IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 3, 2021, accepted June 17, 2021, date of publication June 23, 2021, date of current version July 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3091899

Light Weight IBP Deep Residual Network for Image Super Resolution

HAI LIN[1] AND JUNJIE YANG[2]
[1]Department of Information Science, Zhanjiang Preschool Education College, Guangdong 524300, China
[2]College of Information Science and Technology, Lingnan Normal University, Guangdong 524048, China

Corresponding author: Hai Lin (linhai@zhjpec.edu.cn)

This work was supported by the Zhanjiang Preschool Education College Online Course Construction Project under Grant JPKC20210602.

This work did not involve human subjects or animals in its research.

ABSTRACT Single -image super resolution (SR) is used to reconstruct a high-resolution image with more high-frequency details based on a low-resolution image as input. In recent years, image SR reconstruction based on deep learning methods has shown a considerably better performance than traditional methods. Early deep-learning-based methods deepen convolutional layers and directly reconstruct high-resolution images with complex neural networks. However, with the stacking of modules, network depth and model parameters increase, thereby raising computational resource; hence, it is difficult to apply on low-configuration devices. Furthermore, existing methods ignore the high-frequency details of the image, resulting in unsatisfactory performance. To solve these problems, a lightweight network model that applies the iterative back projection (IBP) mechanism to the network and reduces the dimensionality of the input image features is proposed. The proposed network model consists of three parts, namely, entrance module, main body module, and exit module. It designs a lightweight modular design to reduce the model calculation and control the network depth more easily by adjusting the number of ADB modules. The main part of the model consists of four lightweight accelerating deep residual back projection (ADB) modules. Each ADB module initially decreases the dimensionality of the input image features through a 1×1 convolutional layer to reduce the amount of calculation. Then, IBP is used to back project the image iteratively. Each ADB module only performs the downsampling back projection operation because the image features become larger after upsampling. Through three iterative downsampling back projection units, the high-frequency features of the image are fully explored, and the output image features are then compared with the shallow layer. Image feature fusion is used as input of the next ADB module, and the output part combines the high- and low-frequency image feature output by multiple ADB modules to complete the upsampling by using the PixelShuffle method to generate high-resolution images. Several experiments confirm that the proposed algorithm achieves better SR reconstruction accuracy with faster reconstruction speed than existing image SR methods.

INDEX TERMS Image SR, residual network, light-weight convolutional neural network.

I. INTRODUCTION

Image super resolution (SR) [1] is a widely-used image processing technique for acquiring images with high spatial resolution. Several examples of image SR applications in the real world include face recognition in surveillance videos, improving resolution of images in remote sensing, object detection in scenes (especially small objects), mobile smart devices, and HD television sets. The lightweight model is more suitable for mobile smart devices and surveillance camera, which require less computation and memory. In recent years, with the rapid development of artificial intelligence technology, image SR models based on deep learning have been widely explored. Dong *et al.* [2] proposed the SR convolutional neural network (SRCNN), which initially uses three-layer CNN to fit the nonlinear mapping, to realize image block extraction, feature representation, feature nonlinear mapping, and final reconstruction. This method is a pioneer of deep neural network, which is introduced in SR reconstruction. The experiment results

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

showed that SRCNN performs better than other traditional methods.

In deep learning research, theoretical studies [3] show that in several cases, increasing the depth of network effectively obtains more image hierarchical structure features. VDSR [4] is the first model that applies a very deep network in single-image SR, which has a network architecture of 20 layers. To train a deep model similar to this approach, the authors selected a relatively high learning rate to speed up the convergence and used gradient clipping to prevent gradient explosion. To overcome the difficulty of training deep recursive CNN, DRCN [5] and ResNet [6] adopted the residual structure model. Based on this structure, the authors in [7] proposed SRResNet, which is composed of 16 residual units (one residual unit is composed of two nonlinear convolutional layers with residual learning). Lee *et al.* proposed EDSR [8] that achieved state-of-the-art performance. Compared with the residual unit in the previous work, EDSR eliminates the use of BN, greatly raises the number of output features of each layer, and increases the difficulty of training.

In many SR studies [9], people usually utilize the results of bicubic and HR images to achieve supervised learning. In addition, unsupervised SR methods based on the GAN network, similar to [10], [11], can achieve good results. Especially, [11] can effectively overcome the lack of paired images of HR and LR. This network does not require paired or aligned training datasets; it is composed of unpaired kernel, noise correction network CNet, and pseudo paired stochastic resonance network SRNet, where CNet is used to train unpaired and clear LR images, and SRNet is used to train the generation of paired clear images from LR to HR. The SR reconstruction effect is good; however, it requires high computing cost and hardware configuration of SR image reconstruction due to the use of dual network. Alam *et al.* [12] proposed the network based on adversarial network, which effectively processes the photo-realistic image and reconstructs the image similar to the original image. The method captures the EIA by connecting to a lens array and the camera sensors of a 2D microscope. Then, OVI that contains multiple directional view images that can generate 3D perception to the observer is generated from EIA according to the mapping algorithm. The high-quality resolution enhanced image can be obtained when the directional view image is directly used to feed the SR algorithm. Abbass *et al.* [13] proposed the method that uses bicubic interpolation to enlarge the chromaticity component. Then, they adjusted the chromaticity component by guiding the filter and took the image brightness as the reference when the chromaticity component of bicubic interpolation was the input of the filter to obtain the reconstruction image with a sharper edge and a higher resolution.

To detect and restore small-scale pedestrians with more details from the monitoring image better, Pang *et al.* proposed JCS-Net [14] that consists of two subnets, which are used for SR and classification. In the monitoring image, large-scale pedestrians have more detailed information. The SR subnet can extract relevant image features from large-scale pedestrians to restore the high-frequency features of small-scale pedestrian images and realize the image SR reconstruction of small-scale pedestrians.

CNN is also widely used in motion deblurring of video images in mobile scenes. Liu *et al.* [15] proposed a decoupling end-to-end CNN model based on cooperative learning. This model eliminates the assumption that only a single degradation is available. The model is more suitable for different types of degraded images by decoupling and cooperative learning. It can process each degradation by developing the corresponding restoration network and flexibly process multiple degradation simultaneously to achieve the decomposition and synthesis of multi -image SR in continuous motion and motion deblurring.

