Research and Improvement of Single Image Super-Resolution Based on Generative Adversarial Network

Tianji Zhang

1Department of Informatics, Beijing University of Technology, Beijing 100124, China
zhangtianji001@163.com

Abstract. This paper proposes a new image super-resolution method based on Generative Adversarial Network (GAN). Firstly, the algorithm model includes generating model and discriminant model, generating model to generate high-resolution image, discriminant model to distinguish the image true or false, the original image is true, and the generated image is false. Using alternate training method, the generated model and discriminant model achieve Nash equilibrium, and finally generate high-quality image. Compared with previous super-resolution method based on generative adversarial network (SRGAN), the following changes have been made: modifying the network structure, removing the unnecessary batch normalization layer in the standard residual block, deepening the network layer number and improving the loss function. The experimental results show that compared with the traditional bicubic interpolation method and compared with SRGAN, the proposed algorithm improves the actual image effect, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) in varying degrees.

1. Introduction

With the development of computer vision, the tremendous wealth contained in digital images has gradually been realized. Digital images play an increasingly important role in human life, and play an important role in industry, agriculture, medical treatment, criminal investigation and other fields. Usually, high-resolution images can provide us more detailed information than low-resolution images for human research. However, due to the factors such as hardware cost and technical defects of imaging equipment, the resolution of digital image obtained by people is low, and some details will be lost. Image super-resolution is a technical means of image processing. Through some algorithms, existing low-resolution images are converted into high-resolution images to restore the lost real details. Image super-resolution technology refers to the use of technical means to obtain corresponding high-resolution images from one or more low-resolution images. It is widely used in medical, traffic, criminal investigation, satellite images and other scenes.

Researchers have found that the effect of image super-resolution using the generation antagonism network in depth learning is improved significantly, which brings the effect of image super-resolution to a new height compared with the traditional method. CNN has achieved very good results in single frame resolution reconstruction, and can achieve high peak signal-to-noise ratio (PSNR). They take MSE as the objective function to minimize. Although they can achieve high peak signal-to-noise ratio (PSNR), the reconstructed image will be too smooth and lose details when the image down sampling ratio is high. SRGAN model[1] uses GAN for super-resolution reconstruction to improve the above problems. On this basis, three improvements are made: 1. The structure of the generated model is improved by removing the BN (batch normalization) layer in the standard residual blocks and
optimizing the network structure; 2. Deepening the number of residual blocks and increasing the depth of the network to improve the performance of the model; 3. The loss function is improved by adding the MSE loss based on pixels and the content loss function. The weight factor is added to the number, which improves the quality of the generated pictures. The experimental results show that the model is better than SRGAN in image quality.

2. Research on super-resolution of image based on generative adversarial network

Generative Adversarial Networks (GAN)[2] was first proposed by Good Fellow in 2014. Generative Adversarial Networks (GAN) is a two-player game based on two-player zero-sum game. The sum of interests of both players is a constant and will never change. There are two players in GAN, namely, generation model (G) and discriminant model (D). The generated model is a sample generator, which receives samples/noise, and then outputs imitation samples. The discriminant model is essentially a binary classifier, which receives samples and outputs the result of judgment, that is the authenticity of the samples. The direction of the generation model and the discriminant model is just the opposite. They are competing with each other. Our goal is that when the false samples generated by the generation model are input into the discriminant model, the output of the discriminant model is close to 0.5 value, that is, the discriminant model can not judge the authenticity of the image, and then Nash equilibrium is achieved.

![Generative Adversarial Networks Training](image.png)

Figure 1. Generative adversarial networks training.

The generation model and the countermeasure model are trained alternately and iteratively. The training process is shown in Figure 1.
Firstly, the discriminant model is trained. Suppose we have an initial generation model. We use this generation model to generate false samples by inputting noise. We mark the false samples label as 0, the real samples label as 1, and input the discriminant model. This is a simple supervised binary classification problem. We train the discriminant model.

For the training of the generated model, we need to use the false sample set and the generation-discrimination model, label the generated false sample set as 1, put it into the discriminant model, fix the parameters of the discriminant model, return the error to the generated model, and update the parameters of the generated model.

After the training of generation model, the generated sample set is closer to the distribution of the real sample set than the previous generated sample set. Continue to repeat the <1> <2> step operation. After several iterations, the authenticity of the generated sample will be greatly improved.

The above operation is actually a problem of maximum and minimum optimization, expressed by mathematical formulas:

\[
\min_G \max_D V(D,G) = \mathbb{E}_{x \sim \text{data}}[\log(D(x))] + \mathbb{E}_{z \sim \text{noise}}[\log(1 - D(G(z)))]
\]

2.1 Improved model structure

Increasing the depth and width of the network can improve the performance of the network very well. Many experiments show that for the same field, the model uses deep network and shallow network respectively, and the performance of deep network is generally the worst. It is assumed that the parameters of shallow network can be migrated to the first several layers of deep network, while only one layer behind deep network. An equivalent mapping can achieve the training effect of shallow network. For example, the common VGG network structure improves the network performance greatly by increasing the network depth on the basis of AlexNet. However, simply increasing the depth of the network will lead to the problem of gradient dispersion or gradient explosion. For the above problems, dozens of layers of networks can be trained through regular initialization and batch normalization (BN). However, new degradation problems arise, the number of network layers increases, but the accuracy of training set saturates or even decreases. This can not be understood as over-fitting, because over-fitting should be shown as better performance in training set.

