

## Research Article

# Image Super-Resolution Network Based on Feature Fusion Attention

Changjun Zou  and Lintao Ye

East China Jiaotong University, Nanchang 330013, China

Correspondence should be addressed to Changjun Zou; [zoucj2006@163.com](mailto:zoucj2006@163.com)

Received 30 September 2022; Revised 24 November 2022; Accepted 28 November 2022; Published 14 December 2022

Academic Editor: Jiacheng Yang

Copyright © 2022 Changjun Zou and Lintao Ye. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The residual structure may learn the entire input region indiscriminately because the residual connection can still learn well as the network depth grows. To a certain extent, the attention mechanism can focus the network's attention to the interesting area, enhancing the learning performance of essential areas while decreasing the computational load for the system. As a result, the combination of these two advantages could have substantial research significance, for both improve the efficiency and reduce the computational load. A dense residual connection network that combine feature fusion attention approach in image super resolution process is proposed. The dense residual block is enhanced with pixel and channel attention blocks, and a dual-channel path design incorporating global maximum pooling and global average pooling is utilized. A hybrid loss function is also proposed in order to increase the network's sensitivity to the maximum error between individual pixels. The PSNR/SSIM/ $L_{\infty}$  performance metrics increased after applying the hybrid loss function and our attention techniques. The experimental results demonstrated that our novel approach has several advantages over some recent approaches, as well as showing good outcomes on many testing datasets.

## 1. Introduction

Image super-resolution (SR) is an important task in the field of computer vision [1–5]. In SR process, high-resolution (HR) images can be reconstructed from low-resolution (LR) images, which can break through the resolution limit of the original image. So it can be applied to fields such as enhancing LR film and television works, intelligent monitoring, video processing, HR film, and television production.

In recent years, with the development of convolution neural network (CNN) and deep learning methods, the research of image super-resolution based on deep learning has gradually become a hot spot. The method based on deep learning is good at dealing with this kind of nonlinear problems. With continuous development, convolution neural network is more and more widely used in the field of image processing. Among them, SRCNN [6, 7] proposed a super-resolution network model based on 3 convolution layers, which mainly achieve image feature extraction, representation, feature nonlinear mapping, and final reconstruction. The network's depth is so

constrained because it only consists of 3 convolution layers connected in series. Building a deeper deep learning network is made possible until the residual network [8] emerge. For instance, VDSR [9] has built a deeper super-resolution network combined with residual module [8, 10, 11], and the network performance has also been greatly improved. Since then, residual connections and deeper networks have become one of the hottest areas of super-resolution [12], while dense residual connections [13–16] and cascaded [17] structures with multiple skip connections are becoming more and more popular. This kind of method allows reusing the feature of various network layers, pushing residual information and gradient information to propagate more quickly, and enhancing their performance.

Among them, SRDenseNet [13] is mainly constructed by DenseNet [15] module. And DenseNet not only greatly reduces the amount of network parameters but also alleviates the problem of gradient disappearance to a certain extent by reusing the feature map of shallow networks and bypass the redundant features. This architecture makes full use of the

features extracted from the shallow network and is reused in the deeper layer. The improvement of RDN [16] mainly introduces continuous memory mechanism (CM), global connection, and local residual connection to increase residual propagation. SRResNet [18] mainly removes redundant modules in SRResNet, such as batch normalization (BN) compared with EDSR [19], which can expand the depth of the model to improve the quality of the reconstructed image.

In recently, there are also many attractive methods such as HAN [20] based on heterogeneous graph, RFANet [21] based on residual feature aggregation, and DeFiAN [22] with extra Hessian filtering, they all take advantages of attention mechanism. For SwinIR [23] with Swin Transformer, it is hard to transfer feature information between neighbor window, which may limit the application to some extent. Moreover, in the feature learning process of the traditional deep learning network, with nonlinear mapping such as convolution and activation function alone the network cannot make full use of the high-frequency information and makes it hard to restore the detailed feature. Adding attention mechanism to the network improve the ability of extracting high-frequency information. In recent years, attention mechanism is widely used in image processing. Attention mechanism includes channel attention mechanism and spatial attention mechanism. Channel attention is used to capture the correlation between different channel feature of the model.

After different convolution kernels, each channel will generate new weight information. The convolution kernel output is decomposed into multichannel information components. The attention mechanism assigns each weight value learned to the corresponding channel. The problem of information overload can be resolved elegantly by introducing the attention mechanism, which allows us to concentrate on the information that is more important for the current task among the many inputs, focus the system's attention less on other information, and even filter out unimportant information. At the same time, we can enhance the task's accuracy and processing efficiency. Feature attention is popular in many disciplines, including text image [24], CT image [25] or medical image [26], binocular camera image [27], 3D remote sensing image [28], infrared image [29], and video [30].

Additionally, there is a wealth of study in the fields relevant to attention technology. Such as content attention [31], grid attention [32], spatial attention [33], hierarchical attention [34], shuffle [35], multilevel feature enhancement [36], and residual attention [37].

A comprehensive feature attention system has not yet been developed among these existing attention techniques. Therefore, an integrated multichannel attention network based on channel attention, pixel attention, and residual attention to realize the super-resolution of images and further recover the detailed feature in images is proposed. For this paper, our main contributions are as follows:

- (1) A dense residual connection network based on feature attention mechanism is proposed, including pixel attention and channel attention. Combined with the residual scaling dense connection module

and the new hybrid loss function proposed by us, our new method achieved significant performance improvement especially for the  $L_{\infty}$  error.

- (2) We proposed and verified a residual scaling dense block (RSDB) based on residual scaling. Compared with traditional residual dense block (RDB), our new method incorporates residual scaling layer and feature attention layer to further improve the performance. The experimental results also show that our method has better residual convergence speed.
- (3) A new hybrid loss function  $L_H$  is proposed and verified, which can effectively reduce the  $L_{\infty}$  error of the network. Compared with the traditional  $L_1$  loss function or  $L_2$  loss function, this method cannot only improve the PSNR and SSIM performance but also greatly improve the  $L_{\infty}$  error performance and can be easily adjusted according to user requirements, which is of great value in  $L_{\infty}$  sensitive occasions.

As we know, the attention mechanism can focus the network's attention to the relevant area, enhancing the learning performance of essential areas while decreasing the computational load on the network. The channel attention model can automatically learn the importance of each feature channel and assign different weights to each feature channel. Bigger weights will be assigned to it if the feature information is important, and vice versa. Such requirement is essential for the automobile system in autodiving application. As a result, the combination of the two approaches have substantial research significance. The experimental result infers that the residual dense connection equipped with pixel attention and channel attention could achieve positive influence on the final results.

## 2. Our Method

*2.1. RDAN and the Basic Block.* Attention mechanism in neural network is a resource optimization allocation scheme that assigns computing resources to more important tasks first and solves the problem of information overload when computing resources are limited, especially for the mobile vehicle systems [38, 39]. Generally speaking, the more parameters of the model in the deep learning network, the stronger the expression ability of the model, and the greater the amount of information stored in the model, but this may bring the problem of information overload.

We can focus on the more critical information in the current task among the many input information by incorporating the attention mechanism and even filter out irrelevant information, so as to solve the problem of information overload and improve the efficiency and accuracy of task processing. The attention mechanism has been widely used to tasks like image segmentation and super-resolution in the field of computer vision. All of the visual attention models in use today are spatial, and typically, the feature map of the final convolution layer is weighted. We proposed a model called residual dense attention network (RDAN) based on multichannel feature fusion attention. Figure 1

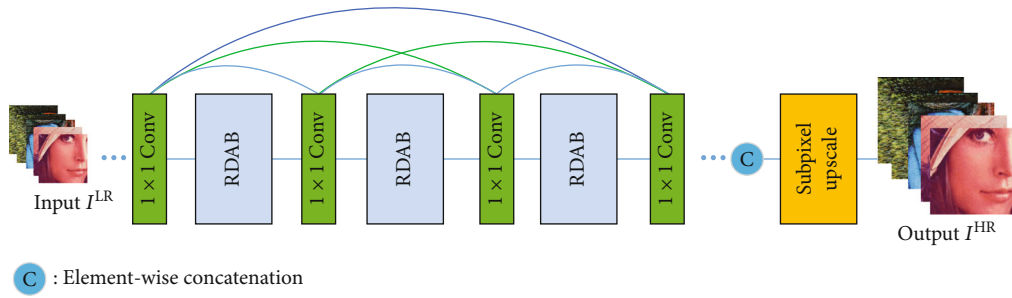


FIGURE 1: Schematic diagram of RDAN feature attention fusion network structure. Our RDAN is constructed by dense connection of residual dense attention blocks (RDAB). The RDAB is composed of the basic block in dense connection. The basic block is equipped with channel attention and pixel attention in sequence. The detailed structure is shown in Figures 2 and 3.

