Neurocomputing 489 (2022) 98–111

Contents lists available at ScienceDirect

Neurocomputing

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

A novel image super-resolution algorithm based on multi-scale dense recursive fusion network ☆

Xiang Lv [a], Changzhong Wang [a,*], Xiaodong Fan [a], Qiangkui Leng [b], Xiaoli Jiang [a]

[a] College of Mathematics, Bohai University, Jinzhou 121013, Liaoning, China
[b] College of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, Liaoning, China

ARTICLE INFO

Article history:
Received 12 October 2021
Revised 8 February 2022
Accepted 11 February 2022
Available online 22 February 2022
Communicated by Zidong Wang

Keywords:
Multi-scale
Recursive fusion
Super-resolution reconstruction
Residual learning
Channel attention

ABSTRACT

With the increasing maturity of convolution neural network (CNN) technology, the image super-resolution reconstruction (SR) method based on CNN is booming and has achieved many remarkable results. Undoubtedly, SR has become the mainstream direction of image reconstruction technology. However, most of the existing SR methods improve the reconstruction performance by increasing the depth of networks, which also increases the number of parameters, number of network computations, and difficulty of training network. To solve the performance complexity dilemma in SR, this paper proposes a network called a multi-scale dense recursive fusion network (MSDRFN). The network is composed of three parts: initial feature extraction module, multi-scale dense fusion group module and recursive reconstruction module. In detail, rough features are first extracted through a shallow feature extraction module, and then are inputted into multi-scale dense fusion blocks (MSDFBs) group. Each MSDFB makes full use of image features in convolution kernels of different sizes to obtain different hierarchical features, and further these output features are inputted into the channel attention mechanism to learn their corresponding weights. All MSDFBs outputs will be restored to high resolution images via the recursive reconstruction module. In addition, the network supplements the information loss with residual learning, which is embodied in one long-jump connection and several short-jump connections. The proposed network is mainly trained in the Pytorch deep learning framework. In comparison experiments on benchmark datasets, the proposed method outperformed the most advanced convolutional methods.

1. Introduction

Image super-resolution reconstruction (SR) improves image and video quality in computer vision applications. SR transforms observed low-resolution images or videos into corresponding high-resolution images or videos [1,2]. Such image processing technology is important for target detection, medical imaging, satellite remote sensing [3], and other applications. However, Image SR is a multi-solution problem because there are multiple mappings between a low resolution (LR) image and its corresponding high resolution (HR) image [4]. When up-sampling factor is large, it is difficult to acquire the high frequency information for recovering SR [5]. In such cases, SR must find the intrinsic interaction between LR and HR images through some prior knowledge of images. A nonlinear mapping describing the intrinsic interaction between LR and HR images can be learned through CNN-based deep learning.

Although existing CNN-based SR models have achieved relatively good results (see the next section), their training is hampered by a large number of parameters and calculations. Most studies focus on deep networks with more complex structures and algorithms, fail to take into account different hierarchical features, which can waste a lot of time and storage space.

To effectively address the complexity problem, this work proposes a network called multi-scale dense recursive fusion network (MSDRFN) that can reconstruct images with super-resolution. This model improves the image-reconstruction performance from those of the existing CNN-based SR methods without consuming excessive computing resources and time. The highlights of this method are full utilization of convolution kernels at different scales, and the embedding of multi-scale dense fusion blocks (MSDFBs) in the network for extracting more information. In addition, for a

This work was supported by National Natural Science Foundation of China under Grants 61976027, 61572082, Liaoning Revitalization Talents Program (XLYC2008002) and Liaoning Provincial Education Department Project (LJKZ1026, LJKZ1030).

* Corresponding author.
E-mail addresses: lvxiang1997@126.com (X. Lv), changzhongwang@126.com (C. Wang), fxd@emails.bjut.edu.cn (X. Fan), jxls209@163.com (Q. Leng), qkleng@126.com (X. Jiang).

https://doi.org/10.1016/j.neucom.2022.02.042
0925-2312/

large number of features with redundant information, a 1×1 convolution kernel is used for dimensionality reduction to remove redundant features and obtain useful information.

Finally, all MSDFBs' outputs are restored to high-resolution images by recursive reconstruction module. The residual learning is used to supplement information loss, and detailed features and contour features are combined to form rich features. In summary, we establish a MSDRFN with three-scale convolution feature-extraction branches and multiple residual learning structures. The method achieves super-resolution reconstruction of a single image with satisfactory performance using a reduced number of parameters. The contributions of this work are as follows:

1) This paper proposes an MSDFB with a new balanced two-stage architecture, which uses residual learning, multi-scale depth recursive fusion and channel attention mechanism to play a role in the flow of information across the network.
2) The hierarchical feature information of images is learned, and the recursive idea is used to connect each hierarchical feature to the image reconstruction module, and the image construction performance, computing resources and time are properly balanced.
3) On the basis of MSDFB group, we proposes a novel MSDRFN network, unlike most of the existing methods, the network directly extracts the features from LR images, and by using multiple MSDFBs extracts the most useful information.
4) Multiple Sub-pixel Convolutional Recursive reconstruction Nets (SpCRNets) are used to recover hierarchical features. The number of HR image features is increased and the better reconstruction effect is achieved. Our proposed MSDRFN is superior to other SR methods.

The remainder of this work is organized as follows. Section 2 summarizes existing SR approaches and briefly reviews the network structures of other approaches related to our approach. Section 3 describes our proposed MSDRFN in detail. Section 4 shows the results and analyses of different experiments. Conclusions and suggestions for future work are presented in Section 5.

2. Related works

2.1. Image super-resolution based on CNN

Most of the existing SR methods are divided into three classes: interpolation-based approaches [6–15], reconstruction-based approaches [16–24], and learning-based approaches [25–35]. Interpolation-based algorithms are a relatively simple image SR methods with low complexity, but their effectiveness is limited and the high-frequency details in images are not easily repaired, leading to complex calculations, blurred images, and low real-time performance during reconstruction. Therefore, Interpolation-based approaches are not suitable for image reconstructions with large magnification factors (such as ×3 and ×4).

The problem has been overcome by deep learning technology based on convolutional neural networks (CNNs). In deep learning, active learning using CNN outperforms other methods. In 2014, Dong et al. [26] introduced the first CNN-based SR method called SRCNN, which directly learned an end-to-end mapping from LR to HR images using a 3-layer CNN. Although SRCNN outperformed previous methods, the interpolation in SRCNN may introduce noise and affect network performance. When the network level deepens, the network speed slows down and the network training becomes more difficult.

Dong et al.'s work [26] triggered much interest in SR methods with deep learning. Two relatively shallow networks based on SRCNN are fast super-resolution CNN (FSRCNN) [27] and efficient sub-pixel CNN (ESPCN) [36]. Both networks outperformed SRCNN in speed and accuracy. FSRCNN improved SRCNN by adopting a faster model with fewer parameters and a deconvolution layer that enlarges images. Therefore, the network can input non-interpolated small images. ESPCN proposed a subpixel convolution up-sampling method that amplifies images by channel number expansion and pixel point rearrangement.

These shallow networks did not meet the requirements of reconstruction performance. Serially developed CNN-based models had since overcome the limitations of the above networks, most of which improved the performance by deepening networks. In 2016, two very deep CNNs denoted as VDSR and DRCN were proposed in [29,30], respectively. Both networks adopted global residual learning (GRL) and obtained successful SR by deep neural networks. One might suppose that a deep CNN is requisite for image SR. GRL can alleviate the training depth, but does not remove many-parameter problem. To reduce the number of parameters, Tai et al. [37] proposed a deep recursive residual network (DRRN), which shared the parameters that proliferate in a deep network. They also added local residual learning, which alleviated the loss of image information passing through the deep network. To obtain the up-sampling factor of large multiples (more than 8 times), Lai et al. designed a network model with a pyramid structure, denoted as LapSRN [32]. Each level of the LapSRN structure produced only the result of a double magnification of the original image.

Many deep networks with various structures, parameter settings, and loss functions were then proposed. Examples are the Generative Adversarial network (GAN) [38] and Enhanced Depth Residual network (EDSR) [39]. These network structures lacked scalability when learning from a single model because the corresponding LR image was obtained from HR image through bicubic down-sampling. EDSR was the super-resolution scheme that won the NTIRE2017 championship challenge. As stated in the accompanying paper, the EDSR was improved mainly by removing some redundant modules, which increased the generalization of the model.

However, most of these methods did not fully utilize the feature information in LR image, instead, they deepened networks to improve performances. As the number of network layers increases, the image feature information can gradually disappear through the transmission process. At this stage, how to fully exploit the existing features in networks had become a key issue in high-quality image reconstruction. This problem had been resolved by two networks with multi-scale convolution modules, namely, Cascading Cross-over network (CMSC) [31] and Multi-scale Residual network (MSRN) [41]. CMSC captured the detailed high-frequency information in the images in a coarser-to-finer way, whereas MSRN reconstructed some multi-scale residual blocks to restore high-quality images.

Some researchers proposed Deep Back-Projection networks (DBPN) in [51], which constructed interdependent up–down sampling modules. Each sampling module represented different image degradation and high-resolution component. Qin et al. [52] proposed a Deep Adaptive Dual-network (DADN) bidirectional SR network, in which one branch of the network was trained for focusing simple image regions, and the other was trained for processing hard image regions. In 2020, Liu et al. proposed two lightweight and effective image super-resolution networks in [53,56], which called Attention Based Multi-Scale Residual Network (AMSRN) and Improved Dual-Scale Residual Network (IDSRN). These two networks were mainly multi-scale feature capture methods based on attention mechanism. But they did not take into account the hierarchical feature information. Shortly afterwards, Zhang et al. [59] made full use of the hierarchical features of residual branches,

combined several residual blocks together, and promoted the forward propagation of features of each residual branch by adding jump connections.

