

Received July 3, 2021, accepted July 19, 2021, date of publication July 26, 2021, date of current version October 18, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3100069

A Super Resolution Algorithm Based on Attention Mechanism and SRGAN Network

BAOZHONG LIU^{ID1} AND JI CHEN^{ID2}

¹Chongqing College of Electronic Engineering, Chongqing 401331, China

²Chongqing Institute of Engineering, Chongqing 400056, China

Corresponding author: Baozhong Liu (jianqiu27@163.com)

This work was supported in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJZD-K201901901, and in part by the Scientific Research Fund of Chongqing Institute of Engineering under Grant 2019gcky03.

ABSTRACT Image super-resolution reconstruction uses a specific algorithm to restore the low resolution blurred image in the same scene to a high resolution image. In recent years, with the vigorous development of deep learning, this technology has been widely used in many fields. In the field of image super-resolution reconstruction, more and more methods based on deep learning have been studied. According to the principle of GAN, a pseudo high-resolution image is generated by the generator, and then the discriminator calculates the difference between the image and the real high-resolution image to measure the authenticity of the image. Based on SRGAN (super resolution general adverse network), this paper mainly makes three improvements. First, it introduces the attention channel mechanism, that is, it adds Ca (channel attention) module to SRGAN network, and increases the network depth to better express high frequency features; Second, delete the original BN (batch normalization) layer to improve the network performance; Third, modify the loss function to reduce the impact of noise on the image. The experimental results show that the proposed method is superior to the current methods in both quantitative and qualitative indicators, and promotes the recovery of high-frequency detail information. The experimental results show that the proposed method improves the artifact problem and improves the PSNR (peak signal-to-noise ratio) on set5, set10 and bsd100 test sets.

INDEX TERMS Super resolution, GAN network, attention mechanism.

I. INTRODUCTION

The task of evaluating high resolution (HR) images from low resolution (LR) images is called super resolution. Super resolution has a good future in various fields, such as military industry, agriculture, or medicine. Super resolution plays an important role in the field of artificial intelligence [1], [2]. This paper focuses on the introduction of Single Image Super-Resolution (SISR) technology, which has been widely used in image compression, medical imaging [3]–[5], remote sensing imaging [6], public security [7] and other fields due to its flexibility, simplicity and high practicability. It is a research hotspot in the field of image processing.

Chen *et al.* [8] first proposed a three-layer convolutional neural network super-resolution method (SRCNN), which jointly optimizes the three stages of feature extraction, nonlinear mapping and image reconstruction in an end-to-end manner, demonstrating that the convolutional neural

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

network can learn the mapping from LR to HR in an end-to-end manner. Zhongfan *et al.* [9] proposed an effective sub-pixel convolutional neural network (ESPCNN). In this paper, the effective sub-pixel convolution is used to replace the up sampling operation, and the simple mathematical operation is used to replace the up sampling, which greatly improves the running speed. It is found that deep network is more effective than shallow network [10]. Shuaikun *et al.* [11] used 20 layers of deeper network in VDSR (super resolution using very deep eep revolutionary networks). At the same time, they used gradient clipping, jump connection and other learning strategies to accelerate network convergence. However, if the network model is too deep, the gradient will disappear or explode, so that the model cannot be trained, and the best effect cannot be achieved through training [12]. Chen *et al.* [13] proposed a dense connection to promote information flow between layers. Renpu *et al.* [14] constructed an information distillation network (IDN), which contains an enhancement unit and a compression unit. This method can maintain good reconstruction accuracy and

achieve real-time performance. Yong *et al.* [15] obtained the structural feature information of low resolution image through coding network, and obtained high-frequency features through multi-path feedforward network composed of stage feature fusion unit. Haitao *et al.* [16] completed the reconstruction task of the generator module by progressively extracting high-frequency features of different scales of the image. Yan *et al.* [17] constructed a detail supplement network to supplement the image features by combining the local similarity of the image, and used a convolution layer to fuse the features obtained by the detail supplement network with the features extracted by the feature extraction network, so as to reconstruct the high-resolution image. For industrial application, Chen *et al.* proposed a new method, carn (cascading residual network), which uses local and global cascading mechanism to integrate multi-layer features, so as to speed up the running speed of the model [18]. Ying *et al.* [19] integrated the prior knowledge of soft edge image into the model to assist image reconstruction, so as to avoid blindly increasing the network depth. To solve the problem that interpolation processing, especially when the super-resolution factor is very high, may cause the image to become smoother, Xie *et al.* proposed a deep recursive fusion network (DRFN), which uses transposed convolution instead of bicubic interpolation for up sampling, and integrates different levels of features extracted from recursive residual block storage. Thus the final high resolution image is formed [20]. In view of the fact that the existing networks seldom mine the correlation of features between layers, the learning ability of convolutional neural networks is reduced.

In order to avoid the over fitting of low-frequency texture under high complex network training, the feature of different frequency is processed hierarchically. Super resolution feedback network (SRFBN), which uses high-level information to refine low-level representation, uses hidden state in a constrained recurrent neural network to achieve this feedback [21]. In order to use a model to solve the super-resolution of arbitrary scale factors (including non integer scale factors). The scale factor is used as the input to dynamically predict the weights of the up sampling filter, and these weights are used to generate high-resolution images of any size [22]–[24]. Chao and Hongyu [25] proposed a DPSR (deep plug and play super-resolution) method. A new single image super-resolution degradation model was designed to replace the blind deblurring kernel estimation, and the energy function was introduced to optimize the degradation model.

