Journal of Network Intelligence
Taiwan Ubiquitous Information
©2021 ISSN 2414-8105 (Online)
Volume **6**, Number **2**, May **2021**

Image super-resolution reconstruction based on improved generative adversarial network

Ying-Li Wang

Department of Electronic Engineering
Heilongjiang University
No.74 Xuefu Road, Harbin, Heilongjiang, China
wangyingli@hlju.edu.cn

Xiao-Jing Li

Department of Electronic Engineering
Heilongjiang University
No.74 Xuefu Road, Harbin, Heilongjiang, China
18846077920@163.com

Hong-Bin Ma*

Department of Electronic Engineering
Heilongjiang University
No.74 Xuefu Road, Harbin, Heilongjiang, China
Corresponding author: mahongbin@hlju.edu.cn

Qi-Tao Ma

Faculty of Engineering
The Hong Kong Polytechnic University
11 Yucai Road, Hung Hom, Kowloon,Hong Kong,China
Jack_coldsweat@163.com

Qun Ding

Department of Electronic Engineering
Heilongjiang University
No.74 Xuefu Road, Harbin, Heilongjiang, China
qunding@aliyun.com

Matin Pirouz

Department of Computer Science
California State University, Fresno, USA
mpirouz@ieee.org

Received May 2020; revised October 2020

Abstract. *This article introduces the super-resolution reconstruction of images based on an improved generative confrontation network, improves the network structure of the generator, and proposes a super-resolution reconstruction algorithm for the recursive residual generation confrontation network. Its discriminator uses PatchGAN as the discriminator. The network solves the bottleneck of low feature information utilization and slow convergence of the generation countermeasure network in the super-resolution algorithm based on convolutional neural network. The reconstruction algorithm is compared with mainstream super-resolution reconstruction algorithms on standard data sets such as Set5 and Set14. The data shows that the algorithm can effectively improve the use of feature information, restore the details of low-resolution images, and improve the quality of image reconstruction.*

Keywords: image super-resolution; convolutional neural network; generative adversarial network; recursive residual network

---

1. **Introduction.** With the continuous informatization of society, the quality of images is extremely important. Generative confrontation networks have excellent results in image generation and resolution, and have high applicability in image super-resolution reconstruction. Single Image Super-Resolution (SISR) is one of the most difficult and challenging technologies in the field of image processing and computer vision. Its goal is to reconstruct a low-resolution picture (LR) into a high-quality high-resolution picture (HR) [1]. In a large number of electronic image applications, people often hope to obtain high-resolution images. High resolution means that the pixel density in the image is high, and it can provide more details, which are indispensable in many practical applications. For example, high-resolution medical images are very helpful for doctors to make a correct diagnosis; it is easy to distinguish similar objects from similarities by using high-resolution satellite images; if high-resolution images can be provided, computer vision the performance of pattern recognition will be greatly improved. Therefore, image super-resolution reconstruction has attracted more and more researchers' attention due to its wide application and practical significance.

The traditional image super-resolution reconstruction methods mainly include interpolation based methods, reconstruction-based methods [2]: convex set projection method (POCS), iterative back projection method (IBP), Bayesian analysis method, etc., and based on traditional machines Learning methods: exempble-based method, support vector regression method, etc. However, due to the loss of a lot of details and excessive blurring, traditional methods still fail to reconstruct satisfactory high-resolution images in many cases. In recent years, with the development of artificial intelligence technology, deep learning methods have been widely used in image super-resolution reconstruction. Super-resolution reconstruction based on deep learning [3] directly learns the end-to-end mapping function from low-resolution images to high-resolution images through convolutional neural networks (CNN). Dong et al. [4] proposed super-resolution reconstruction (SRCNN) based on convolutional neural networks, which is the first image super-resolution reconstruction method based on deep learning, learning the nonlinearity between LR and HR in an end-to-end method the mapping relationship, the structure is shown in FIGURE 1. Inspired by the VGG network, Kim et al. [5] proposed a very deep convolutional network (VDSR) to obtain a highly accurate reconstruction method. The network reaches 20 layers. In order to accelerate the convergence speed, a very high learning rate is used. Residual learning and gradient clipping to solve the gradient explosion problem. Ledig et al. [6] used Generative Adversarial Network (GAN) for SISR problem in 2017 and proposed a Super Resolution Generative Adversarial Network (SRGAN).Experiments show that this method can recover more high-frequency details.