Although many excellent SR methods achieve a high SR performance, several studies show that the following problems remain in existing SR methods: (1) Most SR methods, such as RDN [16] and SRFBN [17], do not focus on the aggregation of residual features and the collection of high-frequency features. Thus, reproducing high-frequency details in HR images is difficult. (2) Many SR methods, such as EDSR [8] and DBPN [18], show high accuracy of deep models, but deploying them to real-world scenarios, for which massive parameters and computational burden may account, is difficult. This network is designed to address several problems on high computational cost and difficulty in controlling network depth. Therefore, a lightweight SR method is urgently required for application to reality with accuracy retention. To solve these problems, a lightweight deep model should be designed for SISR, and the existing deep model should be simplified to reduce the parameters and computation with minimal performance degradation. Therefore, developing lightweight, fast SR methods becomes a new direction of the current SR research.

Any image can be divided into two parts, namely, low frequency and high frequency. The former refers to the area where the image intensity changes slowly, namely, the place where large color blocks are located. The latter represents the area where the image intensity changes sharply, usually the edge of the image. Iterative back projection (IBP) [18] is an early SR algorithm for obtaining high-frequency image features. It iteratively calculates the reconstruction errors and propagates them back to acquire more high-frequency image features to use large filters. In several existing networks, the large filter is not used because it slows down the convergence speed and might achieve local optimal results. However, using iterative projection units enables the network to suppress this limitation and achieve better performance on a large scaling factor even with shallow networks. Various previous works pointed out that the fusion of dense connection of high-frequency features and residual features can effectively assist in reconstructing images. The present paper aims to explore a unified framework that can fully integrate the shallow features and residual features with fewer parameters and lower computational cost. Iterative downsampling back projection (IBP) is used to collect high-frequency features to

realize the lightweight, high accuracy of SR image finally. To achieve this object, a residual network is proposed for lightweight IBP (ADBNet), which emphasizes a lightweight modular design. Thus, model network depth is easier to control. It is composed of three parts, namely, entrance module, main body module, and exit module. The main body consists of four lightweight accelerating deep residual back projection (ADB) modules. Each ADB module initially decreases the dimension of the shallow features of the input image to reduce calculation through 1×1 convolutional layer and then performs downsampling back projection only on the image. Through three iterative downsampling back projection units, the high-frequency features are fully obtained, and the output image features are fused with the shallow features again as the input of the next ADB module. In the output part, high- and low-frequency features from multiple ADB modules are fused, and the PixelShuffle method is adopted to complete the upsampling to generate high-resolution images. Compared with the existing single image SR reconstruction model, the main contributions of this paper are as follows:

a) An image SR model named ADBNet that fully fuses the shallow features and the residual features of the images with less parameters and computation is proposed. The high-frequency features of the image are collected by the downsampling iteration to realize the lightweight SR of the image.

b) Combined with the traditional IBP method, after comparison of a large amount of experimental data, only downsampling back projection is implemented for LR image feature extraction. On the premise of speed, our method adequately extracts the high-frequency features, thereby improving the accuracy of image reconstruction.

c) Our method has a good application prospect in mobile devices, such as smartphone and smart screen, because of the improvement in calculation speed without losing excessive accuracy of HR images.

The remainder of this paper is arranged as follows: Related research background is reviewed in Section II. The detailed architecture of ADBNet is described in Section III. The experimental data and result analysis are presented in Section IV. Finally, a conclusion is drawn in Section V.

II. RELATED RESEARCH BACKGROUND

Many advanced algorithms, such as [16], [17] and [19], are available for high-frequency image feature extraction. Liu *et al.* [20] used full convolution network for object detection; it only consists of convolutions and deconvolutions, and has abandoned fully connected layers to achieve SR reconstruction. The convolutional layer is used to extract features as the encoder, and deconvolutional layer is used to reconstruct the image as the decoder. Multiscale encoder inputs low-resolution LR image and PC edge image with phase into CNN. It also uses multiscale decoder to guide the prediction of edge details of image to reconstruct SR image with multiscale edge details. However, the algorithm has the disadvantages of numerous model parameters and requiring PC edge image and extended time because of the complexity of the end-to-end structure. Final experimental data comparison shows that PSNR and SSIM are relatively low and should be further improved.

The images have a wide range of cross scale block similarity. According to the natural cross scale feature correspondence, high-frequency details can be searched from the LR images. Mei *et al.* [21] proposed the cross scale nonlocal attention module and applied it in the CSNLN network model. Through the SEM unit, combined with local, intrascale nonlocal, and cross-scale nonlocal feature correlation, more image high-frequency information can be mined as much as possible [22]. However, the performance can be slightly improved compared with the experimental data of other attention modules.

Haris *et al.* [18] proposed DBPN, which introduced an effective IBP into SR, to capture the interdependence of LR and HR image pairs. This SR framework attempts to utilize back projection iteratively to calculate the reconstruction error carefully to extract high-frequency features and then fuses it to improve the accuracy of the HR image. Fig. 1 shows the architecture diagram. It alternately connects the upsampling layer and the downsampling layer, and further improves the performance through dense connection, especially when magnified by a factor of 8. However, this method is computationally expensive for image feature extraction, which raises network complexity and running time. The final experimental results show that the performance is not improved remarkably compared with simply using downsampling IBP. Moreover, the dense connection does not focus on the fusion of shallow features and residual features. Similarly, SRFBN [17] adopts the iterative up/down sampling feedback block with more dense connections and learns a better representation. RBPN for video SR [23] extracts contexts from consecutive video frames and integrates these contexts through the back projection module to generate cyclic output frames. The model in this framework can better identify the deep relationship between LR and HR image pairs to provide higher-quality reconstruction results. Nevertheless, the design standard of the back projection module is still unclear. The framework has great potential and requires further exploration because this mechanism has been recently introduced into SR based on deep learning.

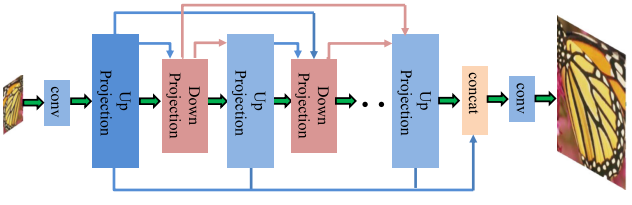


FIGURE 1. The framework of DBPN.