For this paper, the main contributions are as follows:

- (1) Based on SRGAN network, this algorithm changes the structure of the network, introduces rcan [26] (residual channel attention networks) attention mechanism module, and increases the network depth to better express the high frequency [27] features.
- (2) As for the loss function, the square term of MSE (mean square error) increases the influence of large

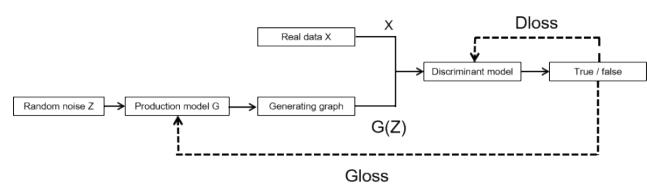


FIGURE 1. Network Framework of SRGAN.

error on image super-resolution, so L1 loss function is used [28].

- (3) The experimental results show that the artifacts produced by SRGAN super-resolution are reduced to a certain extent, and the stability of the model is enhanced, and the experimental results are good.

II. SUPER RESOLUTION RECONSTRUCTION MODEL

SRGAN is a kind of super-resolution network which can generate adversary network. The effect of SRGAN is to improve the super-resolution amplification factor. The research shows that the super-resolution effect is better when the magnification is $\times 4$ or $\times 8$. Because the image obtained from the traditional super-resolution network reconstruction or linear network will produce smoother image quality, the restoration degree of relevant details is not good. However, SRGAN is equivalent to the process of two simulated models fighting each other. G network generates pseudo high-resolution images according to the training set, and D network judges how much error there is between the images generated by G network and the real images according to the real high-resolution images, and continuously carries out the process of cyclic generation and verification. The antagonism formula of G and D is shown in (1).

$$\begin{aligned} & \min_G \max_D V(D, G) \\ &= E_{x-P_{data}(x)}[\log D(x)] + E_{z-P_z(Z)}[\log(1 - D(G(z)))] \quad (1) \end{aligned}$$

where x is the real sample; $D(x)$ is the probability that x is judged to be a real picture after passing through the discriminator; Z is the input noise, and $G(z)$ is the output sample of the generated network.

A. SRGAN NETWORK MODEL

The network of SRGAN is composed of generating network and discriminating network. The low resolution image is amplified by generating network to reconstruct the high resolution image. The discriminating network judges the low resolution image and the high resolution image, and outputs the discriminating probability to judge whether it is true or false. The loss function in the network is added with the identification loss to guide the network to generate high-quality pseudo high-resolution images.

SRGAN generation network is mainly composed of sub-pixel layer and residual block, but the effect image is not good in high-frequency features [29]. Therefore, attention

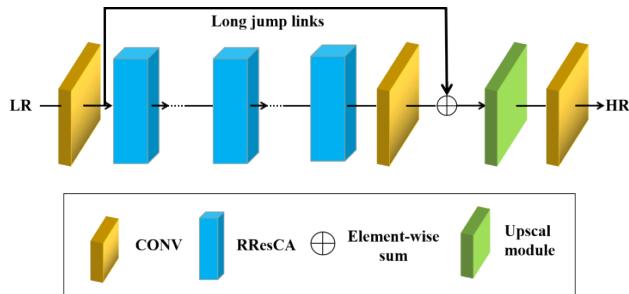


FIGURE 2. SRCAGAN Generates Network Model Diagrams.

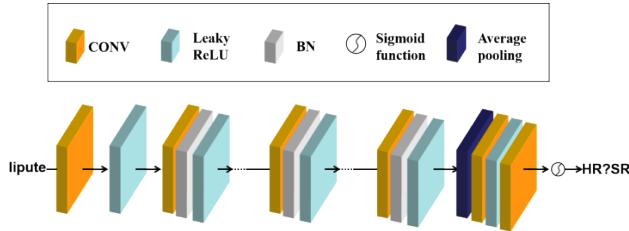


FIGURE 3. SRCAGAN Identification Network Model Diagrams.

mechanism is added in reading the paper, especially in the processing of high-frequency features.

1) GENERATING NETWORK

The network model generated in this paper is shown in Figure 2, and its structure includes six residual modules and long jump connections. At the same time, each residual block contains 10 short jump links and residual channel attention blocks. The structure of residual module including residual module can make the training of super-resolution algorithm deeper. Moreover, residual attention block is introduced to recover high frequency image information better. We will introduce in detail the composition of the attention block of the residual channel in the generation network.

2) AUTHENTICATION NETWORK

In this paper, the training discriminant network is a linear network structure. Referring to Xingchen *et al.* [30], the structure of the discriminant network in SRGAN model is proposed. The specific structure is shown in Figure 3. The first layer of the model is a 2D convolution and a leaky relu activation function, followed by seven modules composed of convolution, BN layer and leaky relu activation function. The last part consists of an average pooling layer, two convolution layers and a leaky relu function. In order to simplify the calculation process and avoid introducing the maximum pooling layer into the network structure, leaky relu is used as the activation function. With the deepening of network model layers, the ability of feature extraction is enhanced, the number of image features is increasing, and the convolution size of each extracted feature is decreasing. The effect of leaky relu activation function is better than that of relu function in negative value, which can avoid sparse gradient. Therefore, leaky relu is selected as the activation function. Finally, the parameters

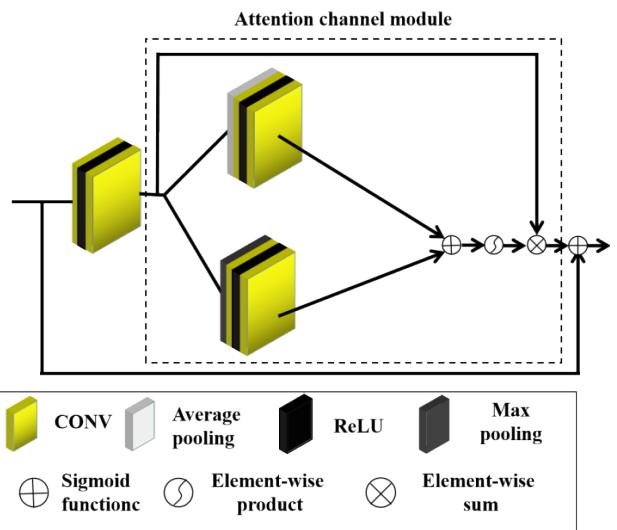


FIGURE 4. ResCA Attention Mechanism Module Diagram.

of the model are passed into the two full connection layers and the last SIGMOD activation function to obtain the similarity probability between the false image generated by the generated network and the real image. After continuous confrontation with the generated network, until the value obtained by the identification network tends to be stable. In order to avoid the gradient disappearing, after leaky relu, the network uses batch normalization to enhance the generalization ability.