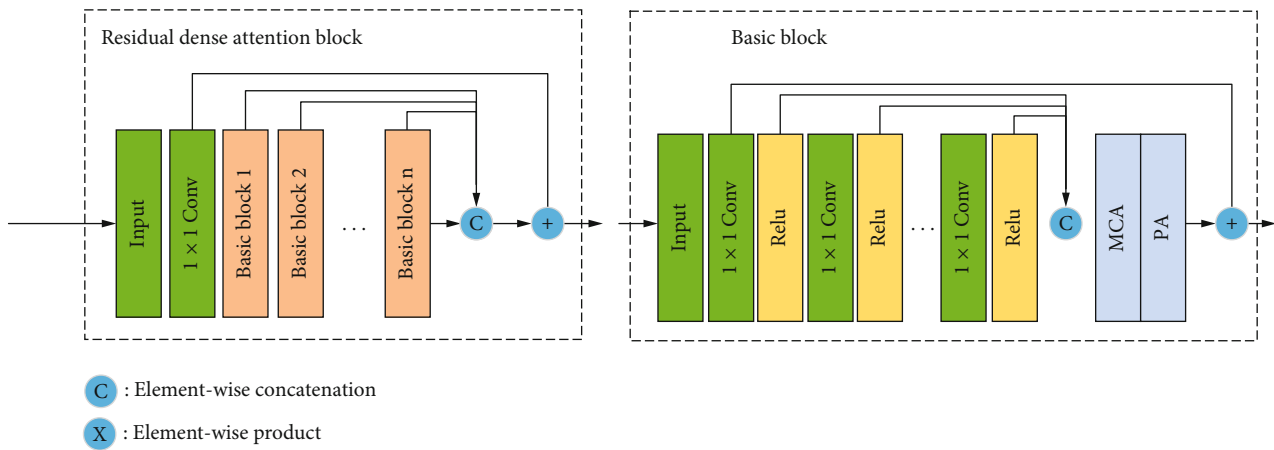


FIGURE 2: Composition of the basic block. The residual dense attention block is shown on the left side. The RDAB contains several basic block connecting in dense. The basic block is depicted on the right side. The basic block is equipped with multichannel attention and pixel attention structure. The basic block is the fundamental component of our RDABs.

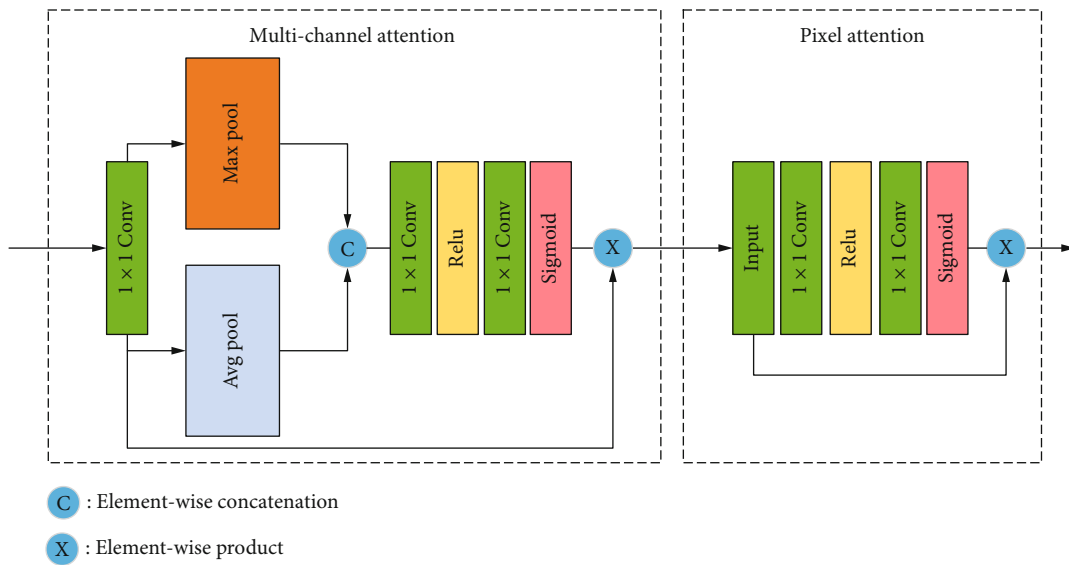


FIGURE 3: Multipath channel attention and pixel attention modules. The multichannel attention block is on the left side. The input is divided into two branches after the first convolution layer. The feature flows into two different pooling channels and then concatenated and activated before flow out this block. While the pixel attention module is on the right side, the PA module has two convolution layers and two activation layer connecting in alternative orders.

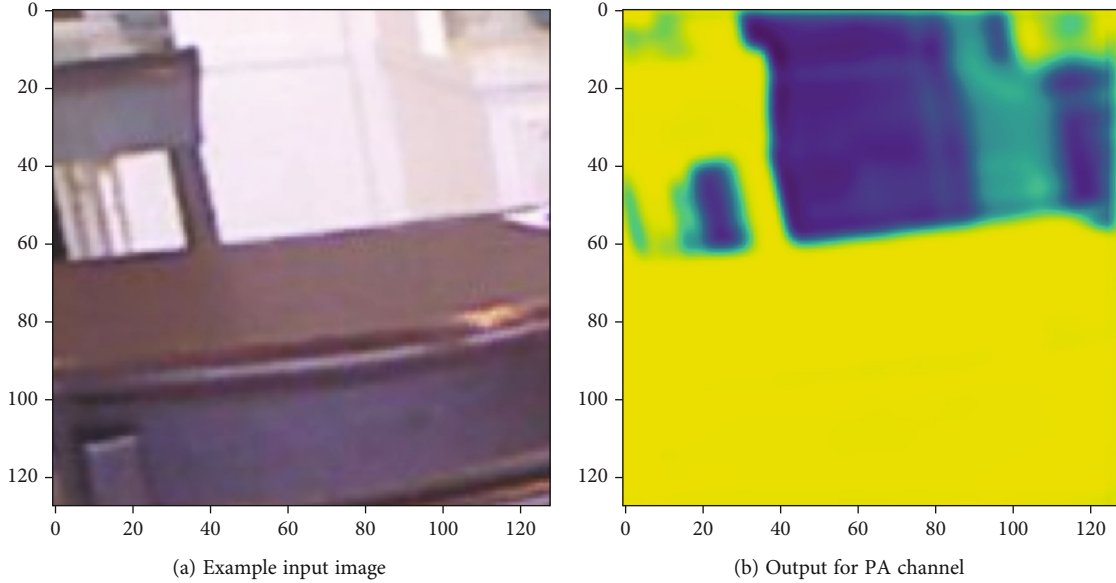


FIGURE 4: Input and pixel attention channel weight (a). As we can see, the foreground is a desk with higher weight on the output (b).

depicts the method's structural breakdown. The multichannel attention module and the pixel attention module can be combined with RDN to restore the detailed feature information.

RDAN is mainly composed of residual dense attention block (RDAB) through dense residual connection. RDAB is composed of basic blocks (BB) through dense connection, and pixel-wise connection is utilized in the dense connection. The specific structure is shown in Figure 2. Combining the dense residual connection module, our network can allow the gradient and feature information propagate more fluently and further improve the network performance.

The basic block is composed of a residual dense block (RDB), multipath channel attention (MCA), and pixel attention (PA) blocks; the attention network adds residual connections after MCA and PA to speed up the propagation of residual and gradient information, as can be seen in Figures 2 and 3. Local residual learning and feature attention modules are part of a basic block structure. Local residual learning modules enable it to skip through less essential area, such as low-frequency regions. The backbone network architecture can focus on more effective critical information thanks to the various local residual connections.

*MCA.* In the previous channel connection attention module [24], average pooling was used to calculate channel weight. But, average pooling tends to smooth edges and ignore detailed information. Hence, we added another maximum pooling path to our multipath channel attention module in order to restore more detailed information. The convolution and activation layers are appended after the average and maximum pooling layers. This strategy enables more efficient detail acquisition and, as a result, increased learning efficiency.

*PA.* The pixel attention module uses a convolution and activation layer in sequence, as well as pixel-wise weighting and residual multiplication. The network treats distinct features and pixels unequally in the attention module, allowing

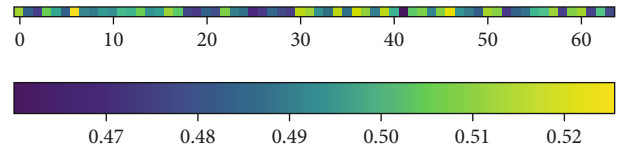


FIGURE 5: Multipath channel attention weight.

for more flexibility in processing diverse types of data and extending the representation potential of convolution neural networks. The feature weights in the features fusion attention (FFA) layer of multiple levels of features structure are automatically learned from the feature attention (FA) module, which assign essential features higher weights. This structure can also store and pass on shallow data to deeper layers.