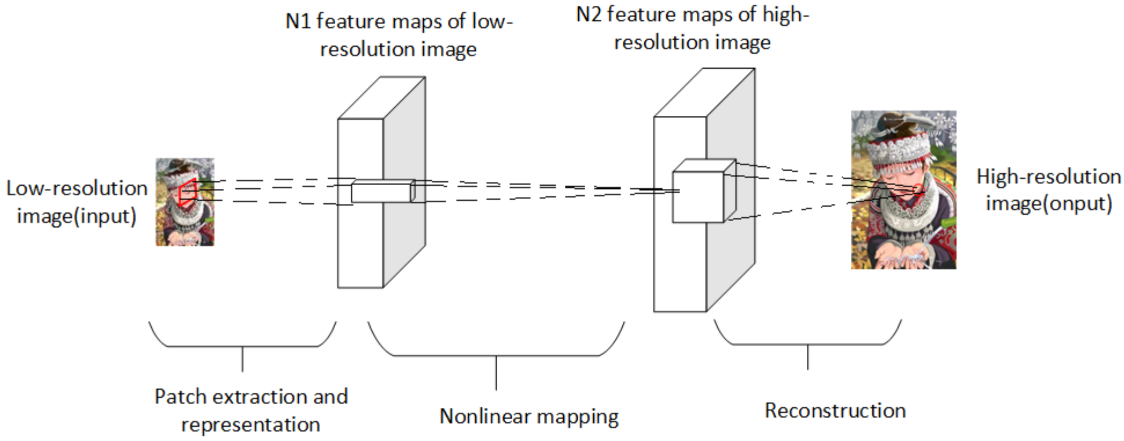

Figure 1. SRCNN network framework.

Although these super-resolution methods based on deep neural networks have made great breakthroughs in the quality of image reconstruction, there are still many shortcomings. Adopting a deep neural network with adversarial training (eg. SRCNN, VDSR) will make the reconstructed image too smooth, which is not in line with human perception of natural pictures. Adopting an anti-neural network method (such as SRGAN) can solve the problem of too smooth reconstructed images, which is more in line with human perception of natural pictures, but due to the very unstable GAN training, the reconstructed pictures are noisy and affect Picture quality reduces the generalization performance of the network. To overcome the above-mentioned deficiencies of the prior art, an image super-resolution reconstruction method based on an improved generating adversarial network is proposed.

2. Based and improved GAN image super-resolution reconstruction.

2.1. Overall structure. With the emergence of generative adversarial networks, Ledig et al. applied GAN to super-resolution reconstruction for the first time and proposed super-resolution reconstruction based on generative adversarial networks. Later, Lim et al. proposed the Enhanced Deep Super-Resolution (EDSR) network [7] modified the design of the residual block based on the structure of SRGAN and removed the batch normalization (BN) layer. Experiments have shown that the removal of the BN layer can save about 40% GPU memory space. This paper improves the network structure of the generator and proposes a super-resolution reconstruction algorithm for recursive residual generation against the network. Its discriminator uses PatchGAN as the discriminator network, which solves the use of feature information in the super-resolution algorithm based on convolutional neural networks. The low rate and the generation of the bottleneck of the slow convergence of the confrontation network. Based on the original SRGAN that can improve the quality of the reconstructed image, this algorithm processes the input image in the form of image blocks, which can reduce the amount of parameters and calculations and solve the problem of slow convergence. The overall flow chart is shown in Figure 2:

2.2. Residual network performance. The proposal of Deep Residual Network (ResNet) is a milestone event in the history of CNN images. Through experiments, we found that when the number of network layers reaches a certain number, the performance of the network will be saturated and the performance of the network will be increased. It will begin

