Image super-resolution reconstruction algorithm based on EDSR and improved SRGAN

Liehao Song

Department of Internet of things, Beijing University of Posts and Telecommunications, Beijing, China

Songliehao2019@bupt.edu.cn

Abstract. Nowadays Image super-resolution reconstruction algorithm has become the dominant research aspect of computer vision, which is about generate high resolution composite image using low resolution original image. SRCNN(Super-Resolution Convolutional Neural Network) algorithm is the pioneer to apply deep learning to image super-resolution reconstruction, while SRGAN(Super-Resolution using Generative Adversarial Network) further explores the fusion of deep learning domain and image super-resolution reconstruction on grounds of SRCNN algorithm, and uses generative adversarial network to fit the extra information after image resolution improvement. Aiming at the defects of SRResNet generation network in SRGAN algorithm, this paper tries to improve the generator in its generation adversarial network by referring to the idea of ENSR(Enhanced Deep Residual Networks for Single Image Super-Resolution) algorithm, and obtains obvious effect. The improved algorithm has a significant improvement in performance.

Keywords: SRGAN, image super-resolution reconstruction, EDSR

1. Introduction

Images are now ubiquitous on every platform on the Internet. Compared with low-resolution images, high-resolution images usually have a better user viewing experience, the higher the resolution of the image, the more comfortable it will be for users to get valuable information in the image. Therefore, in general, web users want the image resolution to be as high as possible, and the network platform that can support higher resolution images will become more competitive in the market, due to that, the method that can provide more high-resolution images to network users at low cost has a very high application value.

However, actually, due to the constraints of imaging equipment, resolution of acquisition equipment, various physical factors during image acquisition and transmission, image degradation and other factors[1], Directly produce ideal high-resolution images with low cost and high efficiency is usually too difficult to realize. And image super-resolution reconstruction technology can help overcome the difficulty. Image super-resolution reconstruction refers to transform a low-resolution image into a higher-resolution image through an algorithm trained in advance[1]. Using the image super-resolution reconstruction algorithm, Problems such as low quality and serious noise pollution caused by various unfavorable conditions of image acquisition, image blur or transportation can be avoided[2]. If we successfully develop an image super-resolution algorithm suits requirement, we just need to transfer

low-resolution images, and transform them through the well-prepared image super-resolution reconstruction algorithm, then we can get the high-resolution images we need. In this way, we successfully circumvent constraints in direct transmission of high-resolution images, and can better meet the needs of users with the existing limited resources and conditions.

This paper first introduces the development history of image super-resolution reconstruction algorithm, then introduces SRGAN algorithm based on SRCNN algorithm, and introduces its structure and main defects in detail, as well as the basic role of Batch Norm operation in SRGAN algorithm, its negative impact on SRGAN algorithm, and the specific principle of reducing the performance of the algorithm. To solve this problem, we explore the influence of using EDSR algorithm, which Batch Norm operation is deleted, as the alternative of SRResNet to be the generator on the performance of SRGAN algorithm. Through experimental verification and testing with multiple datasets, we find that the performance of the algorithm is significantly improved.

2. Super resolution algorithm using deep learning

In 2006, Hinton et al.[3] first proposed deep learning. Deep learning can extract various abstract features of data and use these features to find rules, so as to enhance the performance of the algorithm. The most critical step on image super-resolution reconstruction is how we can make computer restore the information lost when the image resolution decreases as accurately as possible. Deep learning can be used to make the computer learn how to do this independently, which suits this very much, so that the trained computer can produce output images which are very close to the initial image. It makes deep learning extremely fit the need for the further development of image super-resolution reconstruction, also makes more and more researchers begin to try to explore the utilization of deep learning to create a new image super-resolution reconstruction algorithm with more superior performance[4], especially the convolution neural network, serves as an indispensable tool in the further development of the image resolution reconstruction[5].

In 2014, Dong C, et al.[6], as the pioneers, successfully utilized deep learning into image super-resolution reconstruction firstly[7], and their model is named as the SRCNN model, and successfully achieved a superior performance than most of the traditional methods, although the main principle of SRCNN is relatively simple. Since then, deep learning has set off a wave in super-resolution reconstruction. The algorithm firstly expand the low-resolution images to the expected resolution using bicubic interpolation to , and then transform them. Finally the output image with higher resolution is produced by a three-layer convolution neural network.