*2.2. Multipath Channel Attention Module.* Based on the input feature information, the MCA module primarily determines the weights of each channel. Not only global average weights, but also global maximum pooling weights are produced in this module. This method allows for the collection of more specific information. The input and output images in the attention module have dimensions of  $C \times H \times W$  and  $C \times 1 \times 1$ , respectively. After applying global average and maximum pooling to the input data, we can obtain

$$g_a = H_{ap}(F_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j), \quad (1)$$

$$g_m = H_{mp}(F_c) = \text{Max}[X_c(i, j)], \quad (2)$$

where  $H_{ap}$  and  $H_{mp}$  are global average pooling and global max pooling functions, respectively;  $F_c$  is the input;  $X_c$  is

```

Input: Input image, block number, kernel size
Output: Output feature image
1.  $x = \text{conv2d}(\text{input})$ 
2.  $vx_i.\text{clear}(\quad)$ 
3. For  $i = 1$  to  $\text{count} + 1$  do
4.    $vx_i.\text{push}(x)$ 
5.    $\text{out put} = \text{Concat}(vx_i)$ 
6.    $\text{out put} = \text{Relu}(\text{conv2d}(\text{output}))$ 
7. End for
8.  $\text{out put} = \text{Relu}(\text{conv2d}(\text{output}))$ 
9.  $\text{out put} = \text{out put} * \text{scale\_rate}$  {residual scaling}
10.  $\text{out put} = \text{MCA}(\text{out put})$  {Multi-Channel Channel attention layer}
11.  $\text{out put} = \text{PA}(\text{out put})$  {pixel attention layer}
12.  $\text{out put} = \text{Add}(x, \text{out put})$  {residual connection}

```

ALGORITHM 1: Residual dense attention block.

the input value;  $H$  and  $W$  are the height and width dimensions of the input image, respectively. Then, after pooling and point-wise splicing, it goes to the convolution layer and activation layer, we get

$$CA_c = \sigma[\text{Conv}[\delta[\text{Conv}(\text{Concat}(g_a, g_m))]]], \quad (3)$$

where  $\sigma$  and  $\delta$  are the ReLU and Sigmoid activation functions, respectively; Concat is the pixel-wise splicing, and Conv is the  $1 \times 1$  convolution. Finally, the weights and the inputs are multiplied pixel-wise to obtain the final channel-aware output.

$$F_c^* = F_c \otimes CA_c, \quad (4)$$

where  $\otimes$  is the point-wise multiplication, and the detailed structure is shown in Figure 3.

**2.3. Pixel Attention Module.** The pixel attention module concentrates on the pixel weight. The pixel attention module detects weight distribution over the entire image and apply this information to execute targeted weight calculations. The input-output image shape of the pixel attention module changes from  $C \times H \times W$  to  $3 \times H \times W$ . This output also contains the RGB per color component weights; when utilizing the MCA output as the PA input, we get

$$PA = \sigma[\text{Conv}[\delta[\text{Conv}(F_c^*)]]], \quad (5)$$

where  $F_c^*$  is the output of the MCA;  $\sigma$  and  $\delta$  are the ReLU and Sigmoid activation functions, respectively; Conv is the  $1 \times 1$  convolution. Finally, the output of the final pixel-aware module is obtained after multiplying the weights with the input pixel by pixel as follows.

$$F = F^* \otimes PA. \quad (6)$$

Multipath channel attention weight and pixel attention weight example are shown in Figures 4 and 5.

**2.4. Hybrid Loss Function.** RDAN network adopts a hybrid loss function composed of  $L_1$  error and  $L_\infty$  error, which

TABLE 1: Training platform and related parameters.

Item	Parameters
CPU	Intel i9 12900K
GPU	Nvidia RTX3080
CPU memory	64G, DDR5,4800MHZ
GPU memory	GDDR6 12G
Operation system	Windows 10
Deep learning platform	Keras 2.7 with python 3.7

improves the instability of  $L_\infty$  error of the original RDN network. The  $L_\infty$  error is significantly decreased in addition to the PSNR/SSIM performance improvement, demonstrating the evident benefits of the enhanced network. We can refer to the test comparative analysis for further information.

**2.4.1.  $L_1$ -Norm.**  $L_1$ -norm is one of the most common norms, which is defined as follows:

$$\|x\| = \sum_{i=1}^n |x_i|. \quad (7)$$

The  $L_1$ -norm can be used to measure the difference between two vectors, such as the mean absolute error (MAE)

$$\text{MAE}(x_1, x_2) = \frac{\sum_{i=1}^n |x_{1i} - x_{2i}|}{n}. \quad (8)$$

**2.4.2.  $L_\infty$ -Norm.**  $L_\infty$ -norm is mainly used to measure the maximum value of vectors. It is defined as

$$\|x\|_\infty = \sqrt{\sum_i x_i^{\infty}}, (x = x_1, x_2, \dots, x_n). \quad (9)$$

In general, it can be expressed by the following formula:

$$\|x\|_\infty = \max(|x_i|). \quad (10)$$

A very good feature of  $L_\infty$ -norm is that it is independent



TABLE 2: Benchmark tests results, average PSNR/SSIM, bold is the best result, and italic is the second best result.

Method	Set14			BSD100			Urban100		
	X2	X3	X4	X2	X3	X4	X2	X3	X4
Bicubic	30.240	27.550	26.000	29.560	27.210	25.960	26.880	24.460	23.140
	0.869	0.774	0.703	0.843	0.739	0.668	0.840	0.735	0.658
SRCNN	32.450	29.300	27.500	31.360	28.410	26.900	29.500	26.240	24.520
	0.907	0.822	0.751	0.888	0.786	0.710	0.895	0.799	0.722
LapSRN	33.080	29.790	28.190	31.800	28.820	27.320	30.410	27.070	25.210
	0.913	0.832	0.772	0.895	0.797	0.728	0.910	0.827	0.755
SRDenseNet	—	—	28.500	—	—	27.530	—	—	26.050
	—	—	0.778	—	—	0.734	—	—	0.782
CARN	33.520	30.290	28.600	32.090	29.060	27.580	31.920	28.060	26.070
	0.917	0.841	0.781	0.898	0.803	0.735	0.926	0.849	0.784
EDSR	33.920	30.520	28.800	32.320	29.250	27.710	32.930	28.800	<b>26.640</b>
	0.920	0.846	0.788	0.901	0.809	0.742	0.935	0.865	0.803
RDN	34.010	30.570	28.810	32.340	29.260	27.720	32.890	28.800	26.610
	0.921	0.847	0.787	0.902	0.809	0.742	0.935	0.865	0.803
SwinSR	33.070	32.182	31.091	<b>33.345</b>	28.900	<b>31.474</b>	33.856	28.793	24.525
	0.891	0.889	0.847	0.933	0.825	0.854	<b>0.959</b>	0.874	0.781
DeFiAN	33.789	33.747	31.087	32.955	29.050	31.448	34.196	28.537	25.050
	0.908	<b>0.913</b>	0.847	0.931	0.830	<b>0.870</b>	<b>0.959</b>	0.866	0.797
RDAN	<b>35.815</b>	<b>33.881</b>	<b>31.228</b>	33.281	<b>30.133</b>	29.646	<b>34.289</b>	<b>28.816</b>	26.473
	<b>0.923</b>	0.908	<b>0.850</b>	<b>0.939</b>	<b>0.859</b>	0.838	0.958	<b>0.885</b>	<b>0.827</b>

of the dimension of the vectors. This feature has certain advantages in comparing error vectors of different dimensions.

2.4.3.  $L_H$  Hybrid Loss Function. The  $L_H$  hybrid loss function is defined as

$$\begin{cases} L_H = \alpha \cdot L_1 + \beta \cdot L_\infty \\ \alpha + \beta = 1 \end{cases} \quad (11)$$

In practice, we usually use mean absolute error instead of summation of absolute error in order to avoid the correlation between  $L_1$ -norm and vector dimension. However, the direct use of  $L_1$ -norm may reduce the average error, while the absolute error of individual pixels may be still huge. This situation does occur. Because the  $L_1$ -norm only reduces the average error, there is no constrain on the maximum error of a incorporating pixel. Therefore, a new loss function that can not only reflect the overall error, but also effectively reduce the maximum error of a single pixel is required, so as to further improve the quality of image recovery.

The  $L_\infty$ -norm just satisfies this requirement, which allows us to easily compare the max error between a single pixels, which is independent of the number of vector dimension. Therefore, we proposed a hybrid error loss function combining  $L_1$ -norm and  $L_\infty$ -norm, which cannot only ensure the overall error of the image but also effectively reduce the maximum error between individual pixels. Through our tests and analysis, it is found that the recom-

TABLE 3: Parameter size of different method.

Method	Parameters size (M)
SRCNN	16.41
RDN	33.58
SwinIR	36.79
DeFiAN	21.57
Our method	27.96

mended value range of  $\beta$  is [0.002, 0.1]. Too large or too small will deteriorate the performance of the network.

2.5. *Implementation Details.* Residual scaling dense block (RSDB) is composed of RDBs, residual scaling layer, and feature fusion attention layer through dense connection. The detailed pseudocode is show in Algorithm 1. One of the primary modifications from the original RDN model is the utilization of a more effective feature fusion attention layer and residual scaling layer in order to increase the efficiency of hardware resource consumption. In addition, it serves as the fundamental building block of our RDAN method.