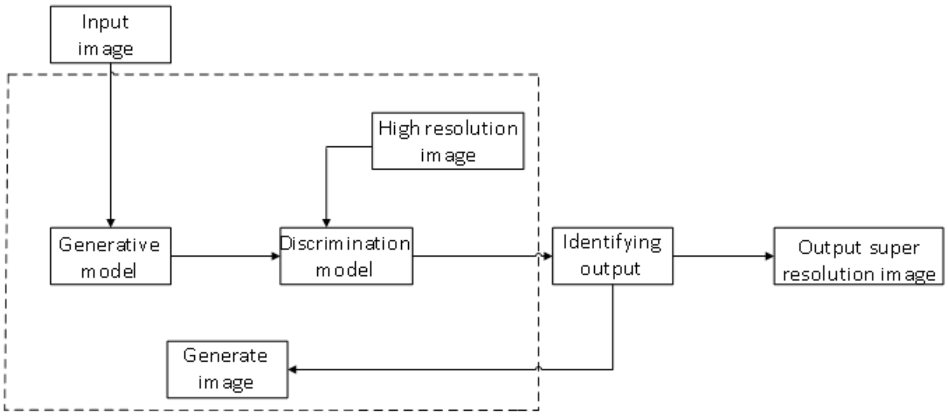

FIGURE 2. Overall flow chart.

to degenerate, but this degradation is not caused by overfitting, because we found that training accuracy and test accuracy are decreasing, which shows that when the network becomes very deep, the deep network becomes difficult to train.

The emergence of ResNet is actually to solve the performance degradation problem after the network depth becomes deeper. The basic structure is shown in Figure 3. For deep residual networks, it is generally assumed that the back layers of the network are all identity mappings. At this time, the network can be simplified into a shallow network. In order to find the identity mapping, these layers are generally used to fit the potential The identity mapping $H(x) = x$, but this is more difficult, so the network is designed as $H(x) = F(x)+x$, where x is the network input and $F(x)$ is the output of the convolutional layer. Therefore, we convert to learning a residual function $F(x) = H(x) - x$. As long as $F(x) = 0$, the identity mapping $H(x) = x$ is formed and the fitting residual is It's easier.

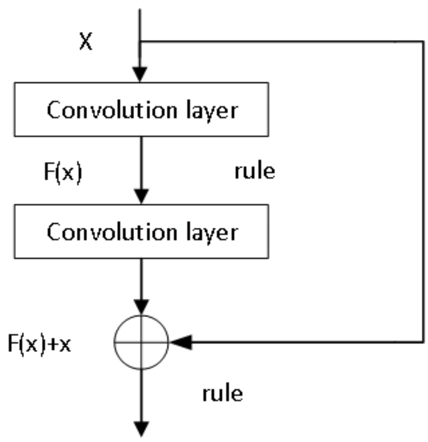

FIGURE 3. Basic structure of Renet network.

2.3. **Improved generator model.**

2.3.1. *Improved multi-cascade structure.* The feature expression ability of convolutional neural network is directly related to the quality of the generated image. The deeper the network model, the stronger the feature expression ability, which can generate better high-resolution images. But deeper models will also lead to slower generation speed and gradient disappearance and gradient explosion problems. This article uses a multi-cascade structure to enlarge the image to the required size step by step. The process of enlarging an image is broken down into multiple stages, each stage being enlarged by a certain ratio. For example, in the case of the required magnification ratio of 4, if the total number of stages N is 2, images with magnifications of 2 times and 4 times will be obtained, and if the total number of stages is 3, images of magnifications of 2 times, 3 times and 4 times will be obtained Image. In this way, high-resolution images of multiple sizes can be generated simultaneously in one model.