Although many methods based on deep learning have achieved unprecedented success in the SR field, they are unsuitable for low configuration devices because of their high

computing cost and large memory occupation. Therefore, studying the lightweight SR network for real application scenarios is very important. FSRCNN [24] designed a convolution network with a funnel structure, which does not require preprocessing and changes the input feature dimension again. In addition, it used a smaller convolution kernel and a series of setting, such as deconvolution upsampling, to realize the lightweight of the network with very few network parameters and very fast image reconstruction speed. Multiscale residual networks [25] designed a global feature fusion structure, which constructs the network by fusing the output of residual blocks with different depths. Global feature fusion costs less computation than local feature fusion [16]. Thus, this structure fuses features with different depths from a global perspective to improve the quality of the reconstructed image and reduce the number of network parameters. Fig. 2 shows the network architecture, where Mn denotes the output of the n-th residual block, I_{LR} denotes the input low-resolution image, and I_{HR}' denotes the reconstructed high-resolution image.

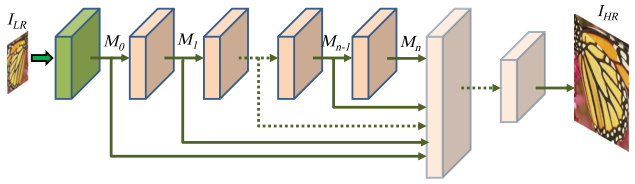


FIGURE 2. Global feature fusion.

At present, the excellent lightweight SR network models are LatticeNet [26], SLUA [27], MAFFSRN [28], and RFDN [29]. The MAFFSRN network initially uses a convolution to extract features and then uses many FFGs to refine and enhance features. Combined with the multiscale upsampling module, the residual image is obtained and finally added with the results of the bicubic image to obtain HR. The RFDN network proposes a novel residual feature aggregation framework for a more effective feature extraction. The network is composed of many RFA modules, which directly aggregate and transform the features of many residual modules and fuse them with local residual branches through the Add mode. Finally, a lightweight attention mechanism ESA is also introduced.

III. ADBNET MODEL AND METHOD

A. MODEL CONSTRUCTION

The network architecture adopts global feature fusion and introduces IBP, which is mainly composed of three modules, namely, entrance module, main module, and exit module. Fig. 3 shows the framework.

B. ENTRANCE MODULE

Low-frequency features of images can be easily detected with the strong expression ability of CNN. Therefore, a shallow CNN model is also qualified for this task. Similar to the previous methods, the entrance module consists of a convolutional layer whose convolution kernel is 3 × 3, with three input channels and 50 output channels. The algorithm can be described in Formula (1):

$$F_0 = CONV^{3\times3}(LR) \tag{1}$$

where LR denotes the input three-channel color image, and F_0 denotes the initial feature. The input image is extracted into shallow image features with 50 channels by CONV function.

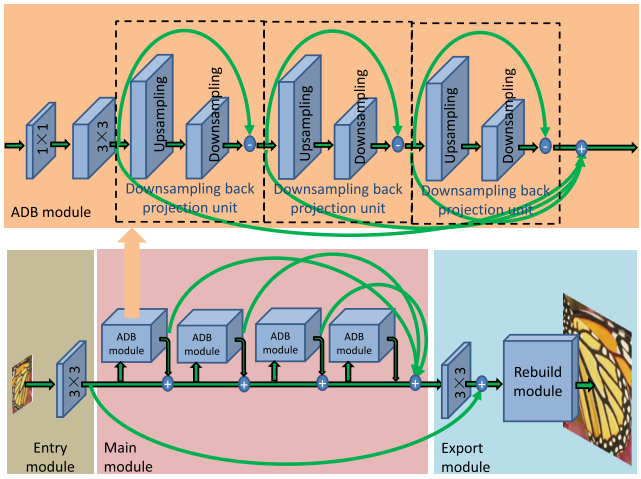


FIGURE 3. The framework of ADBNet.

C. MAIN MODULE

The main module is the key to the proposed method. Its architecture is schematically presented in Fig. 3, which is mainly composed of four lightweight ADB modules. The common residual module stacking design used in the past produces many redundant feature maps due to the deep layers, thereby increasing calculation complexity. The main module of our method adopts residual feature integration, which adds the output features of four lightweight ADB modules. Our method can adjust the number of convolutional layers by controlling the number of ADB modules in the code; thus, it can control the number of network layers and the computational expense. The ablation study reveals that four lightweight ADB modules can reach a good balance between the parameters, calculation, and performance of reconstruction images. The algorithm can be described in Formulas (2)-(6):

$$Ob_1 = ADB(F_0) \tag{2}$$
$$Ob_2 = ADB(Ob_1) \tag{3}$$
$$Ob_3 = ADB(Ob_2) \tag{4}$$
$$Ob_4 = ADB(Ob_3) \tag{5}$$
$$Ob = lrelu(CONV^{1\times1}(Ob_1 + Ob_2 + Ob_3 + Ob_4) \tag{6}$$

where F_0 is the output of the input module; Ob_1, Ob_2, Ob_3, and Ob_4 denote the outputs of four lightweight ADB modules. The output feature maps are added and fused by ADB, and then the dimension is reduced by a 1 × 1 convolutional

layer for less computation. The image feature *Ob* is extracted using *LreLU* function. Finally, *Ob* is fused with F_0 through a 3×3 convolutional layer to obtain the final output *Olr* of the main module.

The *LreLU* activation function can be represented in Formula (7):

$$LreLU(x) = \begin{cases} x, & if \quad x > 0 \\ negative\_slope \times x & otherwise \end{cases} \tag{7}$$

where *LreLU* is a variant of *ReLU*, and the response to the input less than 0 is changed, reducing the sparsity of *ReLU*. The *negative_slope* coefficient ensures a weak output when the input is less than 0, alleviating the dying *ReLU* problem.

The lightweight ADB module collects the high-frequency features of the image mainly through IBP. Each ADB module contains three downsampling back projection units. The algorithm is described as follows:

$$Dc_1 = LreLU(CONV^{1\times1}(F_0)) \tag{8}$$
$$Rc_1 = CONV^{3\times3}(F_0) \tag{9}$$
$$Tc_1 = IBP(Rc_1) \tag{10}$$
$$Tc_2 = IBP(Tc_1) \tag{11}$$
$$Tc_3 = IBP\,(Tc_2) \tag{12}$$
$$Ro = CONV^{1\times1}\,(Dc_1 + Tc_1 + Tc_2 + Tc_3) \tag{13}$$

Each ADB module decreases the dimension through 1×1 convolutional layer to obtain Dc_1, outputs 50-channel features simultaneously through 3×3 convolutional layer, and then outputs 25-channel features. The input F_0 image features are processed by IBP thrice to extract high-frequency features of the LR image, and the image features Tc_1, Tc_2, and Tc_3 are obtained. All acquired image features are fused, and the dimension is reduced by 1×1 convolutional layer to obtain the final output feature *Ro*.

