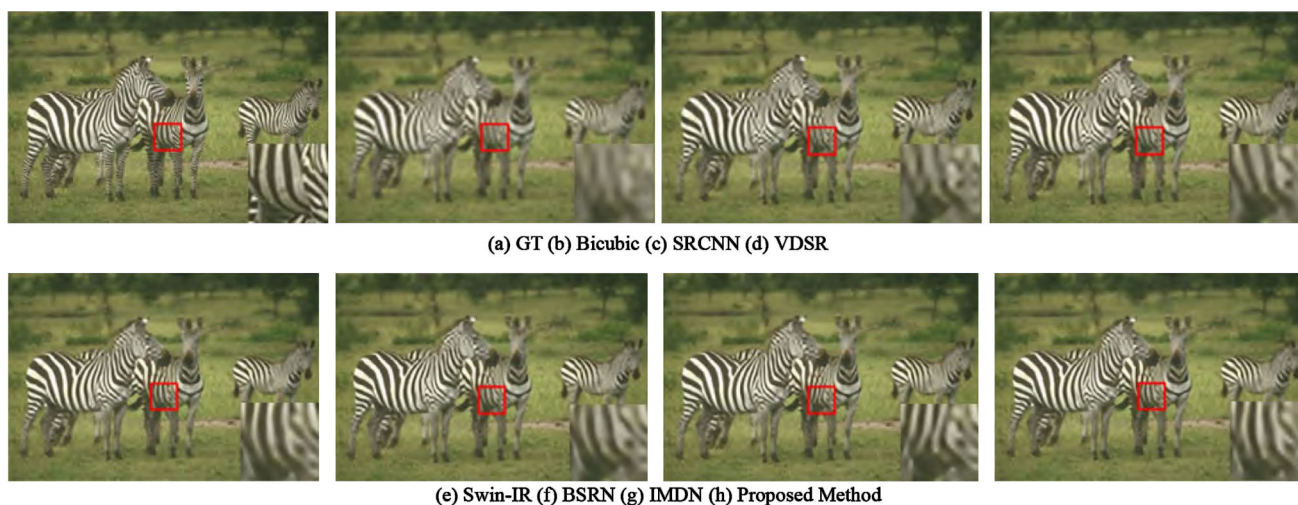


Fig. 7 Super-resolution results of “253,027” in BSD100 dataset for $\times 4$ factor

Fig. 8 Convergence analysis on SFMU and DFMU. The curves for each combination are based on the PSNR on Set5 with $\times 4$ factor in 800 epochs

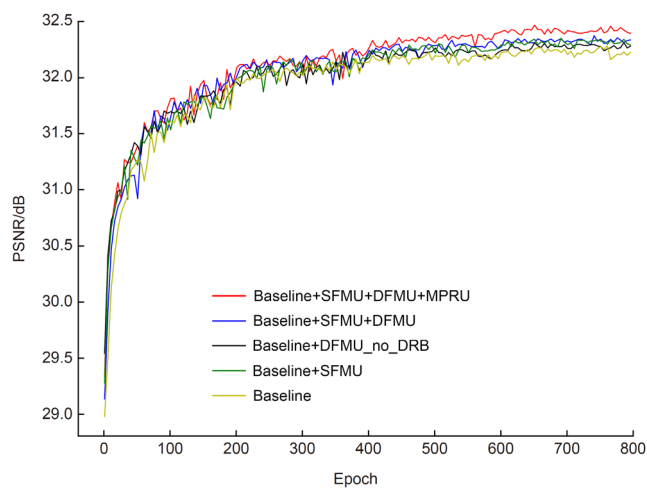
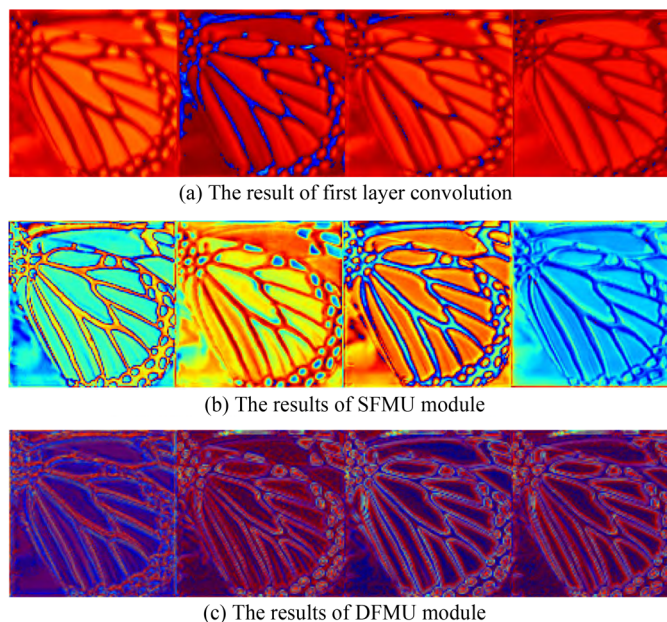