2.3.2. *Improved recursive residual network.* The Renet model proves that the deeper the network, the better the ability to extract features. The structure of the Resnet model is shown in FIGURE 4 (1). There are cross-layer connections to improve the stability of the model, but increasing the depth gradient will introduce some difficulties in gradient propagation. The VDSR model introduces residual learning into the super-resolution algorithm, as shown in FIGURE 4 (2). Because low-resolution images and high-resolution images have many similarities in image information, we can understand the difference between low-resolution images and high-resolution images, which is the residual. Add the residual and the low-resolution image as a matrix to obtain the high-resolution image. As shown in formula , to obtain the residual, x is the input resolution image. Residual learning can be well applied to graphics super-resolution reconstruction algorithms. Because the model only learns the residuals between the two, the model does not need to store the same parts of low-resolution and high-resolution images, reducing model parameters. Promote the transmission of the gradient and prevent the gradient from disappearing or exploding.

$$I^{HR} = I^{SR} + f(x) \tag{1}$$

Convolutional neural networks summarize its functions layer by layer through a deep structure and the functions of different convolutional layers play different roles. The function obtained by the low-level convolutional layer is essentially quite detailed edge or texture information [8]. The function obtained by the intermediate convolutional layer is essentially part of the object contour. The function obtained by the higher-level convolutional layer is essentially the entire structure of the object and different convolutional kernels have begun to form some differences between categories. The attributes obtained by the lower-level convolutional layers are more specific and general and the deeper the network layer, the more abstract the attributes obtained and the greater the difference between categories [9]. This proves that deeper networks tend to work better because the generated functions have stronger distinguishing capabilities. The high-level features have lost some specific details after several sampling operations, so we can not only recursively learn the functions generated by the last convolutional layer, but also recursively learn the functions generated by the shallow convolutional layer [10]. Recursive learning is shown in FIGURE 4 (3). The recursive model combines several layers of convolution functions to improve the expressive ability of the model. Based on the above analysis, this article combines recursive learning and residual learning, as shown in FIGURE 4 (4).

Image super-resolution is based on the frames of low-resolution images to generate corresponding high-resolution images. It can be considered that the input and output of the super-resolution neural network have very similar spatial distributions and the way that BN standardizes the properties of the intermediate convolutional layer is completely abandoned. Due to the need for the original spatial distribution, models with BN layers require convolutional layers or parameters to restore functionality during image reconstruction, which will reduce the quality of the generated images and increase the complexity of the model. Based on the above analysis, this article deletes the BN layer

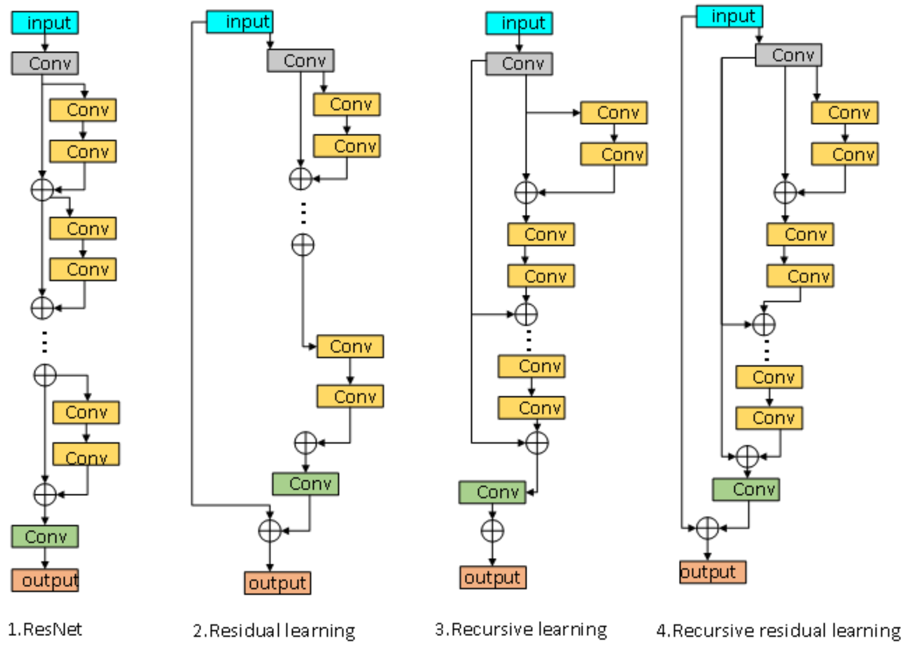

FIGURE 4. Schematic diagram of model structure improvement.