B. ATTENTION MECHANISM

The attention mechanism module focusing on high frequency is called Ca, which is obtained by referring to rca (residual channel attention block) and CBAM [31] (convolutional block attention module). It is composed of an average pooling layer, two convolution layers, and a relu activation function, while the other module is composed of a maximum pooling layer, a convolution layer and a convolution layer. It consists of two convolution layers and a relu activation function. The outputs of the two modules are added to form a residual block. The output is obtained by multiplying the input X by the output y processed by the convolution layer. Through this method, we can make the important channel, namely high-frequency feature weight larger, and reduce the channel weight of the part with small improvement of image quality. The specific structure is shown in Figure 4.

The general CA includes the following three parts:

1. Squeeze is Ca (channel attention) is to compress the image features into a point, SA (spatial attention) is to compress n features into a map.
2. In order to get the connection between channels, we can get the vector of C (CA); Or through the convolution layer, we can get the spatial region relation to get $H \times W \times 1$ (SA).
3. Scaling, the same as sigmoid, converts the attention map into a graph between 0 and 1, and then performs point multiplication with the input.

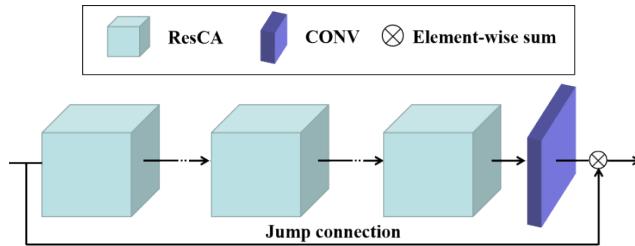


FIGURE 5. Module of RResCA.

Inspired by EDSR residual module [23], because the long jump connection between residual block and residual block can make the main part of the network pay more attention to the information of low resolution features, this paper connects several Resca modules to form rresca module. The model is shown in Figure 5.

It has been proved in reference [32] that the neural network constructed by superimposed residual blocks and long jumps performs better. Therefore, the residual group consists of several rresca modules. But because the network is too deep will cause training difficulties, so the total residual rresca module number is set as 5 and 10, through the comparison of the two to get the best training effect.

At the same time, remove all BN layers of the generated network in SRGAN, because removing BN layer has been proved to significantly enhance the performance of the network in different PSNR oriented tasks, and reduce the computational complexity of network training, including super-resolution and deblurring [33]. The function of BN layer is to normalize all the features of images with the method of mean and variance processing in batch during the model training period. During the test period, the estimated mean and variance obtained from the training model are substituted into the model for testing. However, when there is a big difference between the training data set and the test data set, the BN layer often produces uncomfortable artifacts [34]. This is because the BN layer weakens the generalization ability of the model to a certain extent. It is observed by experience that the BN layer may produce artifacts when the network is deep or under the training of GAN network. These artifacts occasionally appear between iterations and different settings, lacking stability. Therefore, the BN layer is removed for the stability and consistency of training. In addition, the removal of BN layer helps to improve the generalization ability, reduce the computational complexity of the model and the memory usage of the server.

C. LOSS FUNCTION

The loss function of SRGAN is shown in formula (2):

$$l^{SR} = l_x + 2 * 10^{-9} l_{tv} + 10^{-3} l_{Gen}^{SR} \quad (2)$$

Among them: l_x The loss function is called minimum absolute deviation:

$$l_x = l_{l1} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} |(I_H)_{x,y} - (G_{\theta_G}(I_L))_{x,y}| \quad (3)$$

L1 loss function is primitive in error calculation method, it will not have too much penalty for large error term, and the error between each pixel relative to the real image pixel is proportional to the change of L1 loss function and the absolute value of error. According to wGAN network, the L1 loss function is designed as formulas (2) and (3).

Among $(I_H)_{x,y}$ In order to obtain the high resolution image based on the loss function, $(G_{\theta_G}(I_L))_{x,y}$ It is the reconstructed image.

In addition to the relevant experiments of L1 loss function, a series of comparative experiments of MSE loss function are also done. MSE loss function is often used in the reconstruction model of super-resolution network, which directly optimizes the square difference between each pixel of high resolution and low resolution images. The PSNR (peak signal to noise ratio) of the generated image is high.

Although the convergence of MSE loss function is good, the weight of MSE loss function to outliers is usually large, which makes the term with larger error have more influence than the term with smaller error. At the same time, MSE is the method of using the original measurement, which is also the reason why the intuitive feeling of human eyes is not friendly and often causes artifacts(4) Is the formula of MSE loss function.

$$l_{MSE}^{SR} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I_{x,y}^{LR}))^2 \quad (4)$$

In the process of super-resolution reconstruction, as long as there is a little noise in the image, it will have a great impact on the reconstruction effect, because many algorithms will amplify the noise in super-resolution reconstruction. At this time, in order to maintain the smooth integrity of the image, we need to add a regular term to the optimization model. Tvlloss is a regular term usually used. The difference between adjacent pixels in the image can be solved by using tvloss. The formulas are as follows (5):

$$l_{tv} = \int_D \sqrt{u_x^2 + u_y^2} dx dy \quad (5)$$

The specific formula against loss is as follows (6):

$$l_{Gen}^{SR} = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_D}(I^{LR})) \quad (6)$$

III. EXPERIMENT AND ANALYSIS

A. EXPERIMENTAL ENVIRONMENT AND ITS DATA SET

The GPU used in this experiment is NVIDIA geforce GTX Titan x, the programming language is python, IDE is pychar 2017, and the deep learning framework is python.