Each RSDB module has 6 convolution layers in it. All additional connections are made by pixel-wise concatenation, with the exception of the residual connection between the first layer and the last layer, which is done by pixel-wise concatenation. In dense connection, the activation function is ReLu, and the convolution kernel size is 3. The RSDB module excludes the BN, dropout, pooling, and other structures and only includes the convolution layer, activation layer,

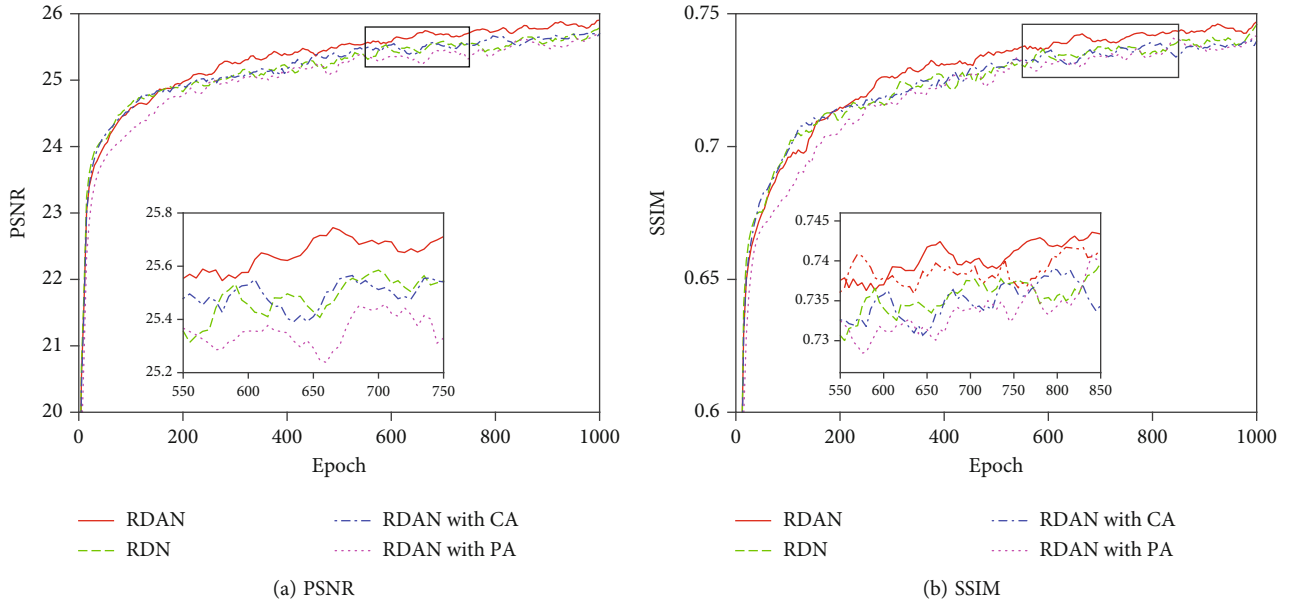


FIGURE 6: Changes in performance metrics of different methods in ablation tests.

residual scaling layer, concat/addition layer, and other essential calculations in order to increase the network’s overall computing efficiency.

In addition, PA, MCA, and other feature fusion attention structures are added to RSDB. In MCA and PA, we reduce the number of feature maps by 8 through convolution layer. The size of MCA and PA convolution kernel is set to  $1 \times 1$  and  $3 \times 3$ , respectively. In multichannel attention, the pooling function adopts average pooling and maximum pooling, respectively. The activation functions are ReLU and Sigmoid, respectively. In MCA output, the number of final output channels is 1. In PA, the final output channel is 3, that is, the RGB component weight of the corresponding color channel.

In the upscale module, the subpixel method is more flexible to realize image scaling, but it is required that the number of feature images input into the subpixel module must be a multiple of the square of the magnification, otherwise integer magnification cannot be achieved. Because this method is different from the upsampling or downsampling method, it will not lose or introduce new pixel information. And the number of feature map can be set through a  $1 \times 1$  convolution layer.

In MCA output, the final output channel number is 1. But in PA output, it is 3, which represents the RGB component weights of the different color channels. During the training process, the preprocessor crops the input images to size  $64 \times 64$  for distinct training sets, with no overlap between images. The training output is the same size as the original input image.

Input and output:  $64 \times 64$  images are used as input for networks with varied magnification, while output high-resolution images vary depending on the magnification and are  $128 \times 128$ ,  $192 \times 192$ , and  $256 \times 256$ , respectively. The input image has been cropped, and it has been cropped to various sizes in accordance with the various needs for magnification.

TABLE 4: Comparison of ablation test performance metrics on the BSD100 dataset (average on last 20 results).

Method	PSNR	SSIM	$L_\infty$ loss
RDAN	25.8387	0.7446	0.0261
Margin	—	—	—
RDN	25.6572	0.7407	0.0267
Margin	-0.70%	-0.52%	2.03%
RDAN with CA	25.6703	0.7385	0.0296
Margin	-0.65%	-0.82%	13.17%
RDAN with PA	25.5777	0.7378	0.0278
Margin	-1.01%	-0.91%	6.26%

### 3. Experiment Results

**3.1. Training Platform, Data, and Evaluation Metrics.** For SR task, there are public datasets for this task in test and validation, mainly include Set5 [40], Set14 [41], BSD100 [42], Urban 100 [43], DIV2K [44], and DTD [45]. The evaluation metrics of training performance includes PSNR and SSIM [46]. PSNR (peak signal to noise ratio) is an objective standard for evaluating images, which is the most common and widely used objective measurement method to evaluate image quality. Therefore, the greater the PSNR value, the less distortion of the image. SSIM (structural similarity) is not only a structural similarity but also a full images quality evaluation metrics. It measures image similarity in three aspects: brightness, contrast, and structure. SSIM value range is  $[0, 1]$ . The larger the value, the smaller the image distortion. The training platform and relevant parameters used in this method are shown in Table 1.

Training settings and results: training optimizer is ADAM, learning rate ( $lr$ )=0.0001,  $\beta_1=0.9$ , and  $\beta_2=0.999$ .

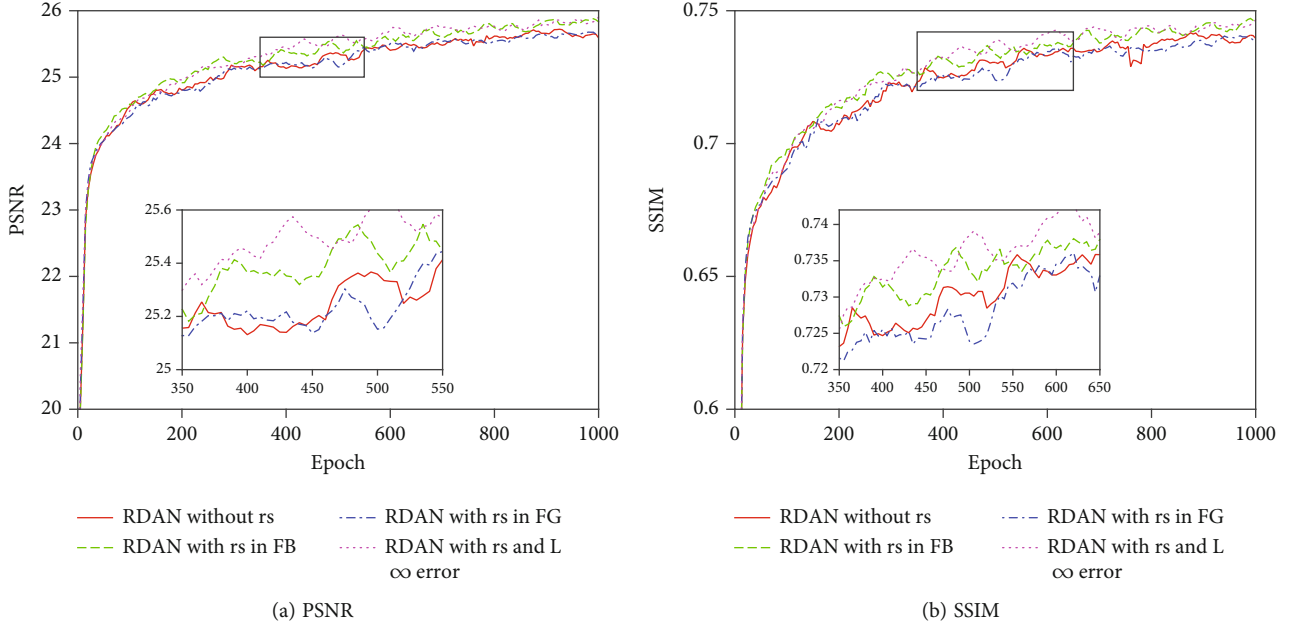


FIGURE 7: Residual scaling ablation comparison test results.

The convolution kernel size is  $3 \times 3$ , the dataset is set according to reference [31], the input image size is  $64 \times 64$ , and the training platform is Keras 2.7. The method proposed by us has achieved ideal results from different datasets and different scaling factor and has better stability compared with the original RDN method.

**3.2. Comparison with Existing Methods.** The comparison of the PSNR/SSIM result with the RDAN method and various recent significant methods is shown in Table 2. Our method has achieved excellent results in PSNR/SSIM/on standard datasets.