Many of the above methods have some problems. Due to the potentially large distribution difference between HR and LR images, the effectiveness of information utilization is affected. To solve this problem, Lu et al. [60] proposed Matching Acceleration and Spatial Adaptation for Reference-based Image super-resolution (MASA) network, where two new modules were designed to solve these problems. The matching extraction module adopted corresponding matching scheme from coarse to fine, which greatly reduced the calculation cost. The spatial adaptive module learned the difference of distribution between LR and HR images and remapped the feature distribution to LR feature distribution in a spatially adaptive way. This scheme makes the network robust when dealing with different reference images.

Recently, Li et al. [61] proposed an adversarial learning strategy of region perception to guide the model to focus on adaptive generation of details of textured regions. Lu et al. [62] proposed a novel efficient super-resolution transformer (ESRT) for fast and accurate image super-resolution. ESRT is a hybrid transformer where a CNN-based SR network was designed in the front to extract deep features. Facing to multi-domain image super-resolution, Rao et al. [63] proposed a deep super-resolution residual StarGAN, a novel and scalable approach that resolved LR images for multiple LR domains using only a single model. Ahn et al. [64] introduced a neural structure search (NAS) method to automate structure construction process. They extended NAS into the super-resolution realm, found a lightweight, densely connected network called DeCoNASNet, and defined a complexity-based penalty to solve the super-resolution problem of images. Villar-corrales et al. [65] believed that some algorithms could amplify noise in the up-sampling step and often failed to reconstruct the high-resolution image from the noise version of a low-resolution image. Therefore, they proposed a joint denoising and super-resolution method. Tran et al. [66] proposed a learning-based approach that was applied to 3D epipolar image (EPI) to reconstruct high-resolution. Through a 2-stage super-resolution framework, the proposed approach effectively addressed various super-resolution problems and enhances the quality of high-resolution EPI volume. Sun et al. [67] proposed a weighted multi-scale residual network to better balance SR performance and computational efficiency. By using multi-scale feature fusion, Liu et al. [68] proposed a cross convolution method of edge detection to locate and represent edge features for super-resolution.

2.2. Residual learning

When training a common network using standard optimization algorithms (such as gradient descent or other popular optimization algorithms) in learning-based SR, the training error first decreases and then increases as the network deepens. In theory, deepening the network should improve the training outcome. The reason for the increase of training error is that the information loss of input features causes the output information to be only composed of the learned features, which lead to gradient vanishing or gradient explosion [5].

Based on this, some scholars put forward the concept of residual learning. Fig.1 is a schematic of convolution operations with residual structure. The residual structure of a network is $R(x) = H(x) + x$, where x is the input item, $R(x)$ is the network output, and $H(x)$ is the residual term. When $H(x) = 0$, the map is identical to the input, i.e., $R(x) = x$.

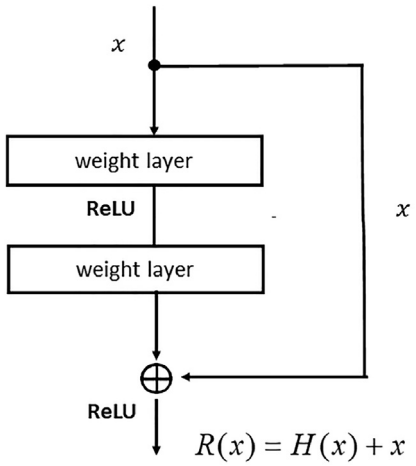


Fig. 1. Residual learning.

2.3. Feature extraction block

To avoid the disappearance or explosion of gradient in training process, He et al. [42] proposed a learning framework with residual blocks (Fig. 2(a)), which retains much of important input information. The primary role of inception block (Fig. 2(b)) is to extract richer features from images using convolution kernels of different sizes. Dense blocks (Fig. 2(c)) are used for a single-size convolution kernel. After passing through a dense block, the network depth increases by multiplying the thickness of dense block by the number of dense blocks. Zhang and Tian [43] presented a residual dense block (Fig. 2(d)) that is composed of both residual and dense blocks, which fully extracts the local and global features.

2.4. Channel attention

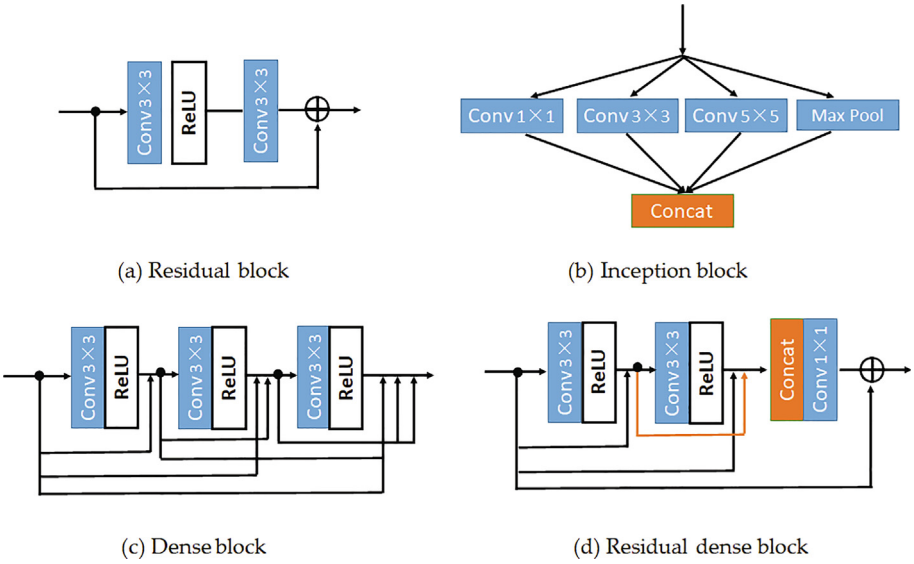
When faced with an image, human eyes can quickly scan the global image and find the target that needs attention, and focus on the key target region, while other non-key target regions are selectively ignored. This phenomenon also exists in life. Human eyes always focus on the object they need to pay attention to first, and other irrelevant objects will be ignored. Given this ability of human vision, researchers hope that computers will also have this form of attention, which can select the information that is more valuable to the immediate goal from a mass of data and continue to pay attention to it.

In 2018, Hu et al. [57] introduced a squeeze and incentive network, which adaptively adjusts channel weights by learning the contribution degree of each channel to an image. In the same year, Zhang et al. [58] proposed a residual channel attention (RCAN) network by combining residual network with channel attention mechanism. In order to eliminate useless information, the network calculates the correlation between different channel components, assigns the correlation as weights to the initial channels, and uses residual learning to supplement the information loss of the network. Experiments show that the channel attention mechanism can improve the performance of network image processing.

3. Proposed method

3.1. Network model architecture

Although the performance of existing CNN-based models has been satisfied to a certain extent, the large number of parameters and calculation in models waste more time and storage space, so



(a) Residual block (b) Inception block

(c) Dense block (d) Residual dense block

Fig. 2. Different feature extraction blocks.

that model training is greatly hindered. Current researches mainly focus on deep networks with complex structures, and do not take into account different hierarchical features. To effectively solve the complexity problem of super-resolution image reconstruction, a multi-scale dense recursive fusion network (MSDRFN), is proposed in the section, which can improve the performance of image reconstruction without consuming too much computing resources and time.

The uniqueness of the proposed method lies in a novel balanced two-stage MSDFB algorithm, which utilizes residual learning, multi-scale deep recursive fusion and channel attention mechanism to play a role in network information flow. The hierarchical feature information of an image is learned and the recursive connection is used to link each hierarchical feature to the image reconstruction module, so as to reasonably balance the performance of image construction and computing resources and time. Fig. 3 shows the architecture of the proposed network, which inputs an LR image and outputs an HR image. The network is composed of three main image processing modules: initial feature extraction module, multi-scale dense fusion group module and recursive reconstruction module.

*Initial feature extraction module.*From the original input LR, the initial feature map P_1 is captured using 3×3 convolution kernel:

$$P_1 = f_1(\mathrm{LR}) \tag{1}$$

where $f_1(\cdot)$ denotes the feature convolution operation. The output P_1 of this stage is inputted to the next stage.

*Multi-scale dense fusion group module.*Multiple multi-scale dense fusion blocks are used to extract rich features from the initial features. We thus have

$$P_2^1 = f_2^1(P_1) \tag{2}$$

$$P_2^k = f_2^k\left(P_2^{k-1}\right) \tag{3}$$

$$\mathrm{C}_1 = \left[P_2^1, \cdots, P_2^k, \cdots, P_2^N\right] \tag{4}$$

where $k = \{2, 3, \cdots N\}, f_2^1$ and f_2^k represent the convolution operations of the first block and the k – *th* block of the multi-scale dense fusion blocks(MSDFBs) group, P_2^1 and P_2^k are their corresponding extraction feature maps. C_1 denotes the connection features with $P_2^1, \cdots, P_2^k, \cdots, P_2^N$, the outputs $P_2^1, \cdots, P_2^k, \cdots, P_2^N, \mathrm{C}_1$ of this stage is inputted to the next stage.

Recursive reconstruction module. In order to make full use of deep learning to improve image resolution, we propose a reconstruction module called Sub-pixel Convolutional Recursive-Reconstruction Net (SpCRNet) as shown in Fig. 4, which adopts 1×1 convolution and sub-pixel convolution layer, and can learn information features of different levels. In this stage, the feature maps $P_2^1, \cdots, P_2^k, \cdots, P_2^N, \mathrm{C}_1$ learned in the multi-scale fusion blocks are respectively inputted into the corresponding sub-module of SpCRNet to obtain better images. Then, all hierarchical features obtained by the reconstruction module are fused. Finally, the initial image is added to the high-resolution image obtained at this stage by bicubic interpolation, and the result is taken as the final output of the network. This stage is simply implemented as:

$$P_3^1 = g_3^1\left(P_2^1\right) \tag{5}$$

$$P_3^k = g_3^k\left(P_2^{k-1}\right) \tag{6}$$

$$P_3 = \sum_{i=1}^{N+1} P_3^i \tag{7}$$

$$\mathrm{HR} = P_3 + bicubic(\mathrm{LR}) = MSDRFN(\mathrm{LR}) \tag{8}$$

where $k = \{2, 3, \cdots N\}, g_3^k$ represents the SpCRNet operation after the k – *th* layer of multi-scale feature extraction block. P_3^k represents the outputs of k – *th* SpCRNet, where *bicublc*(·) denotes bicubic interpolation operation. *MSDRFN*(·) in Eq. (8) represents the operation of the entire network.