Voc2012 data set is used in training, and the magnification is X4. In the test, the bicubic algorithm [35], SRCNN algorithm [36], SRResNet algorithm [37] and SRGAN algorithm are compared horizontally by using set5, set14, urban100 and bsd100 data sets. This experiment is based on SRGAN to change the model, so focus on comparing the

super-resolution PSNR value, SSIM value and image visual effect with SRGAN algorithm.

B. EVALUATION CRITERION

In this experiment, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) were used as two criteria to evaluate images.

1) PSNR

Peak signal-to-noise ratio (PSNR) is one of the most widely used standards to evaluate image quality. The method of mean square error is usually used to judge the image quality. For monochrome $w \times H$ The original HD image of H and the super-resolution image are obtained by MSE formula:

$$M_{MSE} = \frac{1}{WH} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} [X(i, j) - Y(i, j)]^2 \quad (7)$$

For the three color high-definition original image and super-resolution image, each pixel has three channels, so the formula is:

$$M_{MSE} = \frac{1}{3WH} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} [X(i, j)_{R,G,B} - Y(i, j)_{R,G,B}]^2 \quad (8)$$

So the formula of PSNR is as follows:

$$P_{PSNR} = 10 \log\left(\frac{X_{MAX}^2}{M_{MSE}}\right) = 20 \log\left(\frac{X_{MAX}}{\sqrt{M_{MSE}}}\right) \quad (9)$$

Generally, the higher the PSNR, the better the effect of super-resolution image reconstruction. However, since the appearance of GAN network, human visual perception has been introduced into super-resolution. Although the peak signal-to-noise ratio of GAN network is not as good as the previous linear network, the visual perception effect of GAN network is much better than other linear networks.

2) SSIM

Structural similarity is another widely used measurement index in image super-resolution reconstruction, which reflects the real feeling of human visual effect. The SSIM formula is based on the mutual measurement of three indexes between sample X and Y: luminance, contrast and structure. The formula is as follows:

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \quad (10)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \quad (11)$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x + \sigma_y + c_3} \quad (12)$$

$$SSIM(x, y) = [l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma] \quad (13)$$

Commonly $c_3 = c_2/2$, among μ_x Is the average of x , μ_y is the average value of Y . σ_x^2 Is the variance of X , σ_y^2 is the variance of Y , σ_{xy} Is the covariance of X and y . In practical application, it is usually set as follows $\alpha = \beta = \gamma = 1$.

TABLE 1. Comparison of 5 Levels of Attention Module with 10 Levels of Attention module(k=4)

| SRCAGAN-5 | | | SRCAGAN-10 | |
|--------------|---------|----------|------------|----------|
| Set5 | L1 loss | MSE loss | L1 loss | MSE loss |
| PSNR | 29.99 | 29.61 | 30.57 | 29.89 |
| SSIM | 0.8657 | 0.8531 | 0.8730 | 0.8612 |
| Set14 | L1 loss | MSE loss | L1 loss | MSE loss |
| PSNR | 26.39 | 26.25 | 26.58 | 26.29 |
| SSIM | 0.7426 | 0.7286 | 0.7666 | 0.7328 |

C. ANALYSIS OF EXPERIMENTAL RESULTS

In this experiment, objective evaluation results and subjective evaluation results are used to show the super-resolution ability of the model. Not only the network depth and loss function used in this model are compared vertically, but also the representative super-resolution models in different periods are compared horizontally.

1) COMPARISON OF MODELS AND ALGORITHMS

In this paper, CA-10 (10 layer attention mechanism network) is used to compare the super-resolution results of each model when the loss function is MSE and L1. The amplification coefficient K is 4.

In the process of experiment, we try 5, 10, 15 residual blocks of attention mechanism. When we do 15 residual blocks, because of too many parameters, it is easy to cause memory overload during training, so only 5 and 10 residual blocks are selected for comparison.

It can be seen from the table that the performance of PSNR and SSIM of srcaGAN-10 on set5 and set14 is better than that of srcaGAN-5. Therefore, srcaGAN-10 is selected as the comparison model in the subsequent comparison with other algorithm models.

2) OBJECTIVE EVALUATION RESULTS

In this section, we will compare the differences between each algorithm model and our algorithm in three datasets: set5, set14 and bsd100.

Set5 and set14 datasets are low complexity single images, which are based on non negative domain embedding for super-resolution research. That is to say, based on the low resolution image, the high resolution image can be reconstructed by deep learning algorithm. The data set was published in 2012 by Bell Laboratories, France, and is widely used in super-resolution research.

Bsd100 data set is a data set for image segmentation and edge detection. The content includes the image segmentation of 1000 Corel datasets marked by 30 people by hand. 50% of the segmentation is to present the main color image, and the other 50% is obtained through the main gray image. The data set was proposed by UC Berkeley in 2001.