Our method achieved 6 best result in both 9 PSNR metrics and SSIM results. One PSNR result rank second, and two SSIM result rank second. The training setting are according to the literature [31]. The size of different is shown in Table 3.

**3.3. Ablation Experiment.** This ablation experiment is designed to evaluate the impact of feature attention structure on network performance in our novel technique. The results of this experiment, which compare the RDN and RDAN methods with different settings, are displayed in Figure 6 and Table 4. From this result, we can see that the network performance is improved after adding the attention module. Compared with the original RDN method, the performance of PSNR/SSIM/ $L_\infty$  is improved by 0.70%/0.52%/2.03%, respectively. The lesser  $L_\infty$  error is, the better the network performance is. Compared with adding CA module or PA module alone, SNR/SSIM/ $L_\infty$  performance decreases by 0.65%/0.82%/13.17% and 1.01%/0.91%/6.26%, respectively.

Additionally, compared to the original RDN technique, performance cannot be increased by adding CA or PA modules alone. Only by combining the two can the performance be enhanced. The comparison experiments demonstrate that

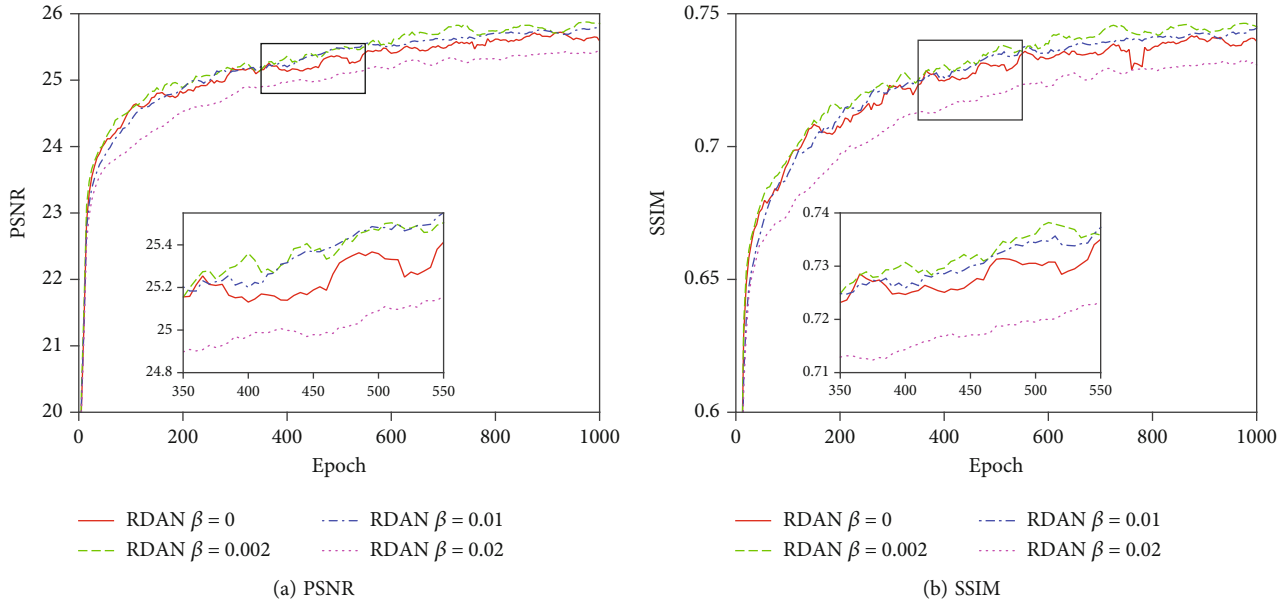
TABLE 5: Residual scaling ablation comparison test performance metrics on the BSD100 dataset (average on last 20 results).

Method	PSNR	SSIM	$L_\infty$ loss
RDAN0	25.6594	0.7397	0.0316
Margin	—	—	—
RDAN1	25.8699	0.7455	0.0270
Margin	0.82%	0.78%	-14.56%
RDAN2	25.7305	0.7415	0.0281
Margin	0.28%	0.24%	-11.08%
RDAN31	25.9647	0.7484	0.0271
Margin	1.19%	1.18%	-14.24%

the addition of the attention structure significantly improved the PSNR/SSIM/ $L_\infty$  error and other performance metrics when compared to the original RDN network.

**3.4. Residual Scaling Layer.** We developed a residual scaling layer in the network to further enhance its residual convergence performance. We designed four groups of comparative tests in order to quantitatively assess and compare the performance of residual scaling layer. They are RDAN1: RDAN with residual scaling in feature block; RDAN2: RDAN with residual scaling layer in feature group; RDAN3: RDAN with residual scaling layer and  $L_\infty$  error loss function, in order to quantitatively assess and compare the performance of residual scaling layer. Figure 7 and Table 5 both show the results. These results demonstrate that the RDAN1 network's performance is superior to the RDAN2. The difference between the two networks is that the residual scaling layer is located in different positions. In RDAN1 network, residual scaling is set in basic block, while residual scaling is set in RSDB in RDAN2 network. But in contrast,

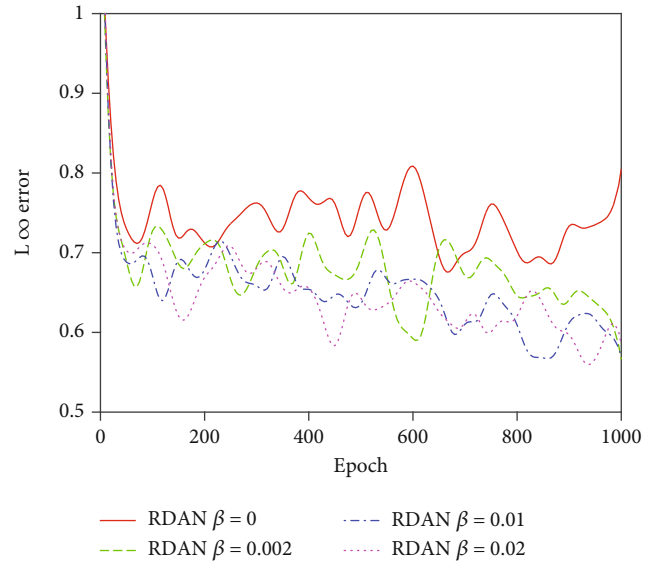


FIGURE 8: Influence of different  $\beta$  on the performance metrics PSNR/SSIM.

the PSNR/SSIM/ $L_\infty$  of these two methods has also been improved compared with the original RDAN0 network with the residual scaling layer. The PSNR/SSIM/ $L_\infty$  performance metrics have been improved by 0.82%/0.78%/14.56% and 0.28%/0.24%/11.08%, respectively. It also shows that the scaling performance in basic block is better than that in RSDB.

Besides, the performance improvement is more pronounced when our  $L_H$  loss function is used, and the performance metrics increase by 1.19%, 1.18%, and 14.24%, respectively. This indicates that residual scaling and  $L_H$  hybrid loss function play an important role in promoting the performance, and these two techniques can be used simultaneously, with particular benefits for PSNR and SSIM performance metrics. It can be seen from this comparative experiment that the position of residual scaling has some impact on the network's performance, and that the performance of residual scaling in basic block is superior to that in RSDB. And combined employment of the  $L_H$  hybrid loss function and residual scaling can achieve to even better benefits.

**3.5. Influence of  $\beta$  in the Loss Function on the Performance Metrics.** We discovered through the previous two sets of comparative experiments that the  $L_H$  loss function can significantly boost the network's performance. Hence, we designed four groups of ablation experiments to assess the performance of ablation on network performance in order to further verify the difference in the loss function. These four sets of comparative tests include values of 0.0, 0.002, 0.01, and 0.02 successively. Figures 8–9 and Table 6 show influence on PSNR/SSIM/ $L_\infty$  error performance of different  $\beta$  value. The outcomes demonstrate that the network performance metrics improves when it is between [0.002, 0.01] after adopting the new hybrid loss function.

FIGURE 9: The effect of  $\beta$  on the performance metrics  $L_\infty$  error.TABLE 6: Influence of  $\beta$  on performance metrics.