3.2. Multi-scale dense fusion block

This subsection describes the complete branches of the MSDFB with three different-scale convolution kernels (Fig. 5). The MSDFB uses local-residual learning, multi-scale dense fusion and channel attention mechanism to capture the feature information. MSDFB is divided into two parts: multi-scale feature extraction, squeeze and excitation mechanism.

Multi-scale feature extraction. As shown in Fig. 5, this part is composed of two multi-scale convolution phases with identical structures. In each phase, three convolution kernels of different scales (7×7 , 5×5 , 3×3) are arranged in parallel for feature extraction. In each parallel path, the second convolution kernel is followed by a Leakage Rectifier Linear Unit (LRELU). On the whole,

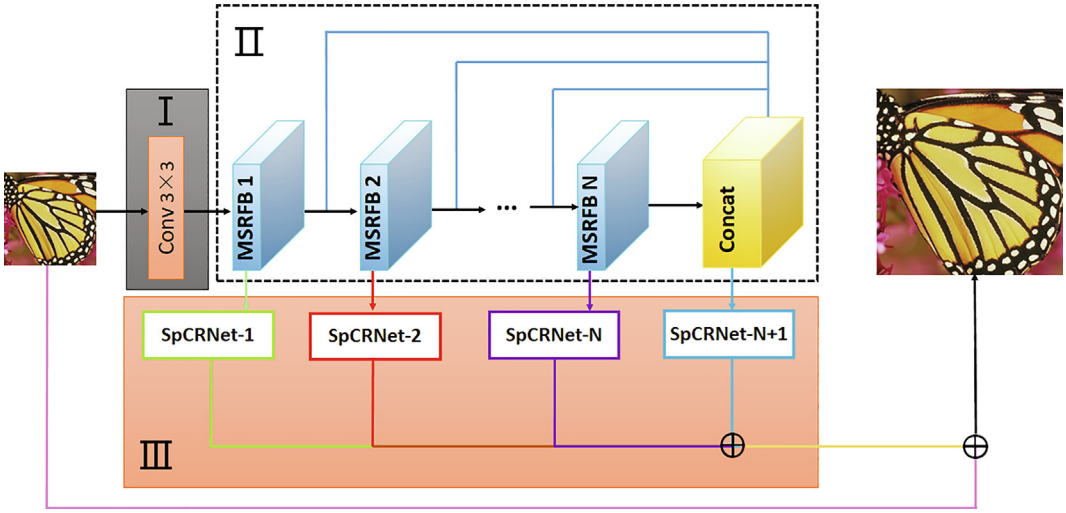


Fig. 3. Architecture of the network model.

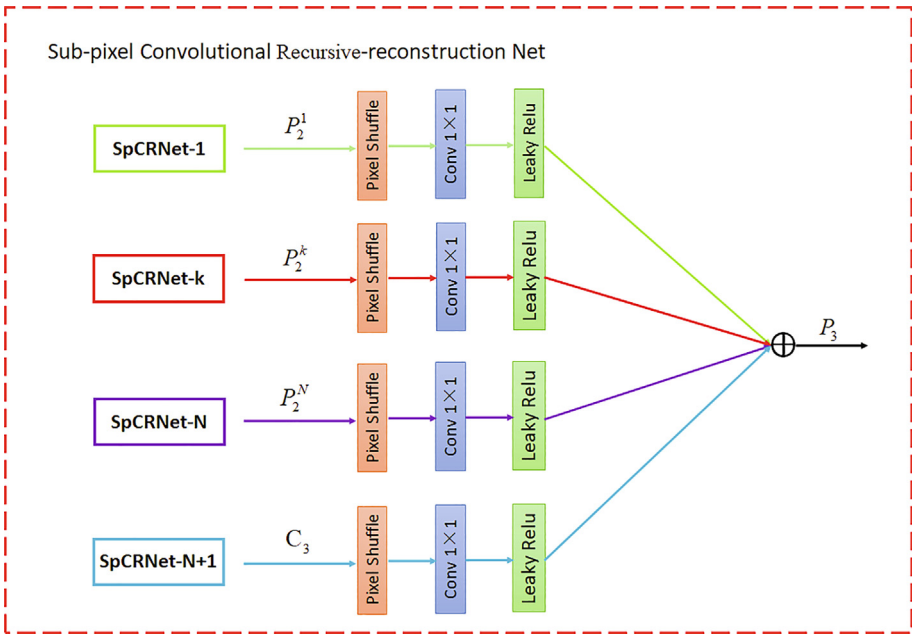


Fig. 4. Sub-pixel Convolutional Reconstruction Net (SpCRNet).

jump connection is used to fuse the initial features with the features extracted in this stage to obtain multi-scale feature information. The MSDFB inputs and outputs are denoted as S_{n-1} and S_n, respectively. The first phase is implemented as follows:

$$A_{11} = W_{11}^{3\times3} \times S_{n-1} + b_{11}^{3\times3} \tag{9}$$

$$A_{12} = f\left(W_{12}^{3\times3} \times A_{11} + b_{12}^{3\times3}\right) \tag{10}$$

$$B_{11} = W_{11}^{5\times5} \times S_{n-1} + b_{11}^{5\times5} \tag{11}$$

$$B_{12} = f\left(W_{12}^{5\times5} \times B_{11} + b_{12}^{5\times5}\right) \tag{12}$$

$$D_{11} = W_{11}^{7\times7} \times S_{n-1} + b_{11}^{7\times7} \tag{13}$$

$$D_{12} = f\left(W_{12}^{7\times7} \times D_{11} + b_{12}^{7\times7}\right) \tag{14}$$

$$C_2 = [A_{12}, B_{12}, D_{12}] \tag{15}$$

Similarly, we can get:

$$A_{21} = W_{21}^{3\times3} \times C_2 + b_{21}^{3\times3} \tag{16}$$

$$A_{22} = f\left(W_{22}^{3\times3} \times A_{21} + b_{22}^{3\times3}\right) \tag{17}$$

$$B_{21} = W_{21}^{5\times5} \times C_2 + b_{21}^{5\times5} \tag{18}$$

$$B_{22} = f\left(W_{22}^{5\times5} \times B_{21} + b_{22}^{5\times5}\right) \tag{19}$$

$$D_{21} = W_{21}^{7\times7} \times C_2 + b_{21}^{7\times7} \tag{20}$$

$$D_{22} = f\left(W_{22}^{7\times7} \times D_{21} + b_{22}^{7\times7}\right) \tag{21}$$

$$C_3 = [A_{22}, B_{22}, D_{22}] \tag{22}$$

$$E_1 = f\left(W_3^{1\times1} \times C_3 + b_3^{1\times1}\right) \tag{23}$$

$$M = S_{n-1} + C_2 + E_1 \tag{24}$$

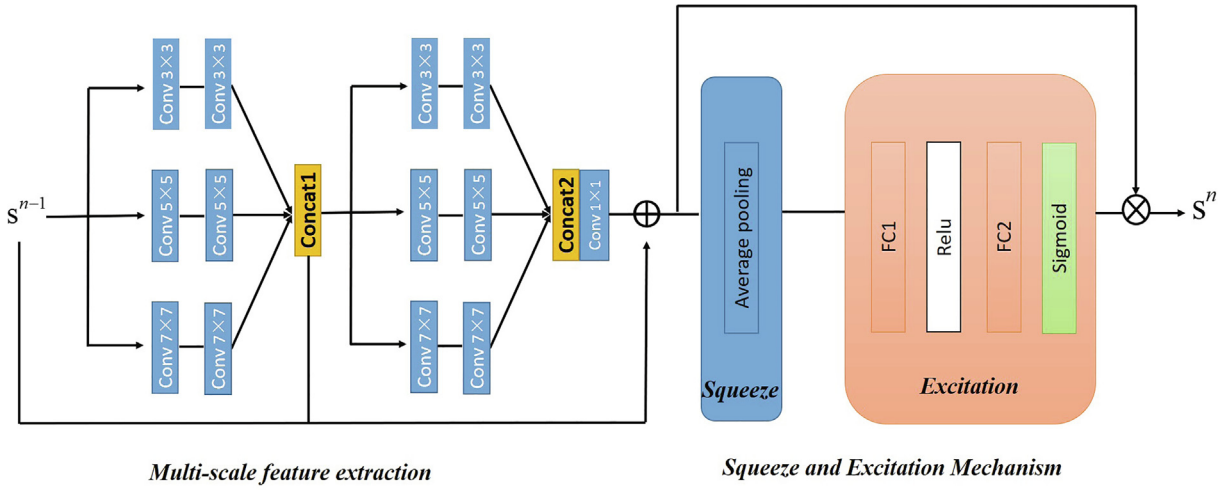


Fig. 5. Multi-scale recursive fusion block (MSDFB).

where A, B and D respectively represent the intermediate results of the three branches of the 3×3 , 5×5 and 7×7 convolution operations. C denotes the concatenate features with A, B and D, W and b represent the weight and bias of neurons, respectively. E_1 denotes the feature map obtained through a 1×1 convolution operation. The subscripts and superscripts of these symbols respectively denote the layer positions and sizes of convolution kernels. M represents the output of multi-scale feature extraction. $f(\cdot)$ is defined by a leakage-constrained linear element function, namely, the LReLU function. The slope of the LReLU's negative domain was set to 0.05. Eq. (15) and (22) represent the concatenate operation.

$$f(x) = \begin{cases} x & x \geqslant 0 \\ kx & x < 0 (0 < k < 1) \end{cases} \tag{25}$$

*Squeeze and excitation mechanism.*This part is used to learn the weights of feature map channels according to loss, so that the weights of effective feature maps are significant, while the weights of ineffective feature maps are small. The whole MSDFB represents the multi-scale dense fusion features, which not only provides rich context information but also learns the weights of the channels.