in the generator model to improve efficiency. The improved generator model structure is shown in FIGURE 5:

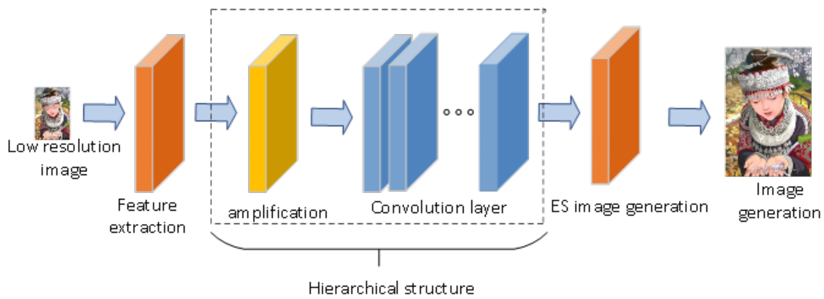

FIGURE 5. The improved generator model.

2.4. **Improved discriminator model.** The discriminator is partly inspired by Patch-GAN, a discriminator based on image blocks proposed by Isola et al. [11] and a 6-layer convolutional neural network is designed as the discriminator network. The discriminator network consists of alternating convolutional layers, batch normalization layers, activation layers and maximum pooling layers. The strides of the convolutional layer and maximum pooling layer are both set to 2. The batch normalization layer can be used to improve discrimination the generalization ability of the server network. The network structure of GAN under normal circumstances has been shown in some people's experiments that it is not suitable for image fields that require high resolution and high detail preservation. Compared with the original discriminator network of SRGAN, PatchGAN discriminator network is suitable for small-sized The image block is calculated, which greatly reduces the number of parameters and the amount of calculation and alleviates the problem of slow convergence of the original GAN network. Isola et al. have confirmed through experiments that when the image block size is greater than 70×70, the effect of image generation is very close to the judgment result of the actual image. Therefore, the algorithm in this paper uses the discriminator network part to improve the quality of the reconstructed image and solve the problem of slow convergence. The improved discriminator is shown in FIGURE 6:

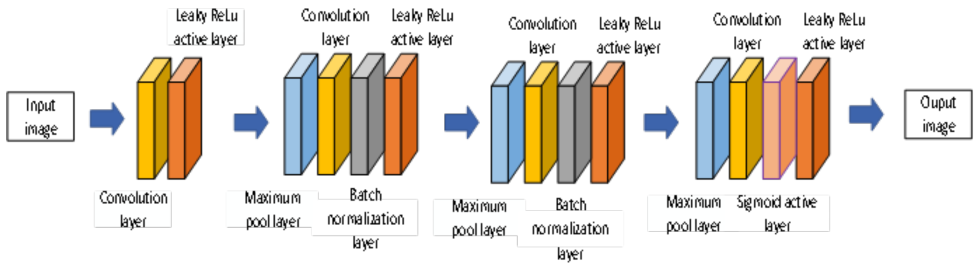

FIGURE 6. The improved discriminator.

2.5. **Loss function.** SRCNN, VDSR and DRCN models all use MSE error, but using the mean square error will produce an image with too smooth edges. In fact, the method based on the mean square error is essentially a high-resolution image corresponding to the low-resolution image and the results are averaged, so the mean square error function cannot find the potential from low-resolution images to high-resolution images Peak distribution. SRGAN uses perceptual loss and confrontation loss to enhance the realism of the recovered pictures. Perceptual loss is the feature extracted by the convolutional neural network [12]. By comparing the difference between the feature of the image generated by the convolutional neural network and the feature of the target image after the convolutional neural network, the meaning and style of the generated image and the target image more similar [13]. The cost function used by traditional methods is generally the minimum mean square error, is

$$l_{MSE}^{SR} = \frac{1}{r^2WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} \left(I_{x,y}^{HR} - G_{\theta_G} \left(I^{LR} \right)_{x,y} \right)^2. \tag{2}$$