Method	PSNR	SSIM	$L_\infty$ loss
RDAN $\beta = 0$	25.6594	0.7397	0.0316
Margin	—	—	—
RDAN $\beta = 0.002$	25.8006	0.7449	0.0278
Margin	0.55%	0.70%	-11.85%
RDAN $\beta = 0.01$	25.7415	0.7425	0.0318
Margin	0.32%	0.38%	0.56%
RDAN $\beta = 0.1$	25.4075	0.7314	0.0306
Margin	-0.98%	-1.12%	-3.16%

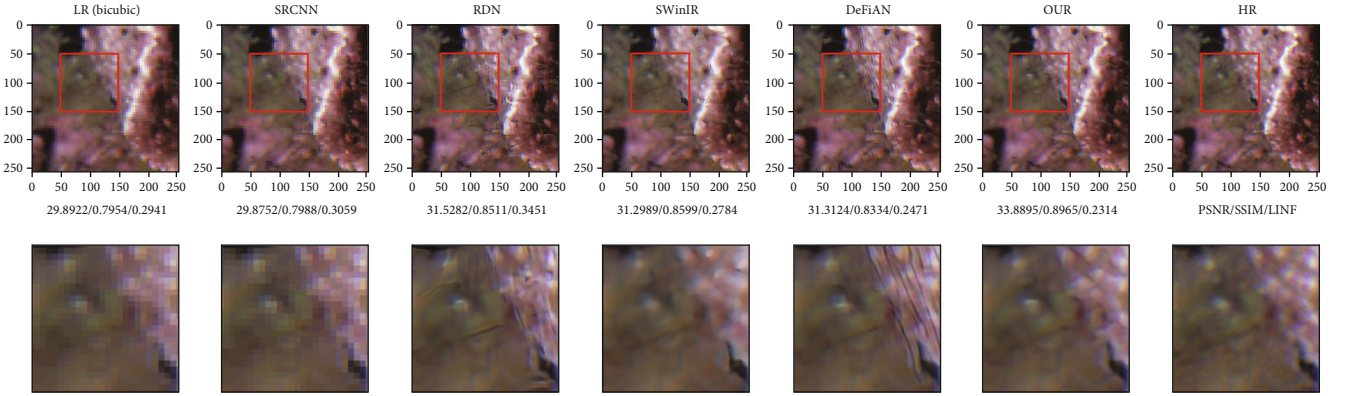


FIGURE 10: Image 0001 from dataset Div2K.

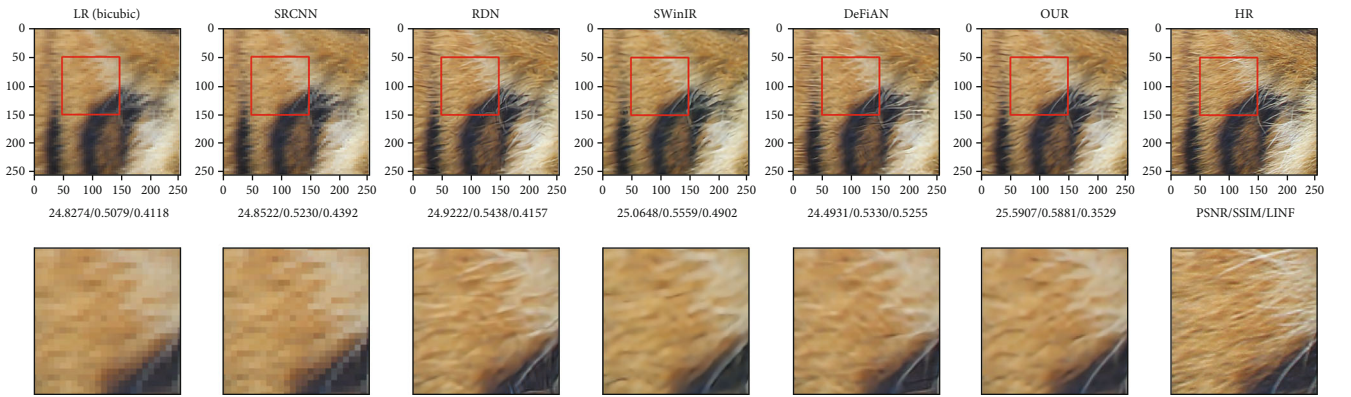


FIGURE 11: Image 0010 from dataset Div2K.

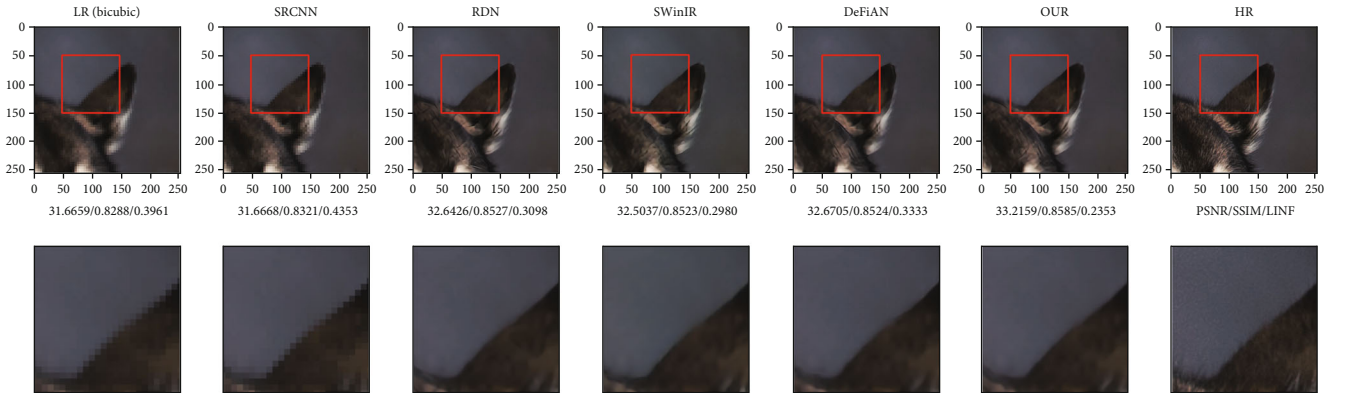


FIGURE 12: Image 0055 from dataset DIV2K.

As can be seen in Figure 9, the network's  $L_\infty$  error is nearly saturated before the adoption of the hybrid loss function, which means that more training cycles will not significantly lower its  $L_\infty$  error. It is clear that the network's  $L_\infty$  error is further decreased when the new hybrid loss function is used, demonstrating the effectiveness of the hybrid loss function in decreasing  $L_\infty$  error. When  $\beta = 0.002$ , PSNR/SSIM performance is enhanced by 0.55%/0.70%, respectively, while  $L_\infty$  performance is enhanced by 11.85%, and the result is quite apparent.

And with  $\beta$  value continues to increase, the PSNR/SSIM metrics increased by 0.32%/0.38%. Since the increase is not obvious, the  $L_\infty$  error decreases by 0.56%. And since  $\beta$  continue to increase, such as  $\beta = 0.02$ , the PSNR/SSIM metrics decreased by 0.98%/1.12%, respectively, and the performance deteriorated significantly. Therefore, it can be seen that when  $\beta$  value in range of  $[0.002, 0.01]$  can effectively improve the network performance from the comparison of this experiment. While when  $\beta$  exceeds this range, the network performance may deteriorate.

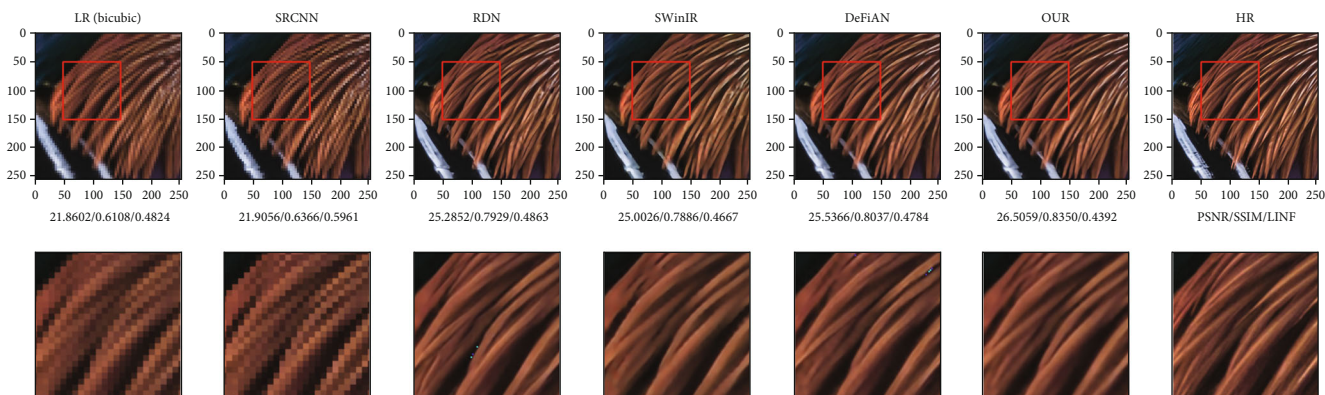


FIGURE 13: Image 0068 from dataset DIV2K.

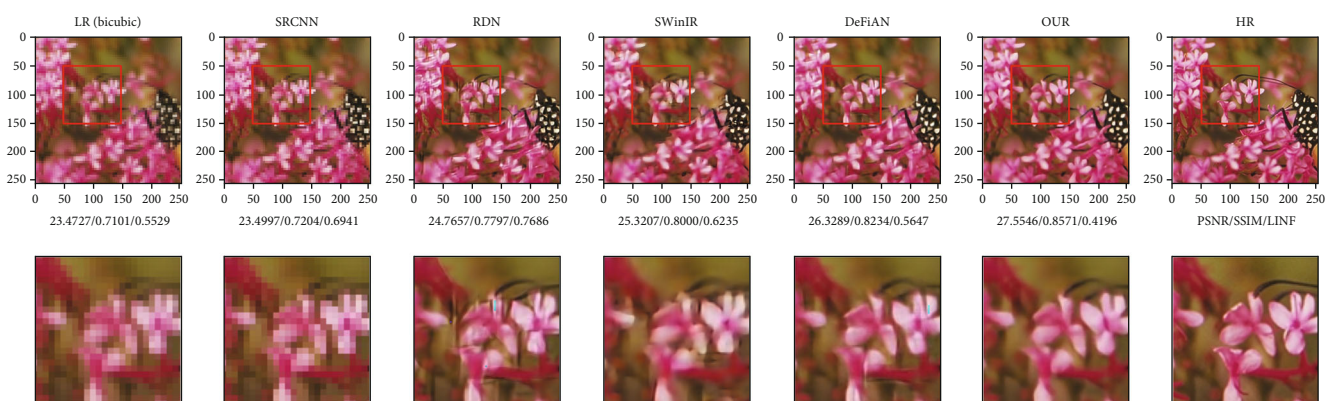


FIGURE 14: Image monarch from dataset Set14.