The second part is implemented as follows:

$$z_k = Ave(M_k) = \frac{1}{p \times q} \sum_{i=1}^{p} \sum_{j=1}^{q} M_k(i,j) \tag{26}$$

$$\eta = \{\eta_1, \eta_2, \ldots \eta_K\} = \delta(\omega_2 \sigma(\omega_1 z)) \tag{27}$$

$$\begin{aligned} S_n &= \eta \odot M = \left\{ \tilde{M}_1, \tilde{M}_2, \cdots, \tilde{M}_K \right\} \\ &= MSDFB(S_{n-1}) \end{aligned} \tag{28}$$

where $M = \{M_1, M_2, \ldots, M_k\}$ represents the input with K different channels, the spatial size is $p \times q$, $Ave(\cdot)$ represents the global average pool operation, z_k represents the average value of the $k-th$ channel feature M_k and $z = \{z_1, z_2, \ldots z_K\}$. $\omega_1 \in R^{\frac{K}{r} \times K}$ and $\omega_2 \in R^{K \times \frac{K}{r}}$ respectively represent the weight vectors of the two fully connected layers, r represents the scale ratio, δ and σ represent Sigmoid function and Relu functions respectively, $\eta = \{\eta_1, \eta_2, \ldots \eta_K\}$ represents the weights of the channels after learning,$\odot$ represents channel-wise multiplication.

3.3. Training and loss functions

When training the network, the usual optimization objective is the mean square error (MSE), which predicts the difference between the obtained and real HR images. In fact, MSE returns the average of many possible scenarios, so the output image is visually ambiguous and unreliable. The mean absolute error (MAE) better reflects the actual situation of prediction error. Therefore, we adopted the MAE as optimization objective function. In addition, in order to better converge the network, we use L1 regularization to balance the goal of fitting training by adding a penalty after the MAE loss. θ, λ represent the parameters of the proposed network and regularization coefficient, respectively. The optimization objective function is then defined as:

$$\mathscr{L}^{SR} = \|MSDRFN(LR) - HR\|_1 + \lambda \|\theta\|_1 \tag{29}$$

Our main objective is to solve the following optimization problem:

$$\widehat{\theta} = \arg\min_{\theta} \mathscr{L}^{SR} \tag{30}$$

$$= \arg\min_{\theta} \left(\sum_{i=1}^{N} \|MSDRFN(LR_i) - HR_i\|_1 + \lambda \|\theta\|_1 \right) \tag{31}$$

4. Experimental results and analysis

We first provided the basic settings of the datasets generated by enhancement techniques, and described the details of the network implementation. The proposed MSDRFN model was then comprehensively evaluated on several common benchmark datasets: Set5 [25], Set14 [45], BSD100 [46], and Urban100 [47] and Manga109 [49]. These datasets included images of various categories, which can be used to verify the performance of the proposed model. We calculated the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) on the luminance channel of YCbCr color [53].

4.1. Datasets

As the training datasets, 200 images were used from the Berkeley Segmentation Dataset (BSD) [40]. We know that blindly increasing depth can result in gradient disappearance or gradient explosion. Therefore, we carried out initial batch standardization on the input data of 200 images, which can ensure network convergence in back propagation more effectively. In the testing stage, we employed the above mentioned Set5 [25], Set14 [45], BSD100 [46], Urban100 [47] and Manga109 [49] datasets. To enrich the test, we selected some images from the test datasets for visual comparison (see Fig. 6).

(a) lenna (b) head (c) butterfly (d) woman

Fig. 6. Test images.

Table 1
Quantitative evaluations (PSNR/SSIM) of Models 1, 2, and 3.

Scale	Dataset	Model 1	Model 2	Model 3
×2	Set5 [25]	38.34/0.9716	**38.49/0.9723**	38.49/0.9721
×2	Set14 [45]	33.97/0.9255	**34.22/0.9267**	34.20/0.9259

Table 2
Quantitative evaluations (PSNR/SSIM) of Models 2, 4, 5, and 6

Scale	Dataset	Model 4	Model 5	Model 6	Model 2
×3	BSDS100 [46]	31.32/**0.8331**	31.25/0.8237	31.27/0.8232	**31.34**/0.8241
×3	Urban100 [47]	28.97/0.8414	28.48/0.8391	28.22/0.8291	**29.33/0.8472**
×3	Manga109 [49]	34.64/0.9436	34.28/0.9339	34.12/0.9326	**35.10/0.9469**

Table 3
Effects of the MSDFB with different numbers on Set5 for ×3SR

MSDFB	2	3	4	5
PSNR	34.49	34.65	**34.88**	34.37
SSIM	0.9320	0.9322	**0.9322**	0.9314

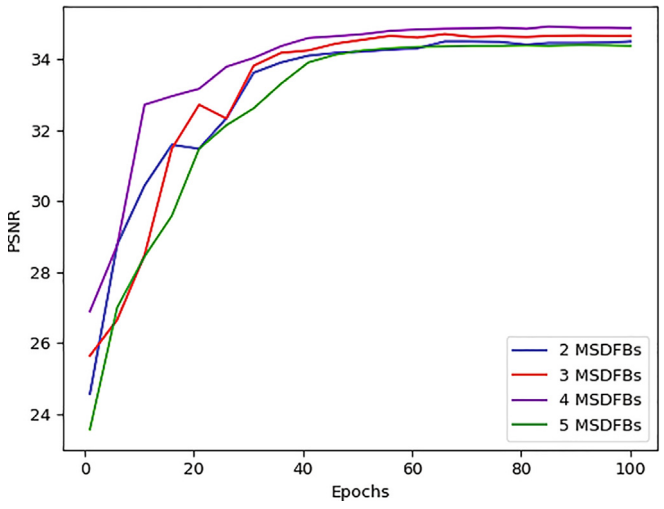


Fig. 7. PSNR comparison of different numbers of MSDFBs on Set5.

Table 4
Effects of the SE Mechanism on Set5 (PSNR (dB)/SSIM)

SEM	×2	×3	×4
PSNR	38.34/0.9715	34.49/0.9320	32.53/0.9229
SSIM	**38.49/0.9723**	**34.88/0.9322**	**32.92/0.9245**

4.2. Details of network implementation

In the proposed MSDRFN network, 64 convolution kernels were used in the initial feature extraction stage, and 4 MSDFBs were used in the multi-scale feature extraction stage to extract hierarchical features. The number of output feature maps of each convolutional layer in MSDFB was set to 32. Then we used the sub-pixel convolution operation to restore the feature map for each MSDFB, and finally merged all the features and processed the residual error to get the final feature map.

The proposed network was optimized using the Adam stochastic optimization method [48]. It was initialized with Gaussian distribution weights, the mean was zero, and the variance was 0.02. In our experiment, each training dataset was divided into 8 training batches and the corresponding test datasets were divided into 2 batches. The regularization coefficient in the loss function, basic learning rate and weight attenuation were both set to 10e-4. Training stopped after 100 epochs because there was no further continuous improvement.

The performance of any image reconstruction must be quantified by specific indices. In the present experiments, the effectiveness of the method was evaluated by the peak signal-to-noise ratio (PSNR) [54] and the structural similarity (SSIM) [55]. These indices are calculated as follows:

$$\mathrm{PSNR} = 10\log_{10}\frac{1}{MSE}, MSE = \frac{1}{mn}\left\|\hat{x} - x\right\|_F^2 \tag{32}$$

where n and m denote the width and height of the input image, respectively, x is the original image, and $\hat{x}$ is the reconstructed image:

$$\begin{aligned} \mathrm{SSIM}(x,\hat{x}) &= L(x,\hat{x}) * C(x,\hat{x}) * S(x,\hat{x}) \\ L(x,\hat{x}) &= \frac{2\mu_x\mu_{\hat{x}}+C_1}{\mu_x^2+\mu_{\hat{x}}^2+C_1} \\ C(x,\hat{x}) &= \frac{2\sigma_x\sigma_{\hat{x}}+C_2}{\sigma_x^2+\sigma_{\hat{x}}^2+C_2} \\ S(x,\hat{x}) &= \frac{\sigma_{x\hat{x}}+C_3}{\sigma_x\sigma_{\hat{x}}+C_3} \end{aligned} \tag{33}$$

Table 5
PSNR/SSIM comparison of the tested methods.