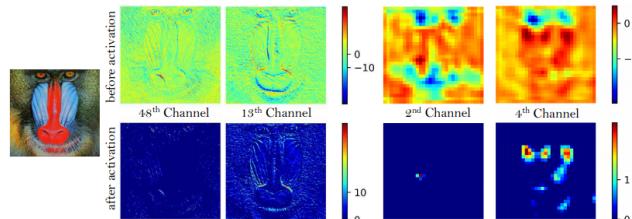
According to table 1, table 2 and table 3, SRResNet performs best in PSNR and SSIM. However, SRResNet is a residual linear network, and the effect of high-frequency

TABLE 2. Comparison of PSNR and SSIM on Set5 and Set14(k=4)

| Set5 | Bicubic | SRCCNN | SRResNet | SRGAN | Ours |
|-------|---------|--------|----------|--------|--------|
| PSNR | 28.51 | 30.12 | 32.07 | 29.39 | 30.20 |
| SSIM | 0.8209 | 0.8630 | 0.9020 | 0.8469 | 0.8726 |
| Set14 | Bicubic | SRCCNN | SRResNet | SRGAN | Ours |
| PSNR | 25.98 | 27.20 | 28.50 | 26.01 | 26.61 |
| SSIM | 0.7485 | 0.7860 | 0.8185 | 0.7380 | 0.7668 |

TABLE 3. Comparison of PSNR and SSIM on BSD100(k=4)

| BSD100 | Bicubic | SRCCNN | SRResN et | SRGAN | Ours |
|--------|---------|--------|-----------|--------|--------|
| PSNR | 25.95 | 26.70 | 27.60 | 25.18 | 26.61 |
| SSIM | 0.6957 | 0.7286 | 0.7619 | 0.6677 | 0.7313 |

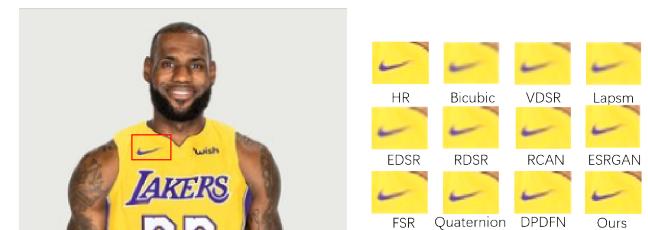
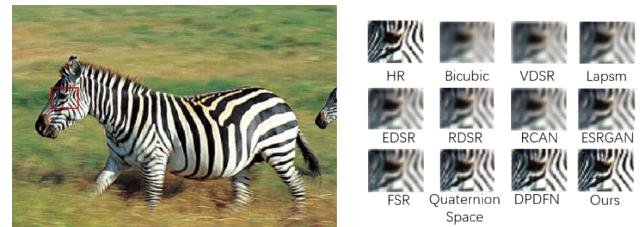
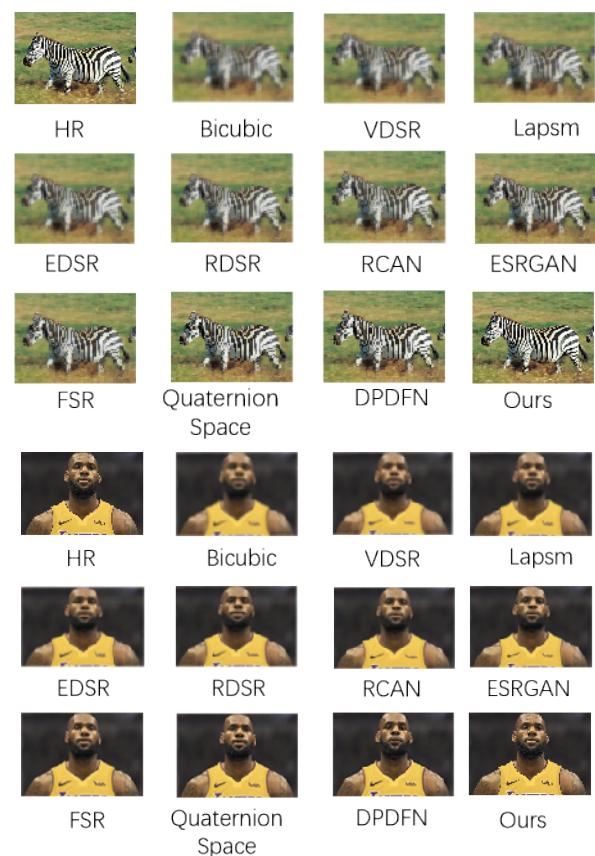
**FIGURE 6.** This paper presents a method of channel visualization.

feature restoration is not as good as that of this paper, because the attention mechanism residual block added in this paper is mainly for the restoration of high-frequency features, and the image displayed is much better than the image generated by SRResNet in human visual effect. The next section will mainly analyze the subjective evaluation results.

We also discuss a variant of srgan, the method proposed in this paper. Compared with the common loss classification methods based on VGg network, this paper proposes a more suitable loss classification method based on sr-minc. It is based on a fine tuned srgan network, which is used to reconstruct details effectively. It focuses on texture. Although the perceptual index gain caused by MinC loss is negligible, we still think that exploring texture centered perceptual loss is very important for super-resolution reconstruction, and the visualization results are shown in Figure 6.

3) SUBJECTIVE EVALUATION RESULTS

In the experiment, SRResNet with the highest PSNR value after reconstruction, SRGAN before improvement and the model in this paper after improvement were selected to compare the visual effects after reconstruction. It can be seen from Figure 7 and 8 that although the PSNR and SSIM values of the images reconstructed by SRResNet have good performance, in the first row of girls' images, the lock decoration on the chest is fuzzy, which is not as good as the visual effect of the images reconstructed by SRGAN and this paper. At the same time, compared with the images of this

**FIGURE 7.** Comparison of the Reconstruction Effect of Each Algorithm (local).**FIGURE 8.** Comparison of the Reconstruction Effect of Each (overall situation).

model, SRGAN images are smoother, and the visual effect is not as good as the details of this model. In the second row of monkey images, the details of the monkey tentacles restored in this paper are better than those of the first two models. Therefore, this paper has some advantages in the restoration of high-frequency feature details, and the restored image is more similar to the original image.