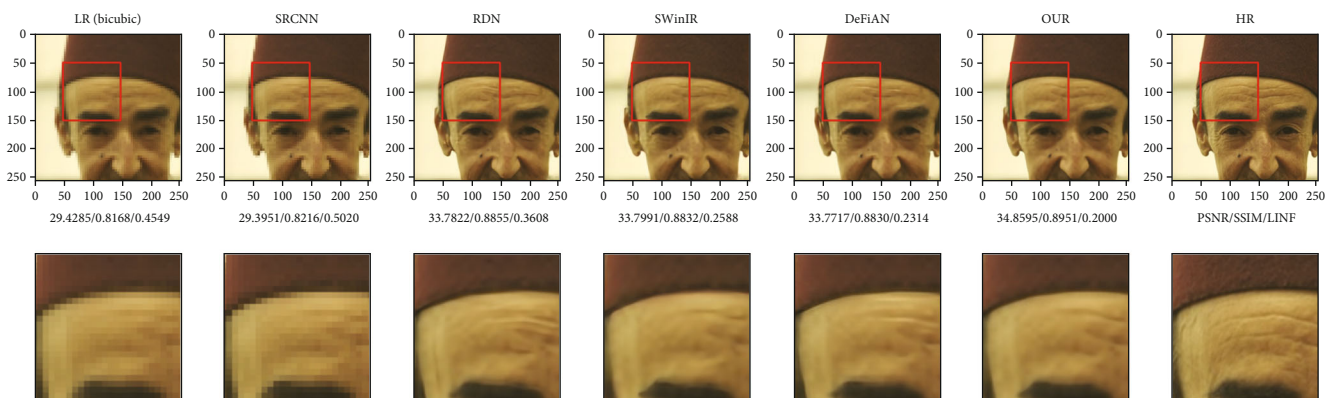


FIGURE 15: Image 189080 from dataset BSD100.

**3.6. Visual Performance Comparison.** The visual performance of various techniques is further compared using the performance of the SRCNN, RDN, SwinIR, and DeFIAN methods, respectively. The results of several methodologies on the datasets DIV2K are shown in Figures 10–13 together with the PSNR/SSIM/ $L_\infty$  error metrics. The PSNR/SSIM/ $L_\infty$  error metrics of the RDAN approach are outstanding in this outcome. Where the loss function  $L_H$  is a hybrid loss function,  $r_s = 0.1$  is the residual scaling factor of the RDAN, and  $\beta = 0.002$ .

From the visual performance, it is clear that the our approach restored results are better than that of SwinIR and DeFIAN methods in Figure 10. For example, the  $L_\infty$  errors of SwinIR and DeFIAN are 0.2784 and 0.2471, respectively, which are significantly bigger than that of the our method with 0.2314  $L_\infty$  error, demonstrating the efficiency with which the RDAN method can reduce the  $L_\infty$  error effectively.

Besides, as show in Figure 11, the PSNR/SSIM of SwinIR method is 25.0648/0.5559, which is better than that of RDN



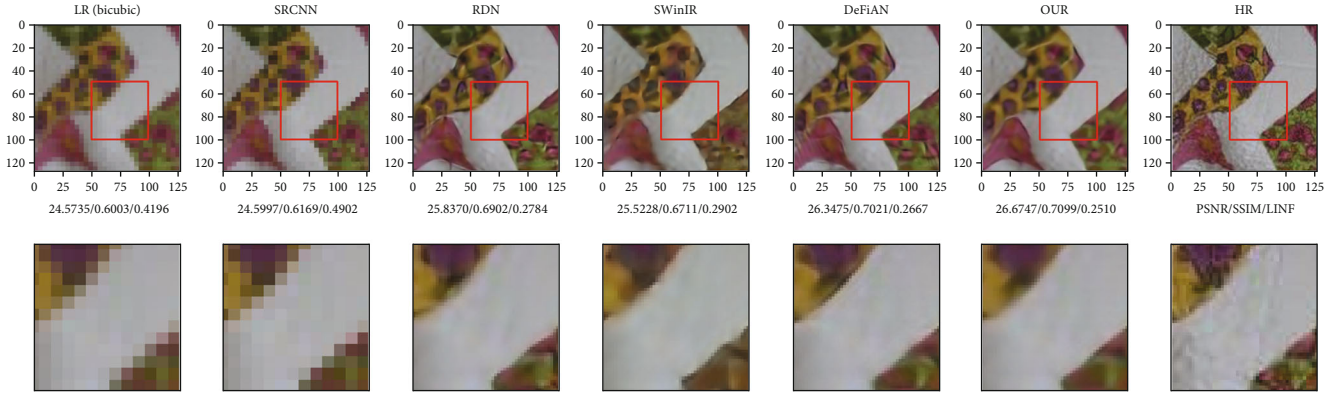


FIGURE 16: Image 7266 from dataset DTD.

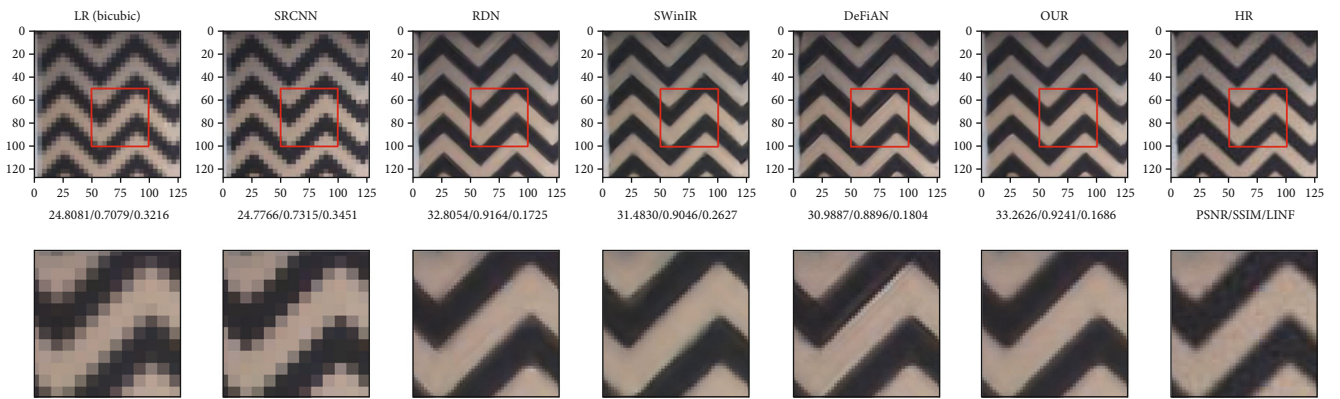


FIGURE 17: Image 7272 from dataset DTD.

method with 24.9222/0.5438, but the  $L_\infty$  error of SwinIR method is 0.4902, which is higher than that of RDN method with 0.4157. The result shows that better PSNR/SSIM result does ensure better  $L_\infty$  performance. Similar result could also be found in Figures 12 and 13.

**3.7. Super-Resolution for Real-World Images.** The training data and the test data for the preceding trials come from the same dataset, and the data from the same dataset have certain similarities even when there is no overlap between the training data and the test data. We employ training and testing on many datasets to further compare the generalization performance of various approaches on unknown datasets, which can enable quantitative comparison and get around the previous restriction that real image comparison could only rely on visual comparison.

In this test, the training dataset is from DIV2K, and the test dataset is from Set14, BSD100, and DTD. Since such images also have high-resolution images for comparison, PSNR, SSIM, and other indicators can be compared quantitatively. As shown in Figures 14–17, the performance of RDN, SwinIR, and DeFiAN methods are compared, respectively.

Taking Figure 16 as an example, the PSNR/SSIM/ $L_\infty$  metrics of SwinIR method and DeFiAN method is 25.5228/0.6711/0.2902 and 26.3475/0.7021/0.2667, respectively, while the performance metrics of our method is

26.6747/0.7099/0.2510, respectively. Compared with SwinIR method and DeFiAN method, the performance is significantly improved.

From this result, we can see that the our method has achieved the best performance, which not only has better PSNR/SSIM performance metrics but also has better  $L_\infty$  error performance. It demonstrates that the RDAN method has better generalization performance on unfamiliar datasets because it not only reduces average error but also maximum error of local pixels.