Table 7 Results of SFMU and DFMU module with $\times 4$ scale on Set5

Baseline	✓	✓	✓	✓
SFMU	×	✓	×	✓
DFMU	×	×	✓	✓
PSNR/dB	32.28	32.35	32.37	32.42
SSIM	0.8962	0.8971	0.8972	0.8982

Fig. 9 Results of each module in the network

represents the results of the network output at the first convolution layer, and Fig. 9(b) and Fig. 9(c) represent the results of the DFMU module output and the MRPDU module output, respectively. From Fig. 9(b) and Fig. 9(c), it can be seen that the DFMU module learns a large number of self-similar features of the image, for example, the circular spots on the butterfly are well-recovered. The MRPDU module, on the other hand, learns more details of the image texture. The experimental results show that the two modules in the network of this paper play a good role in feature enhancement.

J. Parameter and time analysis

In order to further verify the effectiveness of the proposed model, MFFN is analyzed and compared with some current deep learning super-resolution methods that are recognized to achieve better results, including IMDN [14], LDL [36], Swin-IR [39], and BSRN [41], in terms of the number of parameters and the computational volume, and the results of parameters and computational volume are shown in Table 8. The table shows that MFFN achieves better objective metrics while significantly scaling down the number of parameters and the computational effort of the network. The number of MFFN model parameters is approximately equal to 53% of the number of IMDN and Swin-IR parameters and 53%

of their computational effort when scaled up by a factor of $\times 2$ on the Set14 dataset, but the PSNR and SSIM results obtained are very similar. Although the number of parameters and computational effort of MFFN are slightly higher than those of the BSRN method, the obtained PSNR and SSIM values are improved by 0.07 dB and 0.0022 compared to the BSRN method.

This demonstrates that MFFN achieves a better balance between image reconstruction quality and model compression as well as computational efficiency, i.e., MFFN can obtain better PSNR and SSIM results even with fewer parameters. MFFN achieves similar reconstruction quality when compared with the LDL method, which is currently superior in objective metrics, but MFFN parameters are much less.

4 Conclusion

This paper proposes a lightweight multi-level feature fusion network for reconstructing high-quality super-resolution images. In this paper, a double-layer nested residual block (DRB) is designed to extract image feature information. The number of feature channels is first expanded and then compressed, and convolution layers with different receptive fields

Table 8 Comparison of parameter size and computational cost on Set5

Methods	Parameter	Running time (s)	PSNR/dB	SSIM
LDL	710 K	1.3203	33.71	0.9212
Swin-IR	886 K	1.1025	33.94	0.9206
BSRN	156 K	0.3645	33.87	0.9192
IMDN	893 K	0.2532	33.63	0.9177
Proposed	470 K	0.3285	33.94	0.9214

are used to reduce the number of parameters. To effectively transmit feature information in the double-nested residual block, this paper designs the autocorrelation weight unit (ACW), which generates weight information by calculating the feature information, and then uses the weight information to weigh the feature information to ensure that the high-weight feature information is processed. effective delivery. In this paper, two-layer nested residual blocks are formed into residual groups to extract deep hierarchical feature models, and shallow feature mapping unit (SFMU) is constructed to extract multi-scale and multi-level shallow feature information. The multi-path reconstruction unit (MPRU) fuses the deep layer feature information with the shallow layer feature information to reconstruct a high-quality super-resolution image. The experimental results show that the above module design helps reconstruct high-quality images, and the proposed model can effectively enhance image stripes and reconstruct high-quality super-resolution images. Compared with other lightweight models, our model achieves a better balance between performance and model scale.