The IBP downsampling back projection unit includes upsampling (upsampling without back projection calculation), downsampling, back projection calculation, and merging output parts. The algorithm is described as follows:

$$\text{up sampling:} H_0^i = (M^{i-1} * u_t) \uparrow_s \tag{14}$$
$$\text{down sampling:} M_0^i = (H_0^i * d_t) \downarrow_s \tag{15}$$
$$\text{back projection calculation:} E_i^l = (M_0^i - M^{i-1}) \tag{16}$$
$$\text{merging output:} M^i = (M^{i-1} + E_i^l) \tag{17}$$

where i denotes the i-th stage, and so is the sampling factor. The LR image feature M^{i-1} from the previous stage is considered the input, and then the enlarged HR image feature H_0^i is obtained by deconvolution. The enlarged image is downsampled with the *interpolate* function and mapped back to LR to obtain the image feature M_0^i. Then, the high-frequency feature E_i^l between M^{i-1} and M_0^i is obtained by residual calculation. Finally, the output is the fusion of the input image feature and the sum of M^{i-1} and E_i^l.

In conclusion, the design of the main module strictly controls the size and number of convolutional layers by dimension reduction, using convolutional layer with small convolution kernel and image feature optimization with few (25 and 50) channels to realize the light weight of the model. It converges and transforms the high-frequency features of images obtained by multiple lightweight ADB modules and fuses them with local residual branches by using the Add method. Then, the residual features of different levels are aggregated, and more high-frequency features are generated. Thus, the accuracy of the reconstructed image is improved.

D. EXIT MODULE

This section considers the image features from the main module as input, generates 50-channel features through a convolutional layer with 3×3 convolution kernel, and then adds and fuses with the image features of the entrance module. Finally, the HR image is reconstructed by the reconstruction module. The process can be described as follows:

$$COLR = CONV^{3\times3}\,(Ob) + F_0 \tag{18}$$
$$HR = upsampler\,(COLR) \tag{19}$$

where *Ob* is the output of the main module, F_0 is the output of the entrance module, *COLR* is the image feature fused prior to reconstruction, and *upsampler* uses the PixelShuffle [30] method to achieve high-resolution image output. The main idea is to obtain a high-resolution feature map from a low-resolution one through convolution and multichannel reconstruction. This approach is effective for current image SR upsampling. Fig. 4 shows the algorithm flowchart.

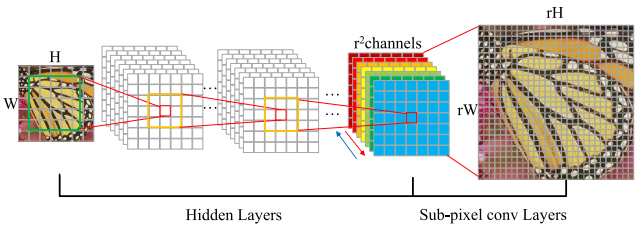


FIGURE 4. Upsampler flowchart.

Its function is to change a low-resolution image of H×W into a high-resolution image of rH×rW by subpixel operation.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL CONFIGURATION

The evaluation dataset is composed of four public datasets, namely, Set5 [31], Set14 [32], BSDS100 [33], and Urban100 [34]. The training data set adopts DIV2K [35], and the input size for training is 96×96 by randomly clipping [36]. DIV2K is a new high-quality dataset for image reconstruction; it contains 800 training images, 100 evaluation images, and 100 test images. The Set5 and Set14 datasets are low-complexity single-image SR datasets based on non-negative neighborhood embedding. The BSDS100 dataset is divided into 200-image training set and

100-image test set. The Urban100 dataset contains 100 challenging urban landscapes with different frequency bands. All training experiments are implemented on Ubuntu 16.4 and eight NVIDIA GTX2080TI GPUs with 8 G memory for each one.

Mean square error is used as the loss function; it is the most commonly used regression loss function to realize image SR with a neural network. The distance from each training point to the optimal fitting line is minimized (or the sum of squares is minimized), and the loss function is as follows:

$$J(w, b) = \frac{1}{2n} \sum_{1}^{n} \|y - a\|^2 \tag{20}$$

$a = f(l) = f(w \times x + b)$, x is the input, w and b are the network parameters, and $f(l)$ is the activation function. Adopting the mean square error as the loss function is beneficial to obtaining a higher PSNR.

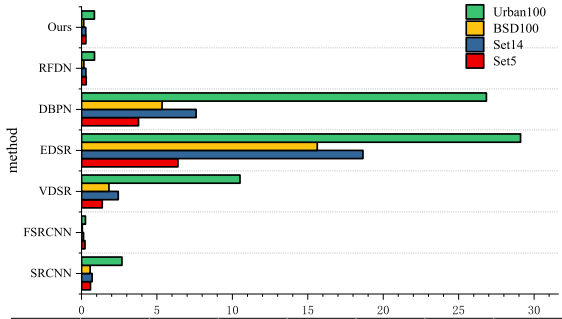
Comparison methods include SRCNN [2], FSRCNN [24], VDSR [4], EDSR [8], DBPN [18], and RFDN [29], among which RFDN is a lightweight, efficient image SR network proposed by Nanjing University and was the champion of the AIM20-ESR competition in 2020.

B. COMPARISON OF AVERAGE RUNNING TIME OF RECONSTRUCTED IMAGES

To validate the running time of our method on the computer with low configuration, the test was conducted on the PC only with integrated graphics, of which the CPU is Intel I3-8300 and the RAM is 4G. Fig. 5 is a bar chart comparing the average running time of the image reconstruction on four evaluation datasets of Set5, Set14, BSD100, and Urban100 when scaling factor is 4. The figure shows that with lightweight design and few parameters, the average time of image reconstruction of our method in the four datasets is very short. The average running times are 0.293, 0.285, 0.146, and 0.853. The higher resolution of the reconstructed image results in a more evident contrast. For example, when reconstructing the image in the Urban100 dataset, the proposed method is 31 times faster than the DBPN method using double sampling IBP, slightly faster than the lightweight RFDN network, 34 times faster than EDSR, 12 times faster than VDSR, and three times faster than the earliest SRCNN. Although FSRCNN is slightly faster than our method, its performance of image construction is poor. Therefore, our method is fast and very suitable for the operation on the platform with a low configuration.