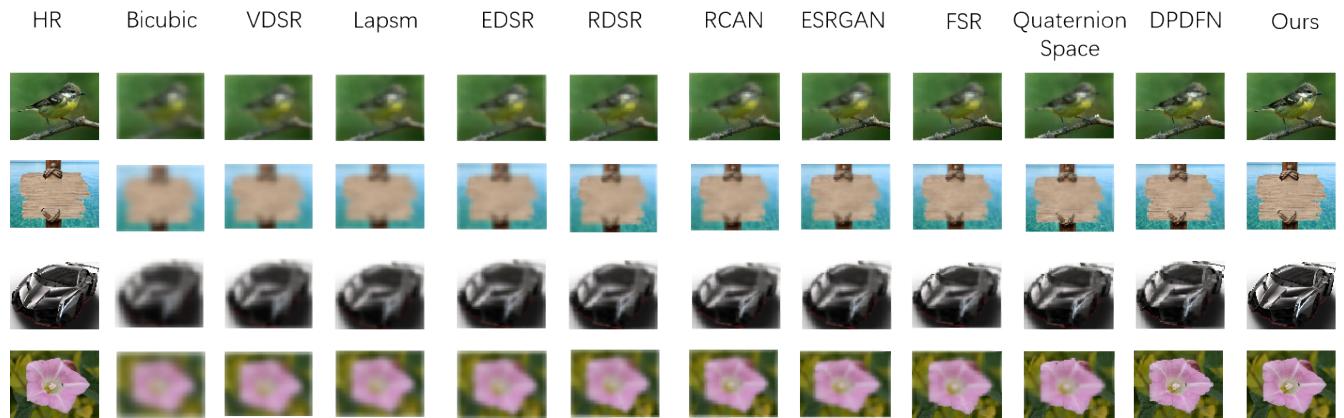


FIGURE 9. Visual comparison of the MHRL datasets for 3 times super-resolution.

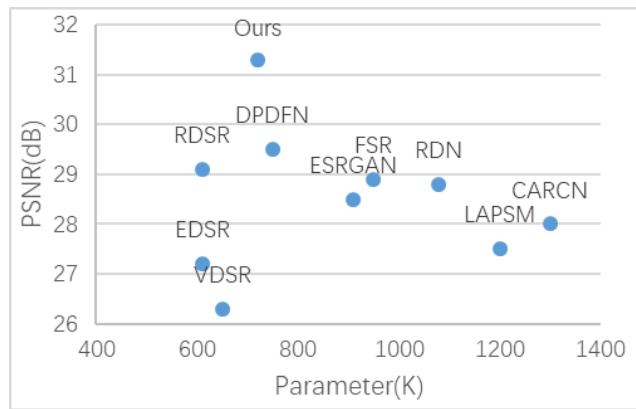


FIGURE 10. Comparison of model parameters.

The saliency maps generated by different methods are shown in Figure 9. These images are from the test images of four data sets. It can be seen that the visual effect of this method is better than other methods, and it can detect saliency objects more accurately. For example, the color contrast between the saliency area in the third image and the background is smaller, so this method can locate saliency area accurately. Most other methods detect saliency images with a lot of background interference; In the third image, the salient regions detected by most methods are incomplete. This method not only detects complete salient regions, but also has fine edges. Whether in the image with complex salient background (lines 1 and 2), salient regions and low background contrast (lines 3 and 4), this method can accurately highlight the whole salient region, And it has good visual effect.

4) MODEL PARAMETERS AND PERFORMANCE COMPARISON

Figure 10 shows the comparison results between the proposed carcn model and other advanced super-resolution models on set5 data set under the condition of 4 times super-resolution and corresponding parameters. The models included VDSR [38], EDSR [39], RDSR [40], DPDFN [41],

ESRGAN [42], FSR [43], RDN [44], LAPSM [45] and CARCN [46]. It can be seen that the PSNR index of the model proposed in this paper is the highest on the test data set, but the parameters are almost only 1 / 4 of EDSR, which also shows that the performance of the model proposed in this paper is better than other models.

IV. CONCLUSION

In this paper, based on SRGAN super-resolution network, artifacts are often produced in high-frequency details, and attention channel network is introduced. At the same time, because the deep learning convolutional neural network is too deep, the training time is too long and the super-resolution efficiency is low. In the discrimination model, the original discriminator of SRGAN is used to guide the training of super-resolution model. In the loss function, L1 loss function is used to further improve the image super-resolution visual effect. On the set5, set14 and bsd100 datasets, the PSNR is improved by 0.72db, 0.60db and 1.42db, respectively, compared with the SRGAN reconstruction results. However, due to the use of L1 loss function, the high-frequency features of the reconstructed image are ignored, which leads to the difference between the reconstructed image and the real image in some details. At the same time, due to the depth of the network, the calculation of the algorithm is too large, and the training efficiency is not high enough. Next, we will study how to further improve the effect of image reconstruction from these two aspects.

REFERENCES

- [1] L. Chao and H. Hongquan, “Lightweight image super-resolution based on convolutional neural network,” *Comput. Modernization*, vol. 3, no. 11, pp. 23–27, Dec. 2020.
- [2] Y. Chen, H. Zhang, L. Liu, J. Tao, Q. Zhang, K. Yang, R. Xia, and J. Xie, “Research on image inpainting algorithm of improved total variation minimization method,” *J. Ambient Intell. Humanized Comput.*, vol. 2, no. 2, pp. 233–244, Jun. 2021, doi: [10.1007/s12652-020-02778-2](https://doi.org/10.1007/s12652-020-02778-2).
- [3] Y. Chen, L. Liu, V. Phonevilay, K. Gu, R. Xia, J. Xie, Q. Zhang, and K. Yang, “Image super-resolution reconstruction based on feature map attention mechanism,” *Int. J. Speech Technol.*, vol. 51, no. 7, pp. 4367–4380, Jul. 2021, doi: [10.1007/s10489-020-02116-1](https://doi.org/10.1007/s10489-020-02116-1).

