Research on Super-Resolution Reconstruction Algorithm of Image Based on Generative Adversarial Network

Rongzhao Jia and Xiaohong Wang*
Liaoning Petrochemical University, Fushun 113001, China

Corresponding author: *wangxiaohong@lnpu.edu.cn

Abstract. Image super-resolution is to use a series of algorithms to improve the original image resolution. The process of obtaining high-resolution image through some low-resolution images is image super-resolution reconstruction technology. There are two main research fields in super-resolution reconstruction, one is to restore the real details of the image, and the other is to not require too much detail, and only pay attention to the overall visual effect of the image. In this paper, an improved super-resolution reconstruction algorithm based on generative adversarial network is proposed. The network model and loss function are improved and optimized. The auxiliary VGG-19 network is used to extract the image features, and the extended convolution is used to expand the receptive field, which makes the image have a better reconstruction effect. Using DIV2k data set as training set and set5, set14, bsd100 data sets as test set, a series of experimental analysis is to prove the feasibility of the procedure method. Compared with the existing mainstream models, the perceptual effect of the image has been improved.

1. Introduction
Image has become an indispensable part of people's life. The procedure of obtaining high-resolution image through some low-resolution images is image super-resolution reconstruction technology. Super resolution reconstruction was mainly used in the field of satellite remote sensing at the earliest stage, and then it was widely used today [1].

The concept of super-resolution was proposed by Harris and Goodman [2,3] of NASA in the 1960s, and then in 1984, Huang and Tsai [4] proposed a algorithm to reconstruct images from multiple low-resolution images. In the 21st century, the super-resolution reconstruction model SRCNN [5,6] model (2015TPAMI) proposed by Dong et al. After that, Kim improved it and proposed a deep recursive convolutional neural network (DRCN) [7] (2016CVPR) model, which increased the network depth and reduced the network parameters. After that, the post generative countermeasure network (GAN) appeared. Inspired by it, ledig [8] and others proposed the similar algorithm (SRGAN) (2017CVPR) based on confrontation network.

2. Our Approach

2.1 Network Model
Figure 1 shows the generation network model used to generate SR images. First, the LR image will pass through a convolution layer and a relu activation function; then it will pass through 16 residual networks with the same layout. Each residual block has two convolution layers, the convolution kernel
size is \(3 \times 3\), including 64 feature maps, and the dilation convolution is used. The expansion rates of the two convolution layers are 3 and 2 respectively, so that the image can learn more feature information. Finally, the upper sampling part, through two sub-pixel convolution layers to improve the resolution.

**Figure 1.** Generate network model diagram

The discriminant network model is to judge the HR image and the SR image sample generated by the generated network. The network model is shown in Figure 2. As the number of network layers increases, the number of features increases and the feature size decreases. The activation function is leakyrelu \((\alpha = 0.2)\), and the maximum pooling layer is avoided. After the completion of the convolution function, the sum of the two sigmoid functions is used to generate the real feature of the image.

**Figure 2.** Discriminant network model diagram

### 2.2 Training Process and Loss Function

The training process of the discrimination network for generating the network is shown in Figure 3. In this paper, the training of the generating network is as follows: the LR image is transferred into the generation network to get the SR image, and then it is transferred into the VGG network to get its feature map. The discrimination result is obtained by using the feature map of the high-resolution image and compared with 1 to get Gan_loss. MSE was obtained by comparing the HR image with the SR image. Lose transfers the HR image and the SR image into the VGG network, and obtains the feature maps of the images. The VGG is obtained by comparing the feature maps of the images loss. The total loss is used for training, and the network parameters are updated after gradient feedback.
Generative countermeasure network is to enhance the perception quality of SR image by continuously improving the generation network to deceive the discrimination network. The objective function used in this paper increases the feature perception loss of VGG-19 network and the countermeasure loss of generative countermeasure network. The formula is as follow.

\[
L_G = 2 \times 10^{-6} L_{VGG} + L_{MSE} + 10^{-3} L_{GAN,G}
\]

\[
L_D = L_{GAN,D}
\]

\[
L_{VGG} = \frac{1}{W_i H_i} \sum_{i=1}^{W_i} \sum_{j=1}^{H_i} \left( \phi_{i,j}(I_{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I_{LR}))_{x,y} \right)
\]

\[
L_{MSE} = \frac{1}{r^2 WH} \sum_{x=1}^{rH} \sum_{y=1}^{rW} (I_{x,y}^{HR} - G_{\theta_G}(I_{LR}))_{x,y}^2
\]
The loss function not only retains the advantages of the existing methods, but also pays more attention to the visual effect of the image, so that the reconstructed image has more texture details and better visual effect.

2.3 Innovation and advantages
Compared with the traditional network, this method adds a VGG-19 network to assist the training of the whole network, and improves the loss function, so that the overall visual effect of the reconstructed image has been significantly improved, the more realistic texture effect is restored, and the image has more abundant High frequency information. Using more intensive residual structure and dilation convolution, the result of image reconstruction is more prominent, and the image reconstruction effect will be better. In traditional generative network, the input of discriminating network is picture, while the improved method in this paper inputs characteristic graph of SR image generated by network, which can make the generated network generate SR image more in line with the characteristics of HR image.

3. Experiments
The training set in the experiment is generated by DIV2k data set. The data set contains 800 images as training set and 100 images as verification set. Because the original image of DIV2k is too large, the network training speed will be too slow, so we need to process the data set. Cut all the images of the original dataset into the same size of 480 × 480. Starting from the 480 × 480 pixel image on the top left of the original image, a picture is obtained every 120 pixels, and the remaining part with a boundary less than 480 pixels is discarded. Each picture needs to be randomly cut to 128 × 128. Bicubic is used to down sample the image by 4 times. These low resolution and high resolution images constitute the data set.