C. ABLATION STUDY

In the study of our method, a variety of combination research on the ADB module was conducted, the backbone network was set without the ADB modules as ADBNet#. Table 1 shows that when the number of ADB modules was raised to seven, the parameter increased by 200.46k, multiadd calculation increased, and PSNR only increased by 0.11 after image reconstruction. When the number of ADB modules



	SRCNN	FSRCNN	VDSR	EDSR	DBPN	RFDN	Ours
Urban100	2.682	0.259	10.505	29.108	26.841	0.862	0.853
BSD100	0.572	0.054	1.819	15.63	5.337	0.159	0.146
Set14	0.702	0.135	2.435	18.659	7.595	0.291	0.285
Set5	0.593	0.228	1.371	6.39	3.779	0.31	0.293

FIGURE 5. Average running time of the image reconstruction of seven methods.

TABLE 1. Ablation study of ADBNet.

Combination	Parameters	Multiadds	PSNR	Time (s)
ADBNet#+ADB (×7)	735.12 K	33.34 G	32.93	0.41
ADBNet#+ADB (×2)	295.53 K	18.3 G	30.83	0.21
ADBNet#+ Upsampling back projection (×4)	3978.37 K	221.5 G	32.97	2.63
ADBNet#+ADB (×4)	534.66 K	25.2 G	32.82	0.29

was reduced to two, the PSNR value was only 30.83, which is lower than that of existing advanced methods, although the parameter was reduced by 239.13K and the calculation of multiadds was reduced. The internal structure of the ADB module was changed from the original downsampling back projection to upsampling back projection and downsampling back projection, and iterated for four times. The parameter was 3443.71k larger than the original, the calculation of multiadds increased sharply, but the PSNR value only increased by 0.15, which shows that the combination (ADBNet#+four ADBs) is a good setting. Moreover, the reconstruction speed was only 0.29 s, which is the fastest on the whole dataset and is suitable for running on the computers with low configuration.

D. COMPARISON OF MODEL COMPLEXITY WITH EXISTING METHOD

To validate the effectiveness of our method on the lightweight aspect, it was compared with the recent three years lightweight SR method in terms of parameters, multiadds, and PSNR. Fig. 6 shows that the parameter of our method is between 400K and 600K, which is similar to that of RFDN, MAFFSRN, and IDN* [37]with less parameter. The multiadd calculation of our method is 20-40 G, which is near that of the RFDN, MAFFSRN, and IDN*. Thus, our method is a lightweight model with low complexity and running time. Table 2 shows the comparison of PSNR and SSIM above our method and existing methods on 2×, 4×, and 8×. The PSNR and SSIM values of our method are the highest among the four datasets on 4× and 8×. Hence, our method can reach a good balance between performance and model complexity.

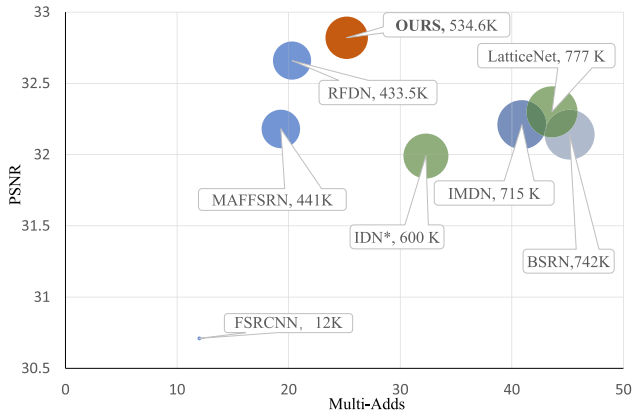


FIGURE 6. Performance comparison of existing lightweight methods on Set5 [3] (4×). Multi adds are calculated on a 720 p HR image.

TABLE 2. Comparison of PSNR and SSIM values of image reconstruction by eight methods.