Until the residual network and hopping connection were proposed by He[3], the network structure makes the deeper network easier to train, so the recognition accuracy can be improved by increasing the number of network layers. The basic structure of the residual block is shown in Figure 2. The residual network adds shortcut and skip connection to the original convolution layer to form the basic residual block. Let the original H(x) to be studied be represented as H(x)=F(x)+x. The residual structure of the residual network makes the learning of H (x) turn to F (x), and the learning of F (x) is easier than that of H (x). Residual network can effectively alleviate the degradation problem of deep network and improve network performance through layer-by-layer accumulation of residual block structure.
variance during training, and uses the estimated mean and variance of the whole training data set during testing. When the statistical data of training and testing data sets are very different, the BN layer will introduce discomfort artifacts, which limits the generalization ability of the model. Experiments show that the removal of BN layer helps to reduce computational complexity and memory usage, and improves the generalization ability of the model.

In terms of network depth, literature\[4-6\] shows that deeper networks can build high-complexity mapping and greatly improve network accuracy. In this paper, the number of residual blocks in the generator of the original model has been increased to 32. No fitting phenomenon has been found in the experiment, which improves the performance of the model to a certain extent. The structure of the improved generation model is shown in Figure 3.

![Figure 3. Structure of generating model.](image)

**2.2 Improved loss function**

The loss function (2.3.1) in the original SRGAN consists of two parts: content loss (2.3.2) based on VGG and countermeasure loss (2.3.3) based on generative-discriminative model. $D_{\theta_D}$ is the probability that an image belongs to a real high-resolution image, and $G_{BG}(I^{LR})$ is a reconstructed high-resolution image. This improves the texture details of the image, but the PSNR and SSIM indicators do not perform well. In this paper, the loss function is improved by adding a loss function based on MSE (2.3.4). Because the reconstruction effect of VGG is different in different layers\[7\] (see the experimental part), we also add a weight factor to the content loss function of VGG. The improved loss function (2.3.5).

\[
\begin{align*}
I_{SR} &= I_{SR}^{VGG} + 10^{-3}I_{Gen}^{SR} \\
I_{SR}^{VGG(i,j)} &= \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left( \Phi_{i,j}(I^{HR})_{x,y} - \Phi_{i,j}(G_{BG}(I^{LR}))_{x,y} \right)^2 \\
I_{Gen}^{SR} &= \sum_{n=1}^{N} -\log D_{\theta_D}(G_{BG}(I^{LR})) \\
I_{MSE}^{SR} &= \frac{1}{r^2WH} \sum_{x=1}^{r^W} \sum_{y=1}^{r^H} \left( (I^{HR})_{x,y} - (G_{BG}(I^{LR}))_{x,y} \right)^2
\end{align*}
\]
\[ I^S_R = I^S_{VGG} + \alpha I^S_{MSE} + \beta I^S_{Gen} \]  

(2.3.5)

3. Experiment

Experimental environment: Under Google Drive platform, the GPU model is Nvidia Tesla K80. Data sets are commonly used super-resolution data sets Set5, Set14, BSD100.

As mentioned above, the training sample set is put into the improved model to calculate various types of losses, and then all types of losses are added to the generated model losses. As described in Chapter 2, the reconstruction effect of VGG in different layers is different. The feature after multiple activation will be very sparse, resulting in low image performance. In addition, the brightness of the image will gradually decrease. The experimental results are shown in Figure 4. After trying different size category loss weights, the weights of VGG(1,1), VGG(1,2), VGG(2,1), VGG(2,2) are 0.45, 0.4, 0.1 and 0.05 respectively as the internal weights of perceptual loss. The weights of \( \alpha \) and \( \beta \) against loss were \( 4 \times 10^{-5} \) and \( 1 \times 10^{-3} \), respectively. As a result, the reconstruction quality is the best at this time.

The experimental visualization results are shown in Figure 5. The objective evaluation indicators of the experiment are in Table 1.

![Figure 4. Reconstruction effect of different layers.](image)

![Figure 5. From left to right: bicubic interpolation, SRGAN, mine, original HR image.](image)

|         | PSNR/SSIM     | bicubic | SRGAN  | mine   |
|---------|---------------|---------|--------|--------|
| **Set5**| 28.41/0.8231  | 29.34/0.8471 | 30.68/0.8741 |
| **Set14**| 25.98/0.7485  | 26.91/0.7796 | 27.36/0.7934 |
4. Conclusion
In this paper, image super-resolution technology based on generating countermeasure network is studied. On the basis of SRGAN, the model structure and loss function are improved. The experimental results show that the image quality generated by my model is improved compared with that generated by SRGAN. The new model not only retains the visualization effect of the original model for detail reconstruction, but also improves the objective evaluation of the image.

References
[1] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, Shi W. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 4681-4690).
[2] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.
[3] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).
[4] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014 Sep 4.
[5] Sun J, Shum HY, inventors; Microsoft Technology Licensing LLC, assignee. Image super-resolution using gradient profile prior. United States patent US 9,064,476. 2015 Jun 23.
[6] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. InProceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 1-9).
[7] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In European conference on computer vision 2016 Oct 8 (pp. 694-711). Springer, Cham.