Scale	Method	Params	Set5 [25]	Set14 [45]	BSDS100 [46]	Urban100 [47]	Manga109 [49]
×2	Bicubic	–/–	33.66/0.9299	30.23/0.8687	29.56/0.8431	26.87/0.8401	30.81/0.9341
	A+ [50]	–/–	36.54/0.9544	32.28/0.9056	31.21/0.8863	29.20/0.8938	35.37/0.9663
	SelfExSR [47]	–/–	36.49/0.9537	32.22/0.9034	31.18/0.8855	29.54/0.8967	35.82/0.9671
	SRCNN [26]	57 k	36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946	35.60/0.9663
	ESPCN [36]	58 k	37.00/0.9559	32.75/0.9098	31.51/0.8939	29.87/0.9065	36.21/0.9694
	FSRCNN [27]	12 k	37.06/0.9554	32.76/0.9078	31.53/0.8912	29.88/0.9024	36.67/0.9694
	VDSR [29]	665 k	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9750
	DRCN [30]	1.7 M	37.63/0.9584	33.06/0.9108	31.85/0.8947	30.76/0.9147	37.63/0.9723
	LapSRN [32]	813 k	37.52/0.9591	33.08/0.9109	31.80/0.8949	30.41/0.9112	37.27/0.9740
	MSRN [41]	4219 k	37.85/0.9599	33.40/0.9154	32.11/0.8988	31.71/0.9240	38.29/0.9761
	DRRN [37]	350 k	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.60/0.9736
	DADN [52]	3303 k	37.87/0.9600	33.39/0.9161	32.11/0.8988	31.77/0.9247	–/–
	EDSR [39]	43.7 M	38.11/0.9601	33.92/0.9195	32.32/0.9013	32.93/**0.9351**	**39.10**/0.9773
	MSRCAN [44]	4.4 M	37.90/0.9601	33.37/0.9158	32.10/0.8989	31.72/0.9242	–/–
	AMSRN [53]	380 k	37.89/0.9601	33.41/0.9163	32.15/0.8992	31.78/0.9249	38.54/0.9765
	IDSRN [56]	5.73 M	37.98/0.9605	33.56/0.9168	32.17/0.8992	31.82/0.9252	38.62/0.9768
	MSDRFN (Ours)	252 k	**38.49**/**0.9723**	**34.22**/**0.9267**	**32.94**/**0.9118**	**32.94**/0.9298	38.98/**0.9786**
×3	Bicubic	-/	30.39/0.8682	27.54/0.7736	27.21/0.7384	24.46/0.7344	26.96/0.8546
	A+ [50]	–/–	32.58/0.9088	29.13/0.8188	28.29/0.7835	26.03/0.7973	29.93/0.9089
	SelfExSR [47]	–/–	32.58/0.9093	29.16/0.8196	28.29/0.7840	26.44/0.8088	27.57/0.7997
	SRCNN [26]	57 k	32.75/0.9090	29.30/0.8215	28.41/0.7863	26.24/0.7989	30.48/0.9117
	ESPCN [36]	58 k	33.02/0.9135	29.49/0.8271	28.50/0.7937	26.41/0.8161	30.79/0.9181
	FSRCNN [27]	13 k	33.18/0.9140	29.37/0.8240	28.53/0.7910	26.43/0.8080	31.10/0.9210
	VDSR [29]	665 k	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9340
	DRCN [30]	1.7 M	33.85/0.9215	29.89/0.8317	28.81/0.7954	27.16/0.8311	32.31/0.9328
	LapSRN [32]	813 k	33.82/0.9227	29.79/0.8320	28.82/0.7973	27.07/0.8272	32.19/0.9334
	MSRN [41]	4219 k	34.14/0.9252	30.07/0.8368	28.98/0.8016	27.73/0.8426	32.96/0.9395
	DRRN [37]	350 k	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.42/0.9359
	DADN [52]	3303 k	34.17/0.9252	30.02/0.8370	28.98/0.8024	27.78/0.8434	–/–
	EDSR [39]	43.7 M	34.65/0.9282	30.52/0.8462	29.25/0.8093	28.79/**0.8655**	34.17/0.9476
	MSRCAN [44]	4.4 M	34.21/0.9255	30.09/0.8376	28.99/0.8020	27.72/0.8397	–/–
	AMSRN [53]	380 k	34.20/0.9254	30.10/0.8376	29.01/0.8025	27.80/0.8435	33.10/0.9417
	IDSRN [56]	5.73 M	34.25/0.9263	30.23/0.8391	29.05/0.8027	27.82/0.8443	33.18/0.9415
	MSDRFN (Ours)	252 k	**34.88**/**0.9322**	**30.61**/**0.8488**	**31.34**/**0.8241**	**29.33**/0.8572	**35.10**/**0.9469**
×4	Bicubic	–/–	28.42/0.8104	26.00/0.7019	25.96/0.6674	23.14/0.6570	24.91/0.7846
	A+ [50]	–/–	30.28/0.8603	27.32/0.7491	26.82/0.7087	24.32/0.7183	27.03/0.8439
	SelfExSR [47]	–/–	30.31/0.8619	27.40/0.7518	26.84/0.7106	24.79/0.7374	27.81/0.8594
	SRCNN [26]	57 k	30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221	27.65/0.8503
	ESPCN [36]	58 k	30.66/0.8646	27.71/0.7562	26.98/0.7124	24.60/0.7360	27.70/0.8560
	FSRCNN [27]	12 k	30.72/0.8660	27.61/0.7550	26.98/0.7150	24.62/0.7280	27.90/0.8610
	VDSR [29]	665 k	31.35/0.8830	28.02/0.7680	27.29/0.7251	25.18/0.7540	28.83/0.8870
	DRCN [30]	1.7 M	31.56/0.8810	28.15/0.7627	27.24/0.7150	25.25/0.7530	28.98/0.8816
	LapSRN [32]	813 k	31.54/0.8855	28.19/0.7720	27.32/0.7280	25.21/0.7553	29.09/0.8893
	MSRN [41]	4219 k	31.80/0.8903	28.29/0.7749	27.42/0.7307	25.65/0.7721	29.73/0.8978
	DRRN [37]	350 k	31.68/0.8888	28.21/0.7721	27.38/0.7284	25.44/0.7638	29.18/0.8914
	DADN [52]	3303 k	31.81/0.8890	28.29/0.7750	27.43/0.7312	25.71/0.7725	–/–
	EDSR [39]	43.7 M	32.46/0.8968	28.80/0.7876	27.71/0.7420	26.64/**0.8033**	31.02/0.9148
	MSRCAN [44]	4.4 M	31.89/0.8907	28.33/0.7753	27.45/0.7308	25.75/0.7733	–/–
	AMSRN [53]	380 k	31.94/0.8911	28.43/0.7763	27.50/0.7319	25.78/0.7731	29.81/0.8990
	IDSRN [56]	5.73 M	31.98/0.8920	28.45/0.7765	27.50/0.7338	25.76/0.7737	30.13/0.9026
	MSDRFN (Ours)	252 k	**32.92**/**0.9161**	**29.04**/**0.8204**	**29.05**/**0.7813**	**28.05**/0.8022	**32.76**/**0.9159**

where μ_x and $\mu_{\hat{x}}$ are respectively the means of images x and $\hat{x}$, σ_x and $\sigma_{\hat{x}}$ are respectively the variances of images x and $\hat{x}$. The constants C_1, C_2 and C_3 are used to avoid the problems incurred by zero denominator. High PSNR and SSIM values typically imply an effective SR reconstruction [5]. In the following discussion, the performances of the networks were evaluated in terms of these two indicators.

4.3. Ablation study

In this subsection, the effects of different network components on the performance of the proposed method were evaluated in ablation experiments with different numbers and sizes of convolution kernels. The details are presented below.

*Initial feature extraction module.*In the initial feature-extraction process, the features of input images were extracted by convolutions of constant width, and the size of the image was unchanged. As is well known, the number of convolution kernels determines the complexity of the network and the number of extracted features. In general, a network with many convolution kernels extracts more features than a network with few convolution kernels. To investigate this property, we set the number of convolution kernels in the initial feature extraction stage as 32, 64 and 128. The corresponding models were denoted as models 1, 2 and 3, respectively. The best performance models were highlighted in bold blue. The comparison results of the four models were shown in Table 1. Clearly, the model with 64 convolution kernels outperformed the model with 32 convolution kernels in the initial feature extraction stage, suggesting that increasing the number of convolution kernels improves the reconstruction performance. However, increasing the convolution kernel number from 64 to 128 actually decreased the reconstruction performance. Therefore, in further experiments, we set the initial feature extraction convolution kernel to 64.

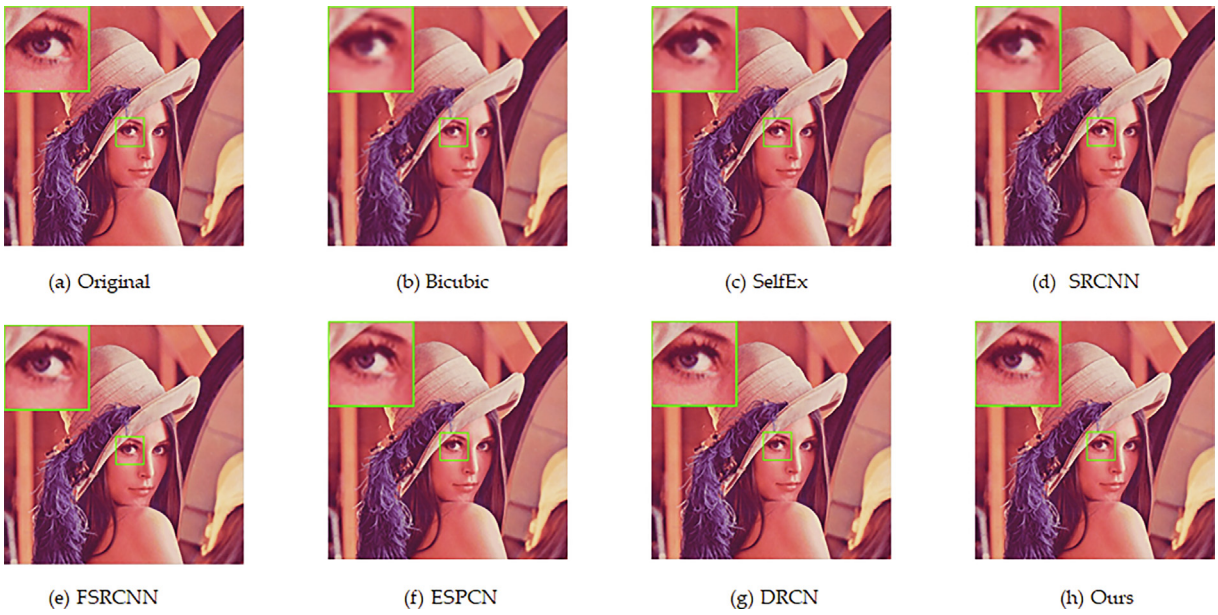


Fig. 8. Visual comparison of image SR constructions on Set14 (scale factor r = 2).