model	MPL	Set5	Set14	BSD100	Urban100
		PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
Bicubic	2×	33.65 / 0.930	30.34 / 0.870	29.56 / 0.844	26.88 / 0.841
SRCNN	2×	36.65 / 0.954	32.29 / 0.903	31.36 / 0.888	29.52 / 0.895
FSRCNN	2×	36.99 / 0.955	32.73 / 0.909	31.51 / 0.891	29.87 / 0.901
MSDEPC	2×	37.39 / 0.957	32.94 / 0.911	31.64 / 0.896	29.52 / 0.895
VDSR	2×	37.53 / 0.958	32.97 / 0.913	31.90 / 0.896	30.77 / 0.914
EDSR	2×	38.11 / 0.960	33.92 / 0.919	32.32 / 0.901	32.93 / 0.935
CSNLN	2×	38.28 / 0.962	34.12 / 0.922	32.40 / 0.902	33.25 / 0.938
DBPN	2×	38.09 / 0.960	33.85 / 0.919	32.27 / 0.900	33.02 / 0.931
SMSR	2×	38.00 / 0.960	33.64 / 0.917	32.17 / 0.899	32.19 / 0.928
MAFFSRN	2×	37.97 / 0.960	33.49 / 0.917	32.14 / 0.899	31.96 / 0.926
IDN*	2×	37.85 / 0.959	33.58 / 0.917	32.11 / 0.898	31.95 / 0.926
BSRN	2×	37.78 / 0.959	33.43 / 0.915	32.11 / 0.898	31.92 / 0.926
IMDN	2×	38.00 / 0.960	33.63 / 0.917	32.19 / 0.899	32.17 / 0.928
LatticeNet	4×	38.15 / 0.9610	33.78 / 0.9193	32.25 / 0.9005	32.43 / 0.930
RFDN	2×	38.26 / 0.962	34.16 / 0.922	32.41 / 0.903	33.33 / 0.940
Ours	2×	38.25 / 0.962	34.10 / 0.920	32.39 / 0.902	33.31 / 0.935
Bicubic	4×	28.42 / 0.810	26.10 / 0.704	25.96 / 0.669	23.15 / 0.659
SRCNN	4×	30.49 / 0.862	27.61 / 0.754	26.91 / 0.712	24.53 / 0.724
FSRCNN	4×	30.71 / 0.865	27.70 / 0.756	26.97 / 0.714	24.61 / 0.727
MSDEPC	4×	31.05 / 0.879	27.79 / 0.758	27.10 / 0.719	24.53 / 0.724
VDSR	4×	31.35 / 0.882	28.03 / 0.770	27.29 / 0.726	25.18 / 0.753
EDSR	4×	32.46 / 0.897	28.80 / 0.788	27.71 / 0.742	26.64 / 0.803
CSNLN	4×	32.68 / 0.900	28.95 / 0.788	27.80 / 0.743	27.22 / 0.816
DBPN	4×	32.47 / 0.898	28.82 / 0.786	27.72 / 0.740	27.08 / 0.795
DRN-S [38]	4×	32.68 / 0.901	28.93 / 0.790	27.78 / 0.744	26.84 / 0.807
SMSR [39]	4×	32.12 / 0.893	28.55 / 0.780	27.55 / 0.735	26.11 / 0.7868
MAFFSRN	4×	32.18 / 0.894	28.58 / 0.781	27.57 / 0.732	26.52 / 0.782
IDN*	4×	31.99 / 0.892	28.52 / 0.779	27.52 / 0.733	25.92 / 0.780
BSRN [40]	4×	32.14 / 0.893	28.56 / 0.780	27.57 / 0.735	26.03 / 0.783
IMDN [41]	4×	32.21 / 0.894	28.58 / 0.781	27.56 / 0.735	26.04 / 0.783
LatticeNet	4×	32.30 / 0.896	28.68 / 0.783	27.62 / 0.736	26.25 / 0.787
RFDN	4×	32.66 / 0.900	28.88 / 0.789	27.79 / 0.744	26.92 / 0.811
Ours	4×	32.82 / 0.912	29.01 / 0.795	27.95 / 0.763	27.16 / 0.823
Bicubic	8×	24.39 / 0.657	23.19 / 0.568	23.67 / 0.547	20.74 / 0.516
SRCNN	8×	25.33 / 0.689	23.85 / 0.593	24.13 / 0.565	21.29 / 0.543
FSRCNN	8×	25.41 / 0.682	23.93 / 0.592	24.21 / 0.567	21.32 / 0.537
MSDEPC	8×	25.58 / 0.707	24.03 / 0.604	24.14 / 0.573	21.32 / 0.558
VDSR	8×	25.72 / 0.711	24.21 / 0.609	24.37 / 0.576	21.54 / 0.560
EDSR	8×	26.97 / 0.775	24.94 / 0.640	24.80 / 0.596	23.07 / 0.620
CSNLN	8×	27.12 / 0.722	25.15 / 0.650	25.02 / 0.603	23.26 / 0.635
DBPN	8×	27.21 / 0.784	25.13 / 0.648	24.88 / 0.601	23.25 / 0.622
DRN-S	8×	27.41 / 0.790	25.25 / 0.652	24.98 / 0.605	22.96 / 0.641
RFDN	8×	27.11 / 0.781	25.01 / 0.642	24.73 / 0.595	23.13 / 0.610
Ours	8×	27.33 / 0.791	25.26 / 0.653	24.95 / 0.613	23.36 / 0.635

E. COMPARISON OF IMAGE RECONSTRUCTION PERFORMANCE

Fig. 7 shows a local magnification of each image selected from the Set5, Set14, BSDS100, and Urban100 datasets and reconstructed with 4× SR by seven methods. According to the comparison of PSNR and SSIM values in Table 2 and the details in Fig. 7, bicubic image quality assessments are the lowest. The entire bicubic reconstruction image is blurry and very smooth, and the features of high and low frequencies are not evident. The PSNR and SSIM values of SRCNN are much higher than that of bicubic, and the reconstruction effect is much better than that of bicubic. Details are evidently enhanced, and high-frequency features have been initially revealed from the low frequency. FSRCNN is a lightweight upgraded version of SRCNN. Although the values of PSNR and SSIM are only slightly higher than those of the former, a great progress is still observed from the comparison of the details in Fig. 7, especially in the comparison of the last graph. Evident square blocks can be seen from the two bright lines at the bottom of the former, and few square blocks can be seen from that of the latter. The PSNR and SSIM values of EDSR are much higher than those of FSRCNN, but the effect of detail improvement is not evident from Fig. 7. The PSNR and SSIM values of DBPN are slightly lower than those of EDSR at 2×. Although a slight improvement is found compared with EDSR at 4× and 8×, DBPN adopts iterative up and down sampling back projection to improve the extraction ability of high-frequency features. The detail comparison in Fig. 7 shows that the display of high-frequency features by DBPN is higher than that of the four previous methods with rich details. However, several mistakes are observed from the four comparison graphs. For example, more superimposed wavy lines are found above the double bright lines in the last picture, and one more prominent white pattern in the middle of the first picture is insufficiently smooth. The PSNR and SSIM values of RFDN are slightly higher than those of DBPN at 2× and 4× and lower than those of DBPN at 8×. According to the detailed comparison graph at 4× in Fig. 7, the high-frequency characteristics of the second graph of RFDN are less than those of DBPN, the first graph is more improved than the previous, and the overall reality of the fourth graph is better than that of DBPN. Our method adopts the global feature fusion network backbone structure, such as RFDN, but the PSNR and SSIM values are slightly lower than RFDN at 2× mainly because the RFDN uses residual module in image feature extraction method, which uses skip connections in the network design to avoid gradients from vanishing and allows the design of very deep networks. In addition, it can fully combine the low-level feature and high-level feature, and achieve a good performance in low magnification. However, with the increase in magnification, our method can better obtain and restore the high-frequency features with IBP iterative downsampling back projection. The PSNR and SSIM values are slightly higher than those of the IBP module at a low magnification, which is decided by the IBP algorithm. The lower magnification results in the less evident effect of high-frequency feature extraction. At 4× and 8× magnification, the PSNR and SSIM values of our method are higher than those of the previous algorithms, and the reconstruction effect is the best with rich details and strong sense of reality. For example, the last contrast graph shows