Funding This work is supported by the Natural Science Foundation of Hunan Province of China under Grant 2020JJ4623, Changsha Major Science and Technology Projects under Grant KQ2102007, KQ1703018, KQ1706064, A Project Supported by Scientific Research Fund of Hunan Provincial Education Department under Grant 22A0701, Scientific Research Project of Hunan University of Information Technology under Grant XXY02ZD01, College Students' Innovative Entrepreneurial Training Plan Program of Hunan University of Information Technology under Grant X202213836002, Smart Manufacturing Barcode Traceability Management System under Grant 20224301020010 and CON202204070272 (Hunan WUJO High-Tech Material Corporation Limited), University-Industry Collaborative Education Program under Grant 202102536008 and 221003279124130, China University Innovation Funding - Beslin Smart Education Project under Grant 2022BL055, A Project Supported by Teaching Reform Research Fund of Hunan Province General Higher Education Schools under Grant HNJC-2022-1335 and 2022 Part-time Vice President of Science and Technology of Changsha Enterprises from Changsha Science and Technology Bureau.

Data availability All data generated or analyzed during this study are included in this published article and its supplementary information files. The authors would like to thank DIV2K, Set5, Set14, BSD100, Urban100, and Manga109 datasets, which allowed us to train and evaluate the proposed model.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Shi, W., Caballero, J., Ledig, C., Zhuang, X., Bai, W., Bhatia, K., et al.: Cardiac image super-resolution with global correspondence using multi-atlas patchmatch. In: Proceedings of Medical Image Computing and Computer-Assisted Intervention. Berlin, Heidelberg: Springer Berlin Heidelberg, p. 916 (2013).
- Luo, Y., Zhou, L., Wang, S., Wang, Z.: Video satellite imagery super resolution via convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **14**(12), 2398–2402 (2017)
- Zou, W.W.W., Yuen, P.C.: Very low resolution face recognition problem. *IEEE Trans. Image Process.* **21**(1), 327–340 (2011)
- Chen, Y., Liu, L., Phonevilay, V., Gu, K., Xia, R., Xie, J., Zhang, Q., Yang, K.: Image super-resolution reconstruction based on feature map attention mechanism. *Appl. Intell.* **51**(7), 4367–4380 (2021)
- Chen, Y., Liu, L., Tao, J., Xia, R., Zhang, Q., Yang, K., Xiong, J., Chen, X.: The improved image inpainting algorithm via encoder and similarity constraint. *Vis. Comput.* **37**, 1691–1705 (2021)
- Xia, R.L., Chen, Y.T., Ren, B.B.: Improved anti-occlusion object tracking algorithm using Unscented Rauch-Tung-Striebel smoother and kernel correlation filter. *J. King Saud Univ. Comput. Inf. Sci.* **34**(8), 6008–6018 (2022)
- Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Proceedings of the European Conference on Computer Vision. Zurich, Switzerland: Springer, Cham, pp. 184–199 (2014).
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
- Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using deep convolutional networks. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Las Vegas, USA, pp. 1646–1654 (2016).
- He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Las Vegas, USA, pp. 770–778 (2016).
- Kim, J., Kwon Lee, J., Mu Lee, K.: Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Las Vegas, USA, pp. 1637–1645 (2016).
- Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Honolulu, USA, pp. 3147–3155 (2017).