The cost function makes the reconstruction result have a higher signal-to-noise ratio, but lacks high-frequency information and an excessively smooth texture appears [14]. Therefore, this article improves the cost function to

$$I^{SR} = I_X^{SR} + 10^{-3} l_{Gen}^{SR}. \tag{3}$$

The first part is a cost function based on content and the second part is a cost function based on adversarial learning. In addition to the minimum mean square error of the pixel space, the content-based cost function also contains a minimum mean square error based on the feature space [15]. This feature is a high-level image feature extracted using the VGG network:

$$l_{VGG/i,j}^{SR} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{ij}} \sum_{y=1}^{H_{i,j}} \left(\Phi_{i,j} \left(I^{HR} \right)_{x,y} - \Phi_{i,j} \left(G_{\theta_G} \left(I^{LR} \right) \right)_{x,y} \right)^2. \tag{4}$$

The cost function of adversarial learning is based on the probability of the discriminator output:

$$l_{Gen}^{SR} = \sum_{n=1}^{N} -\log D_{\theta_D} \left(G_{\theta_G} \left(I^{LR} \right) \right). \tag{5}$$

3. Experiment and data analysis.

3.1. **Experimental preparation.** The method in this article is based on the PyTorch framework and the experimental hardware parameters are: Intel Core i5- 4210U 1.70 GHz, NVIDIA GTX1060 4GB. This article uses 300 images used in VDSR, LapSRN and other models as training data sets and uses methods such as flipping and scaling to expand the data set. This article uses Set5, Set14 and BSD100 as test sets. The test results are compared with Bicubic, SRCNN, VDSR, DRCN, SRGAN and other algorithms. In order to ensure the accuracy of the experiment, all network models are tested at a scale factor of 4 times.

3.2. **Analysis of experimental results of recursive residual network.**

3.2.1. *Subjective effects.* To ensure the comparison effect, select a picture from Set5 and Set14 data sets and compare the actual reconstruction of each algorithm under the condition of 4 times magnification, as shown in FIGURE 7. Observe that the images in each test set can be observed after zooming in on the details. The algorithm proposed in this paper has better reconstruction image quality and the details of the texture are clearer.