## 4. Conclusion

Combining the advantage of residual dense net and feature attention mechanism, we proposed a more effective residual dense attention network (RDAN) for image super-resolution, which primarily consists of pixel attention and channel attention and residual dense structures. By introducing attention mechanism, we can focus on the more critical information in the current task among the many input information, reduce the attention to other area, and even filter out irrelevant information to solve the problem of information overload.

Meanwhile, we proposed a hybrid loss function based on the combination of  $L_1$  and  $L_\infty$  error. The range suggested for the parameters is also given according to a set of comparison experiments. Too big or too small may deteriorate the final

performance. The experiment results show that the new method cannot only obtain better PSNR/SSIM performance but also better  $L_\infty$  performance in our proposed RDAN network based on feature attention mechanism, which verifies that the new network has better performance. Moreover, performance verification on real-world images shows that our proposed method has obvious advantages too, which further illustrates the advantages of our new method. The experimental result infers that the residual dense connection equipped with pixel attention and channel attention could achieve positive influence on the final results. So we can conclude that the attention approach together with residual dense structure could improve the performance of our network if it is properly designed.

### Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

### Conflicts of Interest

We declare that there is no conflicts of interests in this research.

### Acknowledgments

The authors would like to acknowledge the support of the National Natural Science Foundation of China (NO. 62162027); Natural Science Foundation of Jiangxi Province (20212BAB202004); Science and Technology Project of Jiangxi Provincial Department of Education (NO. GJJ210646); Key R and D Projects of Jiujiang City (NO. 2020069).

### References

- [1] H. Chen, X. He, L. Qing et al., "Real-world single image super-resolution: a brief review," *Information Fusion*, vol. 79, no. 3, pp. 124–145, 2022.
- [2] Z. Xu, C. Qi, and G. Xu, "Semi-supervised image super-resolution with attention cycle GAN," *IET Image Processing*, vol. 16, no. 4, pp. 1181–1193, 2022.
- [3] Z. H. Shah, M. Müller, T. C. Wang et al., "Deep-learning based denoising and reconstruction of super-resolution structured illumination microscopy images," *Photonics Research*, vol. 9, no. 5, p. B168, 2021.
- [4] M. Jiang, M. Zhi, L. I. Yang, T. Q. Li, and J. C. Zhang, "Super-resolution reconstruction of MR image with self-attention based generate adversarial network algorithm," *Scientia Sinica Informationis*, vol. 51, no. 6, pp. 959–970, 2021.
- [5] L. X. Zhang, R. M. Dong, S. Yuan, W. J. Li, J. P. Zheng, and H. H. Fu, "Making low-resolution satellite images reborn: a deep learning approach for super-resolution building extraction," *Remote Sensing*, vol. 13, no. 15, p. 2872, 2021.
- [6] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2015.
- [7] D. Chao, C. L. Chen, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference on Computer Vision (ECCV)*, vol. 1, pp. 391–407, Springer, Cham, 2016.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–12, Las Vegas, NV, USA, 2016.
- [9] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1646–1654, Las Vegas, NV, USA, 2016.
- [10] H. Hai, P. Li, N. Zou, C. Wang, and D. Xu, "End-to-end super-resolution for remote-sensing images using an improved multi-scale residual network," *Remote Sensing*, vol. 13, no. 4, p. 666, 2021.
- [11] D. Qiu, L. Zheng, J. Zhu, and D. Huang, "Multiple improved residual networks for medical image super-resolution," *Future Generation Computer Systems*, vol. 116, pp. 200–208, 2021.
- [12] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5835–5843, Honolulu, HI, USA, 2017.
- [13] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4809–4817, IEEE Computer Society, Venice, Italy, 2017.
- [14] L. Zha, Y. Yang, Z. Lai, Z. Zhang, and J. Wen, "A lightweight dense connected approach with attention on single image super-resolution," *Electronics*, vol. 10, no. 11, p. 1234, 2021.
- [15] G. Huang, Z. Liu, V. Laurens, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, Honolulu, HI, USA, 2017.
- [16] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2472–2481, Salt Lake City, UT, USA, 2018.
- [17] N. Ahn, B. Kang, and K. A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network: 15th European conference," in *Computer Vision – ECCV 2018*, pp. 256–272, Springer, Munich, Germany, 2018.
- [18] C. Ledig, L. Theis, F. Huszar et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 7, pp. 105–114, Honolulu, HI, USA, 2017.
- [19] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1134–1140, Honolulu, HI, USA, 2017.
- [20] X. Wang, H. Ji, C. Shi et al., "Heterogeneous graph attention network," in *The World Wide Web Conference*, pp. 2022–2032, San Francisco, CA, USA, 2019.
- [21] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2356–2365, Seattle, WA, USA, 2020.



- [22] Y. F. Huang, J. Li, X. B. Gao, T. Y. Hu, and W. Lu, "Interpretable detail-fidelity attention network for single image super-resolution," *IEEE Transactions on Image Processing (TIP)*, vol. 30, pp. 2325–2339, 2021.
- [23] J. Liang, J. Cao, G. Sun, K. Zhang, and R. Timofte, "SwinIR: image restoration using Swin Transformer," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 1833–1844, Montreal, BC, Canada, 2021.
- [24] C. Fang, Y. Zhu, L. Liao, and X. F. Lin, "TSRGAN: real-world text image super-resolution based on adversarial learning and triplet attention," *Neurocomputing*, vol. 455, no. 2, pp. 88–96, 2021.
- [25] L. Shan, X. Bai, C. Liu, Y. Feng, Y. Liu, and Y. Qi, "Super-resolution reconstruction of digital rock CT images based on residual attention mechanism," *Advances in Geo-Energy Research*, vol. 6, no. 2, pp. 157–168, 2022.
- [26] M. I. Georgescu, R. T. Ionescu, A. I. Miron et al., "Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution," in *Proceeding of the Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 1–10, New Orleans, Louisiana, 2022.
- [27] C. Chen, C. Qing, X. Xu, and P. Dickinson, "Cross parallax attention network for stereo image super-resolution," *IEEE Transactions on Multimedia*, vol. 24, pp. 202–216, 2022.
- [28] X. Dou, C. Li, Q. Shi, and M. Liu, "Super-resolution for hyperspectral remote sensing images based on the 3D attention-SRGAN network," *Remote Sensing*, vol. 12, no. 7, p. 1204, 2020.
- [29] X. Yang, M. Zhang, W. Li, and T. Ran, "Visible-assisted infrared image super-resolution based on spatial attention residual network," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [30] S. T. Xia, "S2A: scale-attention-aware networks for video super-resolution," *Entropy*, vol. 23, no. 11, p. 1398, 2021.
- [31] W. Xu, R. Chen, B. Huang, and Q. Zhou, "Enhanced context attention network for image super resolution," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11665–11673, 2021.
- [32] Y. Yang and Y. Qi, "Hierarchical accumulation network with grid attention for image super-resolution," *Knowledge-Based Systems*, vol. 233, p. 107520, 2021.
- [33] C. Chen, D. Gong, H. Wang, Z. Li, and K. Y. K. Wong, "Learning spatial attention for face super-resolution," *IEEE Transactions on Image Processing*, vol. 30, pp. 1219–1231, 2021.
- [34] Z. Zhang and B. Zhang, "Hierarchical attention maps for super resolution," *Journal of Physics: Conference Series*, vol. 1883, no. 1, p. 012057, 2021.
- [35] Z. Li, Y. Zhang, and Y. Yang, "Residual shuffle attention network for image super-resolution," *Journal of Physics: Conference Series*, vol. 2025, no. 1, article 012069, 2021.
- [36] G. Pandey and U. Ghanekar, "Single image super-resolution using multi-scale feature enhancement attention residual network," *Optik*, vol. 231, no. 1, article 166359, 2021.
- [37] H. Zhang, C. Lan, and Z. Gao, "Residual attention fusion network for single image super-resolution," *Journal of Physics: Conference Series*, vol. 2031, no. 1, article 012013, 2021.
- [38] J. Zhao, X. Sun, Q. Li, and X. Ma, "Edge caching and computation management for real-time internet of vehicles: an online and distributed approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2183–2197, 2021.
- [39] J. Zhao, Q. Li, Y. Gong, and K. Zhang, "Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7944–7956, 2019.
- [40] M. Bevilacqua, A. Roumy, C. Guillemot, and M. Alberi-Morel, "Low-complexity single-image super-resolution based on non-negative neighbor embedding," in *Proceedings of the British Machine Vision Conference 2012*, 2012.
- [41] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces*, pp. 711–730, Springer, Berlin, Heidelberg, 2010.
- [42] 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 *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, pp. 416–423, Vancouver, BC, Canada, 2001.
- [43] J. B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 5197–5206, Boston, MA, USA, 2015.
- [44] R. Timofte, S. Gu, J. Wu et al., "NTIRE 2019 challenge on real image super-resolution: methods and results," in *Proceeding of the Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2019.
- [45] L. Sharan, R. Rosenholtz, and E. H. Adelson, "Accuracy and speed of material categorization in real-world images," *Journal of Vision*, vol. 14, no. 9, p. 12, 2014.
- [46] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.