Three data sets, set5 [9], set14 [10] and bsd100 [11], including 5 images, 14 images and 100 images, were used as test sets to further compare the model performance. Error back-propagation adopts adaptive moment estimation Adam optimizer, which has adaptive learning rate and better iterative effect than gradient descent method. The exponential decay rate $\beta$ of Adam's first-order moment estimation is 0.9, $\beta_2$ is 0.999. In addition, some super parameter settings, batch when the size is set to 16, the initial stage learning rate is 0.0001. When the training reaches 50000, 100000, 200000, 300000 times, the learning rate will be reduced to half of the previous one. The training batch size is set to 16.

In this paper, we do the following experiments to select the feature map of VGG network layer: VGG22 ($I = 2, j = 2$) refers to the second convolution layer before the second pooling layer. Other things are the same, that is, VGG22 is a feature map at a shallow level, and VGG54 is a feature map at a higher level.

It can be seen from Table 1 that the traditional convolution neural network has higher peak signal-to-noise ratio and structural similarity, but the visual effect of the image is much lower than that of the generative countermeasure network method. As shown in Figure 4, the effect of VGG54 is better than that of VGG22, VGG34 and VGG44, and is closest to HR chart. Finally, VGG54 ($I = 5, j = 4$) is selected as the feature extraction graph, its effect is the best.
Figure 4. Comparison of the effects of VGG in "monarch" diagram

Table 1. Comparison of PSNR, SSIM and MOS of VGG network on Set5/Set14 dataset

| Set5/Set14 | MSE       | VGG22     | VGG34     | VGG44     | VGG54     |
|------------|-----------|-----------|-----------|-----------|-----------|
| PSNR       | 30.33/26.87 | 29.81/26.44 | 29.68/26.28 | 29.72/26.31 | 29.66/26.12 |
| SSIM       | 0.8689/0.7608 | 0.8468/0.7509 | 0.8471/0.7481 | 0.8469/0.7468 | 0.8472/0.7466 |
| MOS        | 3.62/3.32   | 3.74/3.57   | 3.77/3.71   | 3.8/3.68   | 3.83/3.72   |

Figure 5. Comparison of GaN and RESNET on "face" images

The following is an experiment on whether to use the generative countermeasure network, just change the optimization goal. It can be seen from Figure 5 that the pictures using the generative countermeasure network have more high-frequency details than those without the generative network, the visual effect of the pictures is also better, and the reconstructed image file is larger. The reconstructed image is better than convolutional neural network in detail texture and visual...
perception. Network training, as shown in Table 2, it can be seen that on set 5 and set 14 test sets, the image perception effect of this method is slightly better than that of her existing methods. Compared with the original low-resolution image, the effect has been greatly improved, so the method in this paper is feasible.

### Table 2. Comparison MOS scores of "baboon", "face" and "Lenna" with different models

| MO   | SRCNN | VDSR | LapSRN | bicubic | OURS |
|------|-------|------|--------|---------|------|
| Baboon | 3.72  | 3.85 | 3.87   | 3.54    | 3.89 |
| Face   | 3.69  | 3.86 | 3.91   | 3.59    | 3.91 |
| Lenna  | 3.75  | 3.9   | 3.92   | 3.57    | 3.93 |
| Average| 3.72  | 3.87 | 3.9    | 3.57    | 3.91 |

4. Conclusion

The image reconstructed by this the method we tried has better visual effect with good PSNR and SSIM values, but there are still some areas to be improved. The research method in this paper is mainly limited by experimental equipment, and some abandonment has been made in the process of data set preprocessing. In the future, more high-performance equipment will be used to expand the number of training sets, increase the size of input image, and increase the total number of experimental iterations, which can make the image reconstruction effect better; although MOS more convenience to reflects the quality of image reconstruction, it will cost more time, the traditional PSNR and SSIM evaluation indicators are only for reference. Therefore, we need to develop stronger evaluation indicators.

### Reference

[1] Gonzalez R C. Digital Image Processing[M]. Electronic Industry Press, 2007: 1-30.
[2] Harris J L. Diffraction and Resolving Power[J]. Journal of the Optical Society of America (1917-1983), 1964, 54(7): 931-933.
[3] Goodman J W. Introduction to Fourier optics[M]. New York: McGraw-Hill, 1968: 1-5.
[4] Tsai R Y, Huang T S. Multiple frame image restoration and registration[J]. Advances in Computer Vision and Image Processing. Greenwich: JAI, 1984: 317-339.
[5] Dong C,. Learning a deep convolutional network for image super-resolution[C]European Conference on Computer Vision. Springer, Cham, 2014: 184-199.
[6] Dong C, et al. Image super-resolution using deep convolutional networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(2): 295-307.
[7] Kim J, Lee J K, Lee K M. Deeply-Recursive Convolutional Network for Image Super-Resolution[J]. 2015.
[8] Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[J]. 2016.
[9] M. Bevilacqua, A. Roumy, C. Guilleminot, et al. Low-complexity single-image super-resolutionbased on nonnegative neighbor embedding[J]. 2012
[10] R. Zeyde, M. Elad, M. Protter. On single image scale-up using sparse-representations[C]. International Conference on Curves and Surfaces, 2010, 711-730
[11] P. Arbelaez, C. Fowlkes, D. Martin. The berkeley segmentation dataset and benchmark[J]. Computer science department, 2007