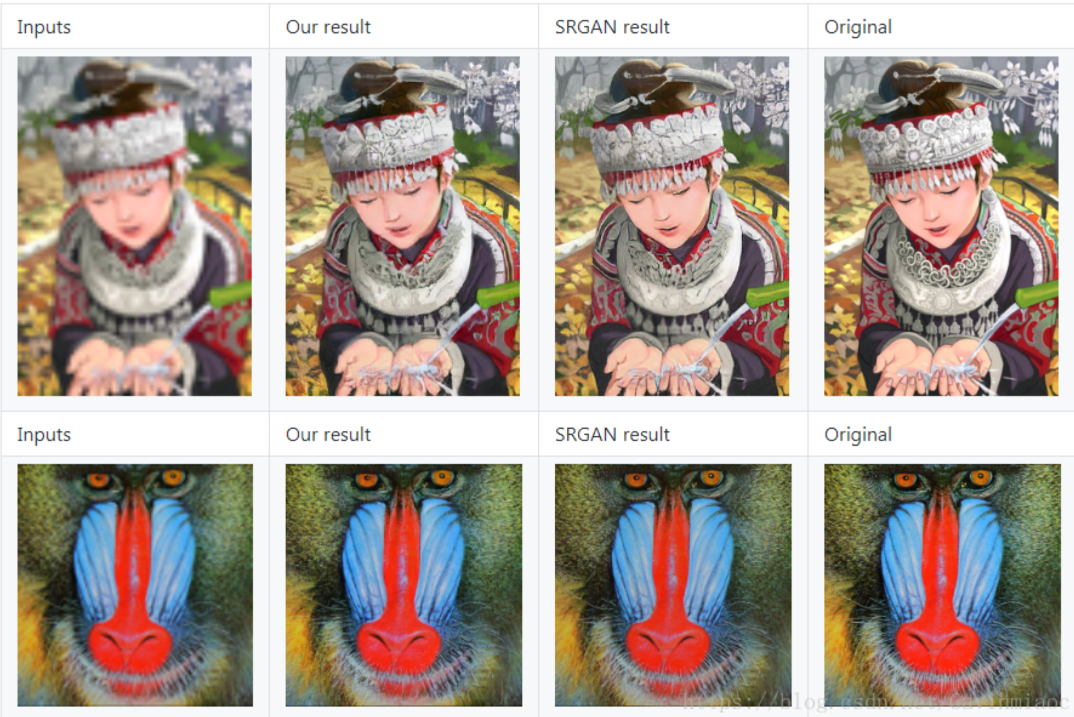

FIGURE 7. Reconstruction effect on Set4 and Set15 data sets

3.2.2. *Objective effect.* This paper uses two commonly used detection indicators in current image processing: Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) as objective evaluation indicators, thus more accurately indicating that the algorithm in this paper is compared with the superiority of other algorithms [16]. PSNR reflects the error between corresponding pixels of two images. The higher the value, the less the distortion of the output image and the better the image reconstruction quality. PSNR is defined by the maximum possible pixel value (L) and mean square error (MSE) of the image [17, 18]. The calculation formula is:

$$\mathrm{MSE} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (X(i,j) - Y(i,j))^2, \tag{6}$$

$$\mathrm{PSNR} = 10\log_{10}(\frac{(2^n-1)^2}{MSR}). \tag{7}$$

Among them, MSE represents the mean square error (MSE) of the current image X and the reference image Y and H and W are the height and width of the image. SSIM is an evaluation index indicating the similarity of two images. The closer the value is to 1, the closer the output image is to the original high-resolution image, that is, the better the reconstruction effect [19]. The calculation formula is:

$$\mathrm{SSIM}(x,y) = l(x,y)c(x,y)s(x,y). \tag{8}$$

Among them, $l(x,y)$ represents the brightness similarity operator, $c(x,y)$ represents the contrast similarity operator, and $s(x,y)$ represents the structural similarity operator. When the magnification factor is 4 times, under the reference data sets Set5, Set14 and BSD100, the PSNR and SSIM values of each algorithm are shown in Table 1 and Table 2. As can be seen from the table, on the set5, set14 and BSD100 data sets, the PSNR and SSIM values of each algorithm are compared. The PSNR and SSIM values of the algorithm in this paper are the best, so the algorithm of this paper is superior to other algorithms.

Table 1. PSNR value comparison

Algorithms	Set5	Set14	BSD100
Bicubic	30.01	27.11	26.18
SRCNN	30.22	27.34	26.49
DRCN	31.66	28.21	27.30
SRGAN	30.17	27.02	26.36
Algorithms in this work	32.55	28.88	27.84

Table 2. Comparison of SSIM values

Algorithms	Set5	Set14	BSD100
Bicubic	0.8611	0.7710	0.7189
SRCNN	0.8653	0.7751	0.7131
DRCN	0.8762	0.7812	0.7255
SRGAN	0.8721	0.7837	0.7387
Algorithms in this work	0.9122	0.7866	0.7579

4. **Conclusion.** Because the image reconstructed by the neural network-based super-resolution algorithm is too smooth, it does not meet the requirements of people's perception of the picture; the image reconstructed based on GAN is unstable and appears noise, which affects the quality of the picture. Combining adversarial networks, an image super-resolution reconstruction method based on improved generated adversarial networks is proposed. Experimental simulations prove that the algorithm in this paper is superior to other algorithms in both intuitive and objective PSNR and SSIM values. It can be seen that the algorithm in this paper can extract low-resolution image features more comprehensively, better improve the reuse rate of feature information, fully restore the texture information of the image and further improve the quality of image reconstruction, which is a better implementation. Algorithm for image super-resolution reconstruction.