- [4] X. Shi, Z. Li, and J. Jiahuan, "Single image super-resolution algorithm based on multi-scale recursive dense network," *J. Shanxi Univ. (Natural Sci. Ed.)*, vol. 43, no. 4, pp. 719–726, 2020.
- [5] Y. Chen, L. Liu, J. Tao, X. Chen, R. Xia, Q. Zhang, J. Xiong, K. Yang, and J. Xie, "The image annotation algorithm using convolutional features from intermediate layer of deep learning," *Multimedia Tools Appl.*, vol. 80, no. 3, pp. 4237–4261, Jan. 2021.
- [6] D. Jiao, D. Wang, H. Lv, and Y. Peng, "Super-resolution reconstruction of a digital elevation model based on a deep residual network," *Open Geosci.*, vol. 12, no. 1, pp. 1369–1382, Nov. 2020.
- [7] L. H. Nian and X. Xin, "Cross resolution pedestrian recognition based on attention mechanism," *J. Beijing Univ. Aeronaut. Astronaut.*, vol. 47, no. 3, pp. 605–612, 2021.
- [8] Y. Chen, H. Zhang, L. Liu, X. Chen, Q. Zhang, K. Yang, R. Xia, and J. Xie, "Research on image inpainting algorithm of improved GAN based on two-discriminations networks," *Int. J. Speech Technol.*, vol. 51, no. 6, pp. 3460–3474, Jun. 2021, doi: [10.1007/s10489-020-01971-2](https://doi.org/10.1007/s10489-020-01971-2).
- [9] S. Zhongfan, Z. Zhenghua, and Z. Jianwei, "Super resolution reconstruction method with arbitrary Magnification Based on spatial meta learning," *Comput. Appl.*, vol. 40, no. 12, pp. 3471–3477, 2020.
- [10] Y. Chen, V. Phonevilay, J. Tao, X. Chen, R. Xia, Q. Zhang, K. Yang, J. Xiong, and J. Xie, "The face image super-resolution algorithm based on combined representation learning," *Multimedia Tools Appl.*, vol. 80, no. 20, pp. 30839–30861, Aug. 2021, doi: [10.1007/s11042-020-09969-1](https://doi.org/10.1007/s11042-020-09969-1).
- [11] H. Shuaikun, C. Honggang, Q. Linbo, and H. Chuanning, "Super resolution reconstruction of raw core image," *Inf. Technol. Netw. Secur.*, vol. 39, no. 10, pp. 1–6, 2020.
- [12] P. Shamsolmoali *et al.*, "Image synthesis with adversarial networks: A comprehensive survey and case studies," *Inf. Fusion*, vol. 72, pp. 126–146, Aug. 2021.
- [13] Y. Chen, J. Tao, L. Liu, J. Xiong, R. Xia, J. Xie, Q. Zhang, and K. Yang, "Research of improving semantic image segmentation based on a feature fusion model," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 24, pp. 176–187, 2021, doi: [10.1007/s12652-020-02066-z](https://doi.org/10.1007/s12652-020-02066-z).
- [14] L. Renpu, Z. Li, M. Chenhui, L. Xuan, and Z. Hao, "Infrared image super resolution reconstruction based on improved depth back projection network," *Infr. Technol.*, vol. 42, no. 9, pp. 873–879, 2020.
- [15] Y. Yong, W. Zheng, Z. Dongyang, and L. Jiaxiang, "Super resolution reconstruction algorithm based on progressive feature enhanced network," *Signal Process.*, vol. 36, no. 9, pp. 1598–1606, 2020.
- [16] W. Haitao, Y. Wenjie, and Z. Guanglei, "Mine image super-resolution reconstruction method based on online multi dictionary learning," *Automat. Ind. Mining*, vol. 46, no. 9, pp. 74–78, 2020.
- [17] W. Yan, L. Ang, and W. Shengquan, "Study on the influence of training set on automatic target recognition under super resolution of remote sensing image," *J. Chongqing Univ. Technol. (Natural Sci.)*, vol. 35, no. 2, pp. 136–143, 2021.
- [18] Y. Chen, L. Liu, J. Tao, R. Xia, Q. Zhang, K. Yang, J. Xiong, and X. Chen, "The improved image inpainting algorithm via encoder and similarity constraint," *Vis. Comput.*, vol. 37, no. 7, pp. 1691–1705, Jul. 2021, doi: [10.1007/s00371-020-01932-3](https://doi.org/10.1007/s00371-020-01932-3).
- [19] L. Ying, Z. Li, and L. qingfan, "Image super-resolution reconstruction based on spatial transformation network," *J. Xi'an Univ. Posts Telecommun.*, vol. 25, no. 5, pp. 45–49, 2020.
- [20] X. Haiping, X. Kaili, and Y. Haitao, "Research progress of image super-resolution method," *Comput. Eng. Appl.*, vol. 56, no. 19, pp. 34–41, 2020.
- [21] L. Xinguo, "Infrared image super-resolution reconstruction based on compressed sensing," *Univ. Electron. Sci. Technol., Chengdu, China, Tech. Rep.*, 9, 2020.
- [22] L. Feng, Z. Lin, and C. Xiaohui, "Research on low resolution face recognition algorithm for security monitoring scenes," *Comput. Appl. Res.*, vol. 38, no. 4, pp. 1230–1234, 2021.
- [23] P. Shamsolmoali, M. Zarepoor, E. Granger, H. Zhou, R. Wang, M. E. Celebi, and J. Yang, "Image synthesis with adversarial networks: A comprehensive survey and case studies," *Inf. Fusion*, vol. 72, pp. 126–146, Aug. 2021.
- [24] Y. Jian, W. Shanshan, and L. Yingwei, "Crowd counting model in public places based on image field division," *Comput. Appl. Res.*, vol. 38, no. 4, pp. 1256–1260, 2021.