Fig. 9. Visual comparison of image SR constructions on Set5 (scale factor r = 3).

*Multi-scale dense fusion block.*From experience, multi-scale modules such as multi-scale feature extraction blocks obtain more abundant features than single-scale modules. To verify the effectiveness of the proposed MSDFB. Here we have done three sets of experiments. We replaced all convolution kernels in three branches of the MSDFB with 3×3 , 5×5or 7×7size kernels, forming models 4, 5 and 6, respectively. Table 2 shows the comparison results of the proposed algorithm and the above models. Among these models, the MSDRFN reconstruction using the multi-scale modules yielded the best results. This was reason why we combined the 3×3 , 5×5and 7×7convolution kernels as a multi-scale feature extraction block.

Number of feature extraction blocks. To explore the impact of multi-scale feature extraction blocks on the entire network, we changed the number of MSDF blocks to check the effect of super-resolution reconstruction experiments. The results of different numbers of MSDF blocks in Set5 were shown in Table 3 and Fig. 7. It is worth noting that the number of MSDF blocks did affect the reconstruction effect, and it is not difficult to find from the table that the model with 4 MSDF blocks worked best. In order to strike a balance between accuracy and resources, we chose to have four MSDF blocks as the final model.

Squeeze and excitation mechanism (SEM). To verify the role of the SEM in MSDF, we conducted a comparative experiment with the feature extraction block without the SEM. The results obtained on Set5 were shown in Table 4. Compared with the original MSDFR network, the PSNR of the network without the SEM was reduced by 0.15 dB or more. This clearly shows that the SEM can improve the reconstruction performance of our model.

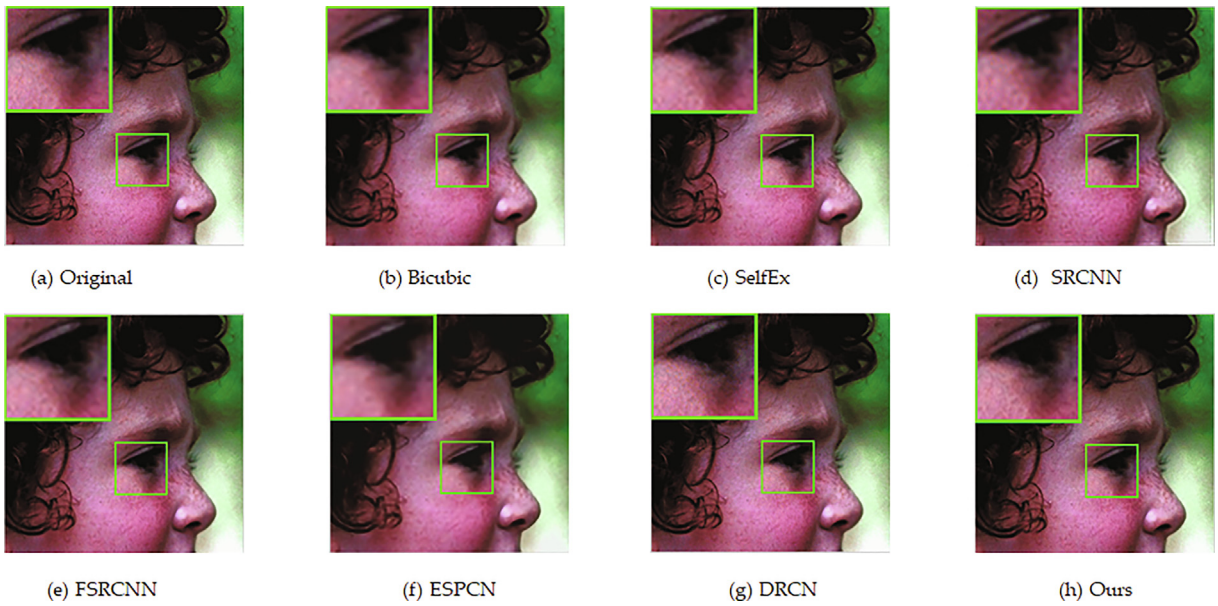


Fig. 10. Visual comparison of image SR constructions on Set14 (scale factor r = 3).

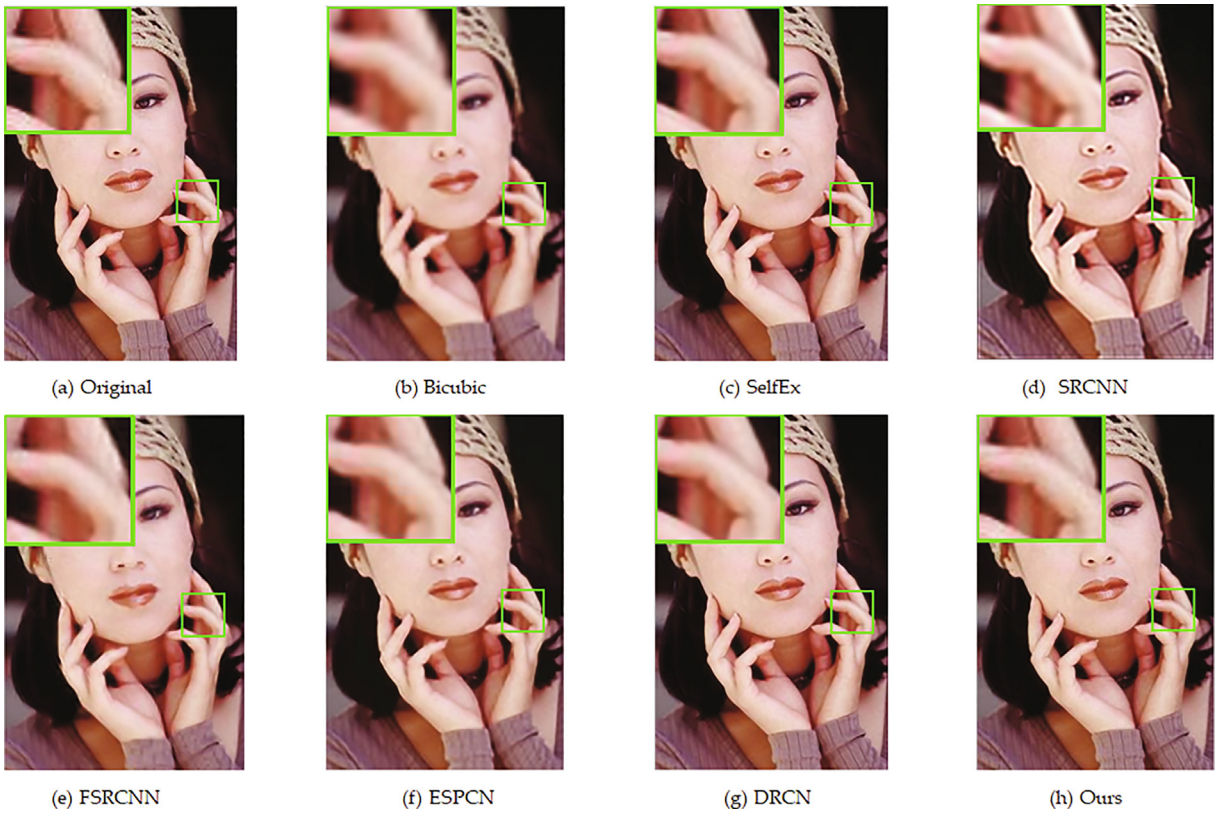


Fig. 11. Visual comparison of image SR constructions on Set5 (scale factors r = 4).

4.4. Model comparison

This subsection compared the performance of the proposed model with those of several existing SR models: Bicubic, A+ [50], SelfExSR [47], SRCNN [26], ESPCN [36], FSRCNN [27], VDSR [29], DRCN [30], LapSRN [32], MSRN [41], DRRN [37], DADN [52], EDSR [39], AMSRN [53], IDSRN [56]. Extensive experiments were carried out on four benchmark datasets: Set5 [25], Set14 [45], BSD100 [46], Urban100 [47] and Manga109 [49]. All reported research results of other models were taken directly from the published papers. The results for scale factors 2, 3, and 4 on the four benchmarks were presented in Table 5. The best performance models were highlighted in bold blue. The second-best performance model was shown in red. It is clear from Table 5 that our method achieved the best performance on these benchmark data sets, demonstrating its effective performance. To better show the reconstruction

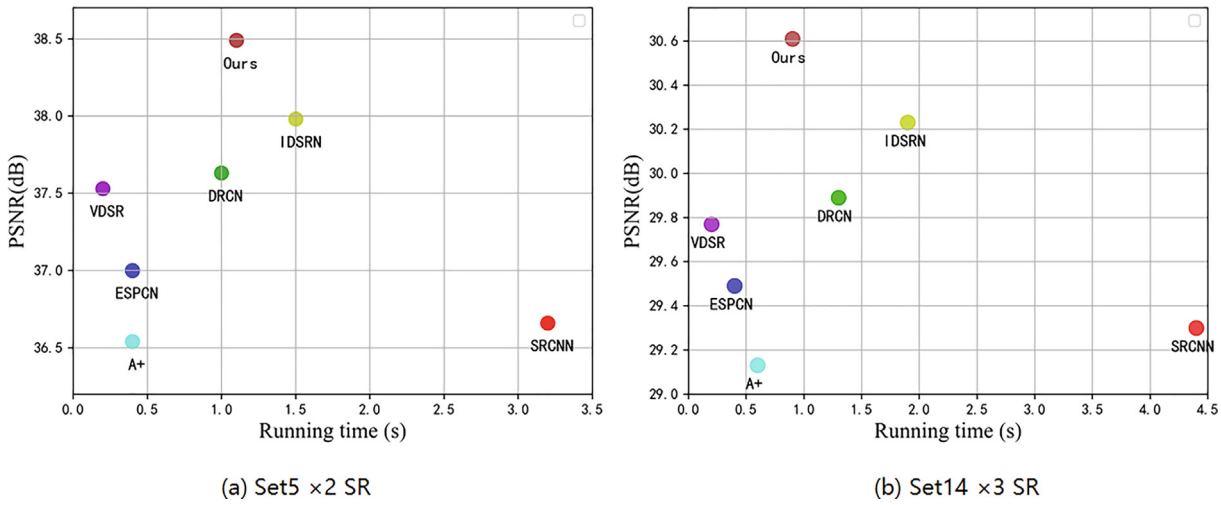


Fig. 12. Comparisons of the accuracy and speed on Set5 with ×2 SR, Set14 with ×3 SR for different approaches.