13. Li, Z., Yang, J., Liu, Z., Yang, X., Jeon, G., Wu, W.: Feedback network for image super-resolution. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Long Beach, USA, pp. 3867–3876 (2019).
14. Hui, Z., Gao, X., Yang, Y., Wang, X.: Lightweight image super-resolution with information multi-distillation network. In: Proceedings of the 27th ACM International Conference on Multimedia. Association for Computing Machinery, New York, USA, pp. 2024–2032 (2019).
15. Ahn, N., Kang, B., Sohn, K.A.: Fast, accurate, and lightweight super-resolution with cascading residual network. In: Proceedings of the European Conference on Computer Vision. Springer, Cham, Zurich, Switzerland, pp. 252–268 (2018).
16. Zhu, F., Zhao, Q.: Efficient single image super-resolution via hybrid residual feature learning with compact back-projection network. In: Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop. IEEE, Seoul, Korea, pp. 2453–2460 (2019).
17. Li, J., Fang, F., Mei, K., Zhang, G.: Multi-scale residual network for image super-resolution. In: Proceedings of the European Conference on Computer Vision. Springer, Cham, Zurich, Switzerland, pp. 517–532 (2018).
18. 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 2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Honolulu, USA, pp. 5835–5843 (2017).
19. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE, Honolulu, USA, pp. 1132–1140 (2017).
20. Timofte, R., Agustsson, E., Gool, L.V., Yang, M., Zhang, L., Lim, B., et al.: Nitre 2017 challenge on single image super-resolution: methods and results. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE, Honolulu, USA, pp. 1110–1121 (2017).
21. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Salt Lake City, USA, pp. 2472–2481 (2018).
22. Liu, J., Zhang, W., Tang, Y., Tang, J., Wu, G.: Residual dense network for image super-resolution. In: Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Seattle, USA, pp. 2356–2365 (2020).
23. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision. Springer, Cham, Zurich, Switzerland, pp. 286–301 (2018).
24. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.: Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Salt Lake City, USA, pp. 4510–4520 (2018).
25. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-ResNet and the impact of residual connections on learning. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence. IEEE, San Francisco, USA, pp. 4278–4284 (2017).
26. Wang, Z., Chen, J., Hoi, S.C.H.: Deep learning for image super-resolution: a survey. *IEEE Trans. Pattern Anal. Mach. Intel.* (2020).
27. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. *arXiv preprint arXIV:1412.6980* (2014).
28. Samlirans, T., Kingma, D.P.: Weight normalization: a simple reparameterization to accelerate training of deep neural networks. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. IEEE, Barcelona, Spain, pp. 901–909 (2016).
29. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on non-negative neighbor embedding. In: Proceedings of the 23rd British Machine Vision Conference. BMVA Press, Guildford, UK, pp. 135.1–135.10 (2012).
30. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Proceedings of International Conference on Curves and Surfaces. Berlin, Germany: Springer, Heidelberg, pp. 711–730 (2010).
31. 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 2001 International Conference on Computer Vision. IEEE, Vancouver, Canada, pp. 416–423 (2015).
32. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Boston, USA, pp. 5197–5206 (2015).
33. Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., et al.: Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools Appl.* **76**(20), 21811–21838 (2017)
34. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
35. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Springer, Cham, Zurich, Switzerland, pp. 391–407 (2016).
36. Liang, J., Zeng, H., Zhang, L.: Details or Artifacts: a locally discriminative learning approach to realistic image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, IEEE, USA (2022) <https://arxiv.org/abs/2203.09195>.
37. Zhang, K., Zuo, W., Zhang, L.: Learning a single convolutional super-resolution network for multiple degradations. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Salt Lake City, USA, pp. 3262–3271 (2018).
38. Mou, C., Wu, Y.Z., Wang, X.T., Dong, C., Zhang, J., Shan, Y.: Metric learning based interactive modulation for real-world super-resolution. In: Proceedings of the 2022 European Conference on Computer Vision, CoRR abs/2205.05065.
39. Liang, J.Y., Cao, J.Z., Sun, G.L., Zhang, K., Gool, L.V., Timofte, R.: Swin-IR: image restoration using swin transformer. In: Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops, Montreal, Canada: IEEE, pp. 1833–1844 (2021).
40. Chen, L.Y., Chu, X.J., Zhang, X.Y., Sun, J.: NAFNet: nonlinear activation free network for image restoration. In: Proceedings of NTIRE 2022 Challenge on Efficient Super-Resolution, <https://github.com/murufeng/FUIR>.
41. Li, Z.Y., Liu, Y.Q., Chen, X.Y., Cai, H.M., Gu, J.J., Qiao, Y., Dong, C.: Blueprint separable residual network for efficient image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, IEEE, USA, pp. 833–834 (2022).
42. Liu, J., Tang, J., Wu, G.S.: Residual feature distillation network for lightweight image super-resolution. In: Proceedings of the European Conference on Computer Vision Workshops, Glasgow: UK, pp. 41–55 (2020).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



recognition, image processing, etc. Email chenyt@hnuit.edu.cn.

Yuantao Chen received the BS degree in Computer Science and Technology from Jiangnan Petroleum Institute. He received the MS degree in Geodetection and Information Technology from Yangtze University. He received the PhD degree in Control Science and Engineering from Nanjing University of Science and Technology, Nanjing, China, in 2014. He is an associate professor at Hunan University of Information Technology. His research interests include pattern



Runlong Xia received the BS degree in Electronic Commerce from Hunan Normal University in 2010. He is a research assistant at Hunan Institute of Scientific and Technical Information. His research interests include news communication and public opinion analysis, etc. Email xiarunlong@vip.qq.com.



Kai Yang received the Master degree in Mechanical Engineering from Jilin University in 2014. He is an engineer at Technology Department of Hunan ZOOMLION Intelligent Technology Corporation Limited. His research interests include mechanical engineering, intelligent control technology, etc. Email yangkai@zoomlion.com.



Ke Zou received the BS degree in Computer Science and Technology from Northeastern University in 2005. He is a director and researcher at Informatized Office, Hunan WUJO High-Tech Material Corporation Limited. His research interests include big data processing, information processing, etc. Email zouk@hnwjgk.cn.