- [25] X. Chao and Z. Hongyu, "Image super-resolution reconstruction method based on deep convolutional neural network," *Sensors Microsyst.*, vol. 39, no. 9, pp. 142–145, 2020.
- [26] C. Lijuan and Q. Wei, "Face image fuzzy texture recognition method based on super resolution laser imaging," *Laser Mag.*, vol. 41, no. 8, pp. 96–100, 2020.
- [27] W. Zhi, L. Anyi, and F. Jinxiong, "Super resolution reconstruction of remote sensing image based on dense convolution neural network," *Surveying Mapping Spatial Geographic Inf.*, vol. 43, no. 8, pp. 4–8, 2020.
- [28] H. Long, "Image super-resolution reconstruction based on improved neighborhood embedding and directed kernel regression," *Digit. Technol. Appl.*, vol. 38, no. 8, pp. 126–127, 2020.
- [29] X. Zhigang, Y. Juanjuan, and Z. Honglei, "Mural image super-resolution reconstruction algorithm based on multi-scale residual attention network," *Prog. Laser Optoelectron.*, vol. 57, no. 16, pp. 152–159, 2020.
- [30] L. Xingchen, J. Juncheng, Z. Li, and H. Qinhan, "Image super-resolution feature concentration network," *Comput. Eng. Appl.*, vol. 38, no. 6, pp. 24–33, 2021, [Online]. Available: <http://kns.cnki.net/kcms/detail/11.2127.tp.20200820.1009.014.html>
- [31] L. Lan, Z. Yun, D. Jia, and M. Shaobin, "Research on super resolution image reconstruction method based on improved residual subpixel convolution neural network," *J. Changchun Normal Univ.*, vol. 39, no. 8, pp. 23–29, 2020.
- [32] Z. Jian-Wen, Y. Qi-Ping, Q. Juan, Y. Xiao-Ping, and C. Zhi-Hong, "Single image super-resolution reconstruction using multiple dictionaries and improved iterative back-projection," *Optoelectron. Lett.*, vol. 15, no. 2, pp. 156–160, 2019.
- [33] W. Wei, H. Tao, L. Xinwei, S. Siwan, J. Xiaoming, and L. Junyuan, "Research on super-resolution reconstruction of white blood cell images," *Comput. Sci.*, vol. 48, no. 4, pp. 164–168, 2021.
- [34] S. Yongxiang, J. Bin, H. Yongxu, Y. Guisheng, L. Qingwu, and Z. Zhiliang, "Super-resolution reconstruction of infrared images based on deep learning," *Appl. Sci. Technol.*, vol. 47, no. 4, pp. 8–13, 2020.
- [35] Z. Yongmei, H. Ruimin, M. Jianzhe, and H. Lei, "High spatiotemporal fusion method of remote sensing based on deep learning and super-resolution reconstruction," *Comput. Eng. Sci.*, vol. 42, no. 9, pp. 1578–1586, 2020.
- [36] S. Yao, A. Onur, G. Ali, and S. K. Warfield, "Isotropic MRI super-resolution reconstruction with multi-scale gradient field prior," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 11766, 2019, pp. 3–11.
- [37] Y. Chang and B. Luo, "Bidirectional convolutional LSTM neural network for remote sensing image super-resolution," *Remote Sens.*, vol. 11, no. 20, p. 2333, Oct. 2019.
- [38] X. Deng and P. L. Dragotti, "Deep coupled ISTA network for multimodal image super-resolution," *IEEE Trans. Image Process.*, vol. 29, pp. 1683–1698, 2019.
- [39] N. Meng, H. K.-H. So, X. Sun, and E. Y. Lam, "High-dimensional dense residual convolutional neural network for light field reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 873–886, Mar. 2021.
- [40] L. Xu and X. Xu, "Regular super-resolution restoration algorithm based on fuzzy similarity fusion for a single haze image over the ocean," *J. Coastal Res.*, vol. 93, no. 1, p. 489, Sep. 2019.
- [41] Y. Yang, M. Zhu, Y. Wang, H. Yang, Y. Wu, and B. Li, "Super-resolution reconstruction of cell pseudo-color image based on Raman technology," *Sensors*, vol. 19, no. 19, p. 4076, Sep. 2019.
- [42] Y. Sui, O. Afacan, A. Gholipour, and S. K. Warfield, "Fast and high-resolution neonatal brain MRI through super-resolution reconstruction from acquisitions with variable slice selection direction," *Frontiers Neurosci.*, vol. 15, p. 709, Jun. 2021.
- [43] Y. Tao and J.-P. Muller, "Super-resolution restoration of spaceborne ultra-high-resolution images using the UCL OpTiGAN system," *Remote Sens.*, vol. 13, no. 12, p. 2269, Jun. 2021.
- [44] Y. Tao, S. Douté, J.-P. Muller, S. J. Conway, N. Thomas, and G. Cremonese, "Ultra-high-resolution 1 M/pixel CaSSIS DTM using super-resolution restoration and shape-from-shading: Demonstration over oxia planum on Mars," *Remote Sens.*, vol. 13, no. 11, p. 2185, Jun. 2021.
- [45] J. Rongzhao and W. Xiaohong, "Research on super-resolution reconstruction algorithm of image based on generative adversarial network," in *Proc. J. Phys., Conf.*, vol. 1, 2021, p. 1944.
- [46] L. Wan and G. Zhang, "Super-resolution reconstruction of unmanned aerial vehicle image based on deep learning," in *Proc. J. Phys., Conf.*, Jun. 2021, vol. 1948, no. 1, Art. no. 012028.