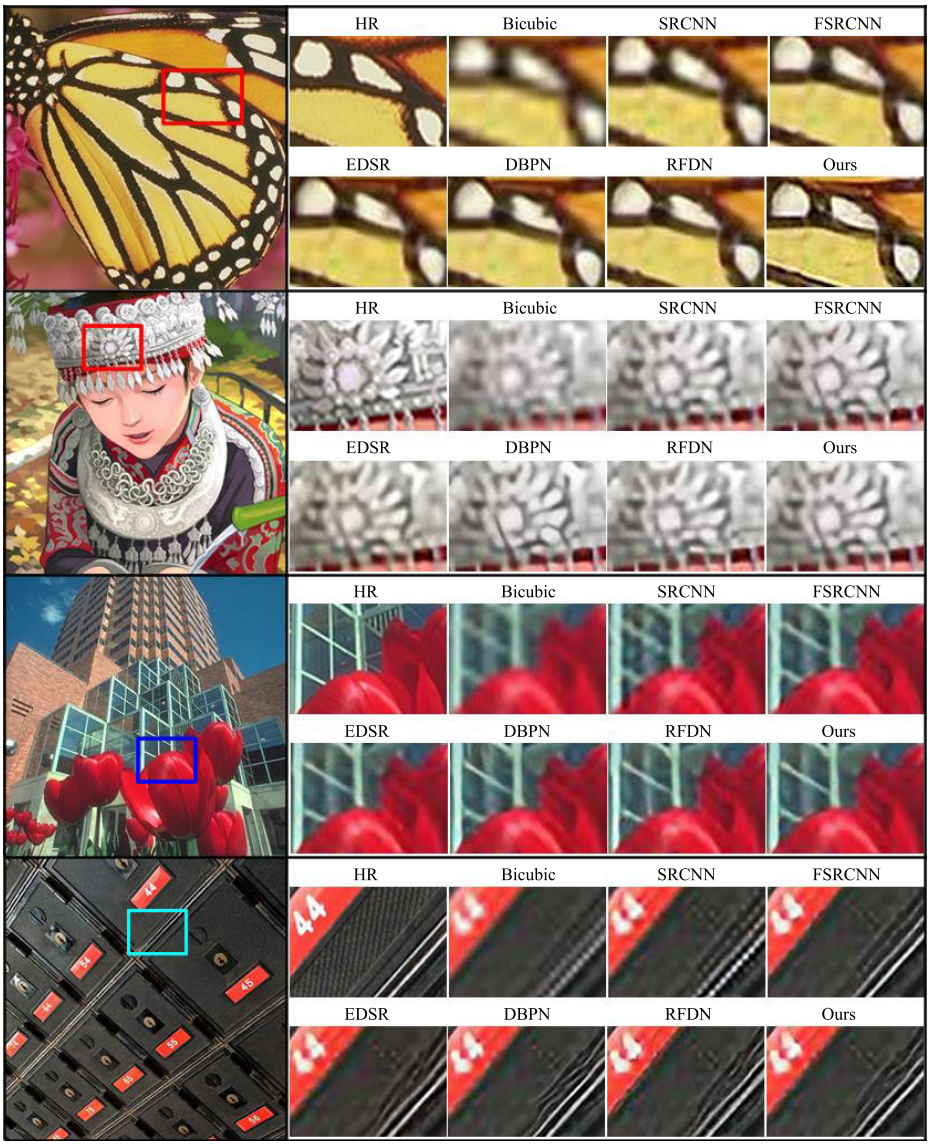


FIGURE 7. Reconstruction performance of different methods on the magnification of 4 times.

that the high-frequency features are evident, and the curve above the double bright line is closer to the HR graph than other algorithms. Our method simultaneously requires less computation and memory to adapt to the actual application better.

F. THE COMPARISON OF VISUAL FEATURE MAP BETWEEN IMAGE FEATURE EXTRACTION AND THE AGGREGATION OF HIGH-FREQUENCY FEATURE

To observe the network model better for extracting the image feature and aggregating the high-frequency feature, the visual convolution graph is adopted to analyze EDSR, DBPN, RFDN, and ADBNet at 8× SR. Fig. 8 shows that EDSR adopts residual structure, and the convolution layer of 0-31 residual units is selected for visualization. The convolution layers in the selection have numerous invalid black blocks. Moreover, the difference between image feature extraction and high-frequency aggregation of several units in the 11 feature maps selected from each unit is not very evident. The DBPN network adopts dense connection and iterative upsampling and downsampling back projection. Several convolution layers are selected for visualization from down1 to down5 and up1 to up7 units, and the results show that invalid black blocks in the extracted convolution layers are remarkably reduced. Further aggregation of high-frequency features is found evidently from the feature maps selected from each unit, but the final image feature extraction is not the best. RFDN adopts a lightweight network structure with global feature fusion and residual feature aggregation. The convolution layers of B1.c1_d to b1.c5_d units is visualized, and the finding reveal that the number of invalid black blocks of selected convolution layers is similar

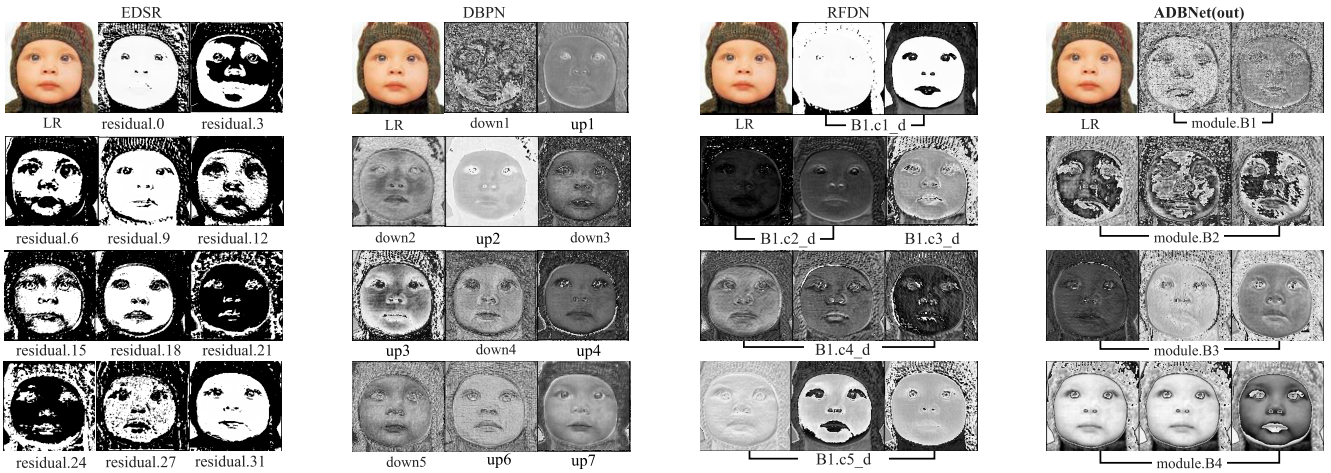


FIGURE 8. The comparison of visual feature map between image feature extraction and the aggregation of high-frequency feature of 4 networks.

to that of DBPN. However, the aggregation of high-frequency features is evidently inferior to that of DBPN. Our network ADBNet adopts global feature fusion and iterative down-sampling back projection. The convolution layers of module.B1 to module.B4 is visualize, and no valid black blocks are found in the visualized convolution. Moreover, each feature map has evident image features, and the aggregation of high-frequency features is evidently better than the three previous networks. We can conclude that the network architecture adopting global feature fusion and iterative down-sampling back projection not only realizes the lightweight of the network but also highlights the high-frequency feature aggregation of the image.

V. CONCLUSION

An ADBNet network is proposed to solve the problems of massive model size, complex model structure, and slow running speed in current mainstream image SR methods. The proposed ADBNet achieves the light weight and high efficiency of the network by reducing feature dimension and controlling network depth, using small convolution kernels and global feature fusion. IBP is combined to enhance the extraction of image feature and aggregation of high-frequency feature to improve the reconstruction accuracy (PSNR and SSIM) of the output image and obtain more detailed image in visualization. The comprehensive benchmark evaluation shows that our ADBNet model is effective in terms of model complexity and running speed. In the future, we will delve into the transformation of image SR.

REFERENCES

[1] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.

[2] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.

[3] G. Montúfar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks," 2014, *arXiv:1402.1869*. [Online]. Available: http://arxiv.org/abs/1402.1869

[4] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1646–1654.

[5] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1637–1645.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[7] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4681–4690.

[8] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 136–144.

[9] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Mar. 23, 2020, doi: 10.1109/TPAMI.2020.2982166.

[10] Y. Chen, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, and D. Li, "Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2018, pp. 91–99.

[11] S. Maeda, "Unpaired image super-resolution using pseudo-supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 291–300.

[12] M. S. Alam, K.-C. Kwon, M.-U. Erdenebat, M. Y. Abbass, M. A. Alam, and N. Kim, "Super-resolution enhancement method based on generative adversarial network for integral imaging microscopy," *Sensors*, vol. 21, no. 6, p. 2164, Mar. 2021.

[13] M. Y. Abbass, K.-C. Kwon, M. S. Alam, Y.-L. Piao, K.-Y. Lee, and N. Kim, "Image super resolution based on residual dense CNN and guided filters," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 5403–5421, Feb. 2021.

[14] Y. Pang, J. Cao, J. Wang, and J. Han, "JCS-Net: Joint classification and super-resolution network for small-scale pedestrian detection in surveillance images," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 12, pp. 3322–3331, Dec. 2019.

[15] H. Liu, J. Qin, Z. Fu, X. Li, and J. Han, "Fast simultaneous image super-resolution and motion deblurring with decoupled cooperative learning," *J. Real-Time Image Process.*, vol. 17, no. 6, pp. 1787–1800, Dec. 2020.

[16] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.

[17] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 3867–3876.

[18] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.

[19] N. Wang, Y. Zhang, and L. Zhang, "Dynamic selection network for image inpainting," *IEEE Trans. Image Process.*, vol. 30, pp. 1784–1798, Jan. 2021.
[20] H. Liu, Z. Fu, J. Han, L. Shao, S. Hou, and Y. Chu, "Single image super-resolution using multi-scale deep encoder–decoder with phase congruency edge map guidance," *Inf. Sci.*, vol. 473, pp. 44–58, Jan. 2019.
[21] Y. Mei, Y. Fan, Y. Zhou, L. Huang, T. S. Huang, and H. Shi, "Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 5690–5699.
[22] N. Wang, S. Ma, J. Li, Y. Zhang, and L. Zhang, "Multistage attention network for image inpainting," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107448.
[23] M. Haris, G. Shakhnarovich, and N. Ukita, "Recurrent back-projection network for video super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 3897–3906.
[24] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 391–407.
[25] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 517–532.
[26] X. Luo, Y. Xie, Y. Zhang, Y. Qu, C. Li, and Y. Fu, "LatticeNet: Towards lightweight image super-resolution with lattice block," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 272–289.
[27] D. Wu, L. Ding, S. Yang, and D. Tao, "SLUA: A super lightweight unsupervised word alignment model Via cross-lingual contrastive learning," 2021, *arXiv:2102.04009*. [Online]. Available: http://arxiv.org/abs/2102.04009
[28] A. Muqeet, J. Hwang, S. Yang, J. Heum Kang, Y. Kim, and S.-H. Bae, "Multi-attention based ultra lightweight image super-resolution," 2020, *arXiv:2008.12912*. [Online]. Available: http://arxiv.org/abs/2008.12912
[29] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 2359–2368.
[30] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1874–1883.
[31] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 135.1–135.10.
[32] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.*, 2010, pp. 711–730.
[33] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2. Jul. 2001, pp. 416–423.
[34] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5197–5206.
[35] R. Timofte, E. Agustsson, L. V. Gool, M.-H. Yang, and L. Zhang, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 114–125.
[36] L. Zhang, L. Song, B. Du, and Y. Zhang, "Nonlocal low-rank tensor completion for visual data," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 673–685, Feb. 2021.
[37] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 723–731.
[38] Y. Guo, J. Chen, J. Wang, Q. Chen, J. Cao, Z. Deng, Y. Xu, and M. Tan, "Closed-loop matters: Dual regression networks for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5407–5416.
[39] L. Wang, X. Dong, Y. Wang, X. Ying, Z. Lin, W. An, and Y. Guo, "Exploring sparsity in image super-resolution for efficient inference," 2020, *arXiv:2006.09603*. [Online]. Available: http://arxiv.org/abs/2006.09603
[40] J.-H. Choi, J.-H. Kim, M. Cheon, and J.-S. Lee, "Lightweight and efficient image super-resolution with block state-based recursive network," 2018, *arXiv:1811.12546*. [Online]. Available: http://arxiv.org/abs/1811.12546
[41] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2024–2032.

HAI LIN received the bachelor's degree in computer science and application. He is currently a Lecturer. He has published three articles in provincial journal. He has presided over two college-level projects with the college. His research interests include artificial intelligence and digital image.

JUNJIE YANG was born in Hubei, China, in March 20, 1969. He received the B.S. degree in physics from Huazhong Normal University, in 1991, and the M.S. degree in automation of electric power systems and the Ph.D. degree in system analysis and integration from the Huazhong University of Science and Technology (HUST), in 1998 and 2006, respectively. His research interests include modern optimization theory and algorithm and its application in optimal operation of hydropower energy systems.