REFERENCES

[1] Z. Y. Xing, E. L. Xiao and X. Z. Jian, Infrared image super-resolution reconstruction of dual discriminant generative confrontation network[J], *Small Microcomputer System*, vol. 41, no. 5, pp. 662-667, 2020

[2] Y. N. Wang, W. T. Li and J. Ren, Single-frame image super-resolution algorithm based on generative confrontation network [J], *Foreign Electronic Measurement Technology*, vol. 39, no. 1, pp. 26-32, 2020

[3] C. Li, Y. Zhang and C. H. Huang, Improved generation of confrontation network image super-resolution reconstruction[J], *Computer Engineering and Applications*, vol. 56, no. 4, pp. 191-196, 2020

[4] C. Dong, C. C. Loy, K. He et al., Image super- resolution using deep convolutional networks[J], *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295-307, 2015

[5] J. Kim, J. K. Lee and K. M. Lee, Accurate image super-resolution using very deep convolutional networks[C], *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646-1654, 2016

[6] C. Ledig, L. Theis, F. Huszar et al., Photo- realistic single image super-resolution using a generative adversarial network[C], *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4681-4690, 2017

[7] B. Lim, S. Son, H. Kim et al., Enhanced deep residual networks for single image super-resolution[C], *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 136-144, 2017

[8] K. K. Tseng, R. Zhang, C. M. Chen, and M. M. Hassan, DNetUnet: a semi-supervised CNN of medical image segmentation for super-computing AI service, *The Journal of Supercomputing*, https://doi.org/10.1007/s11227-020-03407-7 , 2020.

[9] E. K. Wang, C. M. Chen, M. M. Hassan and A. Almogren, A deep learning based medical image segmentation technique in Internet-of- Medical-Things domain, *Future Generation Computer Systems*, vol. 108, pp. 135-144, 2020

[10] E. K. Wang, C. M. Chen, F. Wang, M. K. Khan and S. Kumari, Joint-learning segmentation in Internet of drones (IoD)-based monitor systems, *Computer Communications*, vol. 152, pp. 54-62, 2020

[11] P. Isola, J. Y. Zhu, T. Zhou et al., Image-to-image translation with conditional adversarial networks[C], *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125-1134, 2017

[12] T. Y. Wu, X. N. Fan, K. H. Wang, C. F. Lai, N. X. Xiong and J . M . -T. Wu, A DNA computation based image encryption scheme for cloud CCTV systems, *IEEE Access*, vol. 7, pp. 181434-181443, 2019

[13] E. K. Wang, X. Zhang, F. Wang, T. Y. Wu and C. M. Chen, Multilayer Dense Attention Model for Image Caption, *IEEE Access*, vol. 7, pp. 66358-66268, 2019

[14] F. Q. Zhang, T. Y. Wu and G. Y. Zheng, Video salient region detection model based on wavelet transform and feature comparison, *EURASIP Journal on Image and Video Processing*, pp. 58, 2019

[15] T. Y. Wu, X. N. Fan, K . H. Wang, J. S. Pan and C. M. Chen, Security analysis and improvement on an image encryption algorithm using Chebyshev generator, *Journal of Internet Technology*, vol. 20, no. 1, pp. 13-23, 2019

[16] T. Y. Wu, X. N. Fan, K. H. Wang, J. S. Pan, C. M. Chen and J. M. -T Wu, Security Analysis and Improvement of An Image Encryption Scheme Based on Chaotic Tent Map, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 4, pp. 1050-1057, 2018

[17] T. Y. Wu, K. H. Wang, C. M. Chen and J. M. -T. Wu and J. S. Pan, A Simple Image Encryption Algorithm Based on Logistic Map, *Advances in Intelligent Systems and Computing*, vol 891, pp 241-247, 2018

[18] X. Huang, P. N. Matin and Q. Ding, Research on Image Encryption Based on Hyperchaotic System, *Journal of Network Intelligence*, Vol. 5, No. 1, pp. 10-22, 2020.

[19] F. Q. Zhang, T. Y. Wu, J. S. Pan, G. Y. Ding and Z. Y. Li, Human Motion Recognition Based on SVM in VR Art Media Interaction Environment, *Human-centric Computing and Information Sciences*, 9: 40, 2019