Table 6
FLOPs comparison of the state-of-the-art methods

Methods	SRCNN	FSRCNN	DRCN	VDSR	EDSR	Ours
×2	10.6 M	8.7 M	114.9 M	681.9 M	495.4 M	263.5 M
×3	14.6 M	8.7 M	196.3 M	961.6 M	687.4 M	330.2 M
×4	20.8 M	8.7 M	272.4 M	1255.7 M	894.3 M	425.2 M

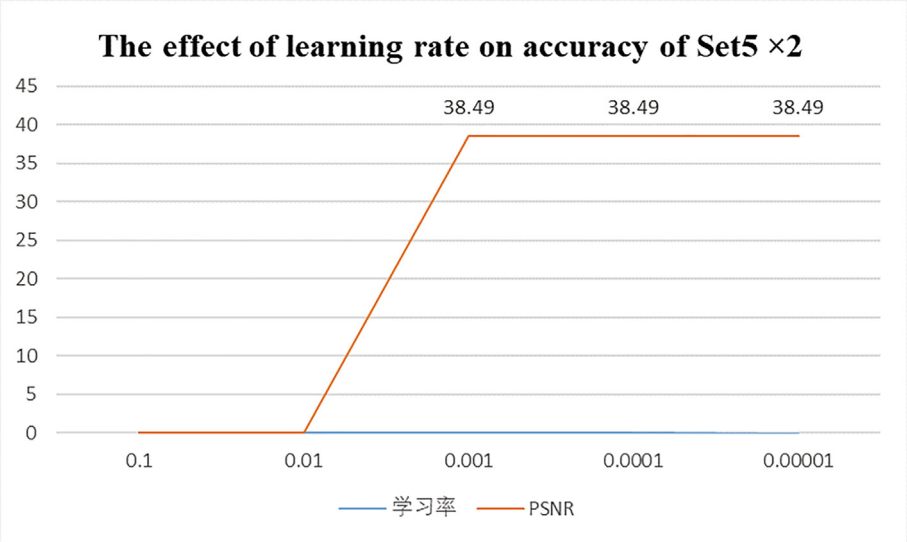


Fig. 13. The effect of learning rate on accuracy of Set5 ×2 SR.

performance of the proposed network, the visual effects of some models on some commonly used images were compared in Figs. 8–11. Clearly, the image reconstruction quality was higher in the proposed model than in most of the existing models. The higher reconstruction effect by the MSDRFN model can be attributed to suppressed distortion. In general, the sensory effect imparted by MSDRFN was comparable to those of the existing methods.

4.5. Execution time and FLOPs

To better illustrate the superiority of the proposed algorithm, this section compared the amount of network parameters, the running time and FLOPs between the proposed algorithm and some representative methods. The results about the amount of network parameters were shown in Table 5. The results on the running time and FLOPs were shown in Fig. 12 and Table 6. From Table 5, it is easily seen that the proposed algorithm has very lightweight parameters but best reconstruction performance compared with other representative algorithms, this shows that the network has less computation and lower complexity.

Fig. 12 shows the tradeoff between the execution time of Set5 ×2SR and Set14 ×3SR and the performance indicator PSNR. Points with different colors represent the average PSNR and running time of different methods for images in different data sets. Table 6

shows FLOPs comparison results of different methods. It can be clearly seen from the Fig. 12 and Table 6 that the proposed algorithm in this paper achieved high-quality reconstruction performance with relatively fast speed and low FLOPs, and was superior to other representative algorithms.

4.6. Parameter studies

In this section, we mainly discussed the influence of parameter learning rate and L1 regularization coefficient on the accuracy. In the experiment, we first set the learning rate to 10e-1, 10e-2, 10e-3, 10e-4, 10e-5, respectively. As can be seen from the experimental results (see Fig. 13), 10e-1 and 10e-2 are very ineffective because the gradient oscillates around the minimum and does not even converge when the learning rate is set too high. However, for 10e-3, 10e-4, 10e-5 there are the best performance and convergence, therefore, in this work, the learning rate takes the middle value 10e-4. For the L1 regularization coefficient, several experiments have been done, and it is found that the convergence speed is the fastest when set to 10e-4, and the convergence speed is slower when set to other values, but the results are consistent.

5. Conclusion

In this paper, we proposed a new multi-scale dense recursive fusion network (MSDRFN) for super-resolution image reconstruction, which improves the performance of existing CNN-based SR methods without consuming too much computing resources and time. The advantages of this method are that the convolution kernels of different scales are fully utilized and multi-scale dense fusion block (MSDFB) is embedded into the network to capture the image information of different scales comprehensively. On the basis of multi-scale feature extraction, channel attention mechanism is added to play a role in information flow between networks. In addition, the extracted features of MSDFBs are processed hierarchically, and the high-resolution images are recovered using sub-pixel Convolutional Recursive-Reconstruction Net (SpCRNet) respectively. Finally, residual learning is used to fuse the hierarchical features, and the resulting feature map is added to the original input image of bicubic interpolation to reconstruct the final high-resolution image.

At present, most methods based on CNN focus on training complex and deep networks. However, the proposed method in this paper is a relatively shallow and simple algorithm. The proposed method can not only extract image features of different scales, but also use residual learning and hierarchical feature processing to supplement information loss, and combine detail features and contour features to form rich features.

In extensive experiments and analyses on five benchmark datasets, the proposed network demonstrated high reconstruction performance both qualitatively and visually. Although our model is both efficient and accurate, its performance could be still improved in several ways. For example, how to realize the weight share of the four feature extraction blocks to reduce the number of model parameters without reducing the SR reconstruction performance remains an open problem. In future work, we will seek more general and effective models for image reconstruction. We will also explore effective and practical models for applications other than SR tasks.

CRediT authorship contribution statement

Xiang Lv: Formal analysis, Writing - original draft. **Changzhong Wang:** Supervision, Methodology. **Xiaodong Fan:** Conceptualization, Investigation. **Qiangkui Leng:** Validation, Software. **Xiaoli Jiang:** Writing - review & editing.

Declaration of Competing 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.

Acknowledgement

This work was supported by National Natural Science Foundation of China under Grants 61976027, 61572082, Liaoning Revitalization Talents Program (XLYC2008002) and Liaoning Provincial Education Department Project (LJKZ1026, LJKZ1030). We also would like to express our sincere gratitude to the editor and anonymous reviewers for their valuable comments, which have greatly improved this paper.

References

[1] S. Borman, R.L. Stevenson, Super-resolution from image sequences CA review, Proc. Midwest Symp. on Circuits and Systems, Notre Dame (1998) 374–378.
[2] S. Farsiu, D. Robinson, M. Elad, P. Milanfar, Advances and challenges in super resolution, Int. J. Imaging Syst. Technol. 14 (2) (2010) 47–57.
[3] K. Jiang, Z. Wang, P. Yi, et al., A Progressively Enhanced Network for Video Satellite Imagery Super-Resolution, IEEE Signal Process. Lett. 25 (11) (2018) 1630–1634.
[4] J. Zhao, C. Chen, Z. Zhou, et al., Single image super-resolution based on adaptive convolutional sparse coding and convolutional neural networks, J. Vis. Commun. Image Represent 58 (1) (2019) 651–661.
[5] F. Cao, B. Chen, New architecture of deep recursive convolution networks for super-resolution, Knowl.-Based Syst. 178 (2019) 98–110.
[6] F. Cao, M. Cai, Y. Tan, Image interpolation via low-rank matrix completion and recovery, IEEE Trans. Circuits Syst. Video Technol. 25 (8) (2015) 1261–1270.
[7] M. Bevilacqua, A. Roumy, C. Guillemot, M.-L.A. Morel, Single-image super-resolution via linear mapping of interpolated self-examples, IEEE Trans. Image Process. 23 (12) (2014) 5334–5347.
[8] T. Blu, P. Thvenaz, M. Unser, Linear interpolation revitalized, IEEE Trans. Image Process. 13 (6) (2004) 710–719.
[9] K. Hung, W. Siu, Robust soft-decision interpolation using weighted least squares, IEEE Trans. Image Process. 21 (3) (2012) 1061–1069.
[10] R. Keys, Cubic convolution interpolation for digital image processing, IEEE Trans. Acoust. Speech Signal Process. 29 (6) (1981) 1153–1160.
[11] T. Lehmann, C. Gonner, K. Spitzer, Addendum: B-spline interpolation in medical image processing, IEEE Trans. Med. Imaging 20 (7) (2001) 660–665.
[12] X. Li, M. Orchard, New edge-directed interpolation, IEEE Trans. Image Process. 10 (10) (2001) 1521–1527.
[13] W. Tam, C. Kok, W. Siu, Modified edge-directed interpolation for images. J. Electron. Imaging, 19(1) (2010) 013011-1-013011-20.
[14] L. Zhang, X. Wu, An edge-guided image interpolation algorithm via directional filtering and data fusion, IEEE Trans. Image Process. 15 (8) (2006) 2226–2238.
[15] X. Zhang, X. Wu, Image interpolation by adaptive 2-d auto-regressive modeling and soft-decision estimation, IEEE Trans. Image Process. 17 (6) (2008) 887–896.
[16] S. Dai, M. Han, W. Xu, et al., Softcuts: A soft edge smoothness prior for color image super-resolution, IEEE Trans. Image Process. 18 (5) (2009) 969–981.
[17] K. Hung, W. Siu, Single image super-resolution using iterative wiener filter. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 2012, pp. 1269-1272.
[18] A. Marquina, S. Osher, Image super-resolution by TV-Regularization and Bregman iteration, J. Sci. Comput. 37 (3) (2008) 367–382.
[19] M. Protter, M. Elad, H. Takeda, et al., Generalizing the nonlocal-means to super-resolution reconstruction, IEEE Trans. Image Process. 18 (1) (2009) 36–51.
[20] J. Sun, J. Sun, Z. Xu, H. Shum, Gradient profile prior and its applications in image super-resolution and enhancement, IEEE Trans. Image Process. 20 (6) (2011) 1529–1542.
[21] L. Wang, S. Xiang, G. Meng, et al., Edge-directed single-image super-resolution via adaptive gradient magnitude self-interpolation, IEEE Trans. Circuits Syst. Video Technol. 23 (8) (2013) 1289–1299.
[22] H. Xu, G. Zhai, X. Yang, Single image super-resolution with detail enhancement based on local fractal analysis of gradient, IEEE Trans. Circuits Syst. Video Technol. 23 (10) (2013) 1740–1754.
[23] K. Zhang, X. Gao, D. Tao, X. Li, Single image super-resolution with non-local means and steering kernel regression, IEEE Trans. Image Process. 21 (11) (2012) 4544–4556.
[24] K. Zhang, X. Gao, D. Tao, X. Li, Single image super-resolution with [24]multiscale similarity learning, IEEE Trans. Neural Networks Learn. Syst. 24 (10) (2013) 1648–1659.
[25] M. Bevilacqua, A. Roumy, C. Guillemot, et al., Low-Complexity Single Image Super-Resolution Based on Nonnegative Neighbor Embedding, in: Proc. British Machine Vision Conference (BMVC), 2012, pp. 1–10.

[26] C. Dong, C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: Proc. European Conference on Computer Vision (ECCV), 2014, pp. 184–199.
[27] C. Dong, C.C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, in: Proc. European Conference on Computer Vision (ECCV), 2016, pp. 391–407.
[28] C. Dong, C.C. Loy, K. He, X. Tang, Image super resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2016) 295–307.
[29] J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2016, pp. 1646–1654.
[30] J. Kim, J.K. Lee, K.M. Lee, Deeply-recursive convolutional network for image super-resolution, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2016, pp. 1637–1645.
[31] Y. Hu, X. Gao, J. Li, et al., Single Image Super-Resolution via Cascaded Multi-Scale Cross Network, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2018.
[32] W. Lai, J. Huang, N. Ahuja, M. Yang, Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks, IEEE Trans. Pattern Anal. Mach. Intell. (2017).
[33] W. Lai, J. Huang, N. Ahuja, M. Yang, Deep Laplacian pyramid networks for fast and accurate super resolution, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2017, pp. 624–632.
[34] W. Han, S. Chang, D. Liu, M. Yu, M. Witbrock, Image super-resolution via dual-state recurrent networks, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2018, pp. 1654–1663.
[35] Y. Zhang, K. LI, et al., Image Super-Resolution Using Very Deep Residual Channel Attention Networks, European Conference on Computer Vision (ECCV) (2018).
[36] W. Shi, J. Caballero, F. Huszr, et al., Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2016, pp. 1874–1883.
[37] T. Ying, Y. Jian, X. Liu, Image super-resolution via deep recursive residual network, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition CVPR, 2017, pp. 2790–2798.
[38] C. Ledig, L. Theis, F. Huszar, Caballero, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2017, pp. 105–114.
[39] B. Lim, S. Son, H. Kim, et al., Enhanced deep residual networks for single image super resolution, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2017, pp. 1132–1140.
[40] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2001, pp. 416–423.
[41] J. Li, F. Fang, K. Mei, G. Zhang, Multi-scale residual network for image super-resolution, in: Proc. Proceedings of European Conference on Computer Vision (ECCV), 2018, pp. 527–542.
[42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2016, pp. 770–778.
[43] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2018, pp. 2472–2481.
[44] F. Cao, H. Liu, et al., Single Image Super-resolution via Multi-scale Residual Channel Attention Network, Neurocomputing 358 (09) (2019) 424–436.
[45] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse representations. International conference on Curves and Surfaces. Springer, Berlin, Heidelberg, 2010. pp. 711-730.
[46] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2001, pp. 416–423.
[47] J. Huang, A. Singh, N. Ahuja, Single image super-resolution from transformed self-exemplars, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2015, pp. 5197–5206.
[48] D. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, Computer Science, 2014.
[49] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, et al., Sketch-based Manga Retrieval using Manga109 Dataset, Multimedia Tools Appl. 76 (20) (2017) 21811–21838.
[50] R. Timofte, V.D. Smet, L.V. Gool, A +: adjusted anchored neighborhood regression for fast super-resolution, Proc. Asian Conference on Computer Vision (ACCV) (2014) 111–126.
[51] M. Haris, G. Shakhnarovich, N. Ukita, Deep Back-Projection Networks For Super-Resolution, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2018, pp. 1664–1673.
[52] J. Qin, Z. Xie, Y. Shi, W. Wen, Difficulty-aware image super resolution via deep adaptive dual-network, in: 2019 IEEE International Conference on Multimedia and Expo (ICME), 2019, pp. 586–591.
[53] H. Liu, F. Cao, C. Wen, et al., Lightweight multi-scale residual networks with attention for image super-resolution, Knowl.-Based Syst. 203 (4) (2020) 106103.
[54] Z. Wang, A. Bovik, A universal image quality index, IEEE Signal Process Lett. 9 (3) (2002) 81–84.
[55] Z. Wang, A. Bovik, H.R. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[56] H. Liu, F. Cao, Improved dual-scale residual network for image super-resolution, Neural Networks 132 (8) (2020) 84–95.
[57] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2018, pp. 7132–7141.
[58] Y. Zhang, Y. Tian, Y. Kong, et al., Residual densenet work for image super-resolution, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2018.
[59] J. Liu, W. Zhang, Y. Tang, et al., Residual Feature Aggregation Network for Image Super-Resolution, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2020.
[60] L. Lu, W. Li, X. Tao, et al., MASA-SR: Matching Acceleration and Spatial Adaptation for Reference-Based Image Super-Resolution, in: Proc. IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2021.
[61] W. Li, K. Zhou, L. Qi, et al. Best-Buddy GANs for Highly Detailed Image Super-Resolution, 2021.
[62] Z. Lu, H. Liu, J. Li, et al. Efficient Transformer for Single Image Super-Resolution, 2021.
[63] U.M. Rao, A. Munir, C. Micheloni. A Deep Residual Star Generative Adversarial Network for multi-domain Image Super-Resolution, 2021.
[64] J. Ahn, N. Cho, Neural Architecture Search for Image Super-Resolution Using Densely Constructed Search Space: DeCoNAS[C]In: Proc, in: International Conference on Pattern Recognition (ICPR), 2021.
[65] A. Villar-Corrales, F. Schirrmacher, C. Riess. Deep learning architectural designs for super-resolution of noisy images, 2021.
[66] T. Tran, J. Berberich, S. Simon. 3DVSR: 3D EPI Volume-based Approach for Angular and Spatial Light field Image Super-resolution, 2022.
[67] L. Sun, Z. Liu, et al. Lightweight Image Super-Resolution via Weighted Multi-Scale Residual Network. IEEE/CAA J. Automatica Sinica pp. 99(2021):1–10.
[68] Y. Liu Y, Q. Jia,X. Fan X, et al. Cross-SRN: Structure-Preserving Super-Resolution Network with Cross Convolution, 2022.

Xiang Lv received the B.Sc. degree in mathematics and Applied Mathematics from Bohai University, Jinzhou, China, in 2020. She is currently pursuing the M.Sc. degree in applied mathematics with Bohai University. Her current research interests include deep learning, computer vision and image processing.

Changzhong Wang received the M.S. degree from Bohai University, Jinzhou, China, the Ph.D. degree from Harbin Institute of Technology, Harbin, China, in 2005, and 2008 respectively. He is currently a Professor with Bohai University. His research interests are focused on granular computing, machine learning and deep learning. He has authored or coauthored more than 50 journal and conference papers in the areas of machine learning, data mining, and image processing.

Xiaodong Fan received the M.S. degree from Harbin University of Commerce in 2000, the M.S. degree in Mathematics from Guizhou University in 2007 and the Ph.D. degrees in Mathematics from Beijing University of Technology in 2013, respectively. He is currently an associate Professor with Bohai University. His research interests are focused on machine learning and optimization.

Qiangkui Leng received the B.S. and M.S. degrees in Computer Science from Bohai University in 2004 and 2011 respectively, and the Ph.D. degree in Computer Science from Beijing University of Technology. He is currently a Professor with Liaoning Technical University. His research interests include machine learning and pattern recognition.

Xiaoli Jiang was born in Suihua City, Heilongjiang, in 1975, in P. R. China.In 2012, she received the M.S. degree in College of Mathematics and Physics from Harbin Engineering University, and she received the Ph.D. degree in System Engineering from Harbin Engineering University in 2014. At present, she is a lecturer in College of Mathematics and Physics in Bohai University in Applied Mathematics in P. R. China. Her research interests include image processing, machine learning, system control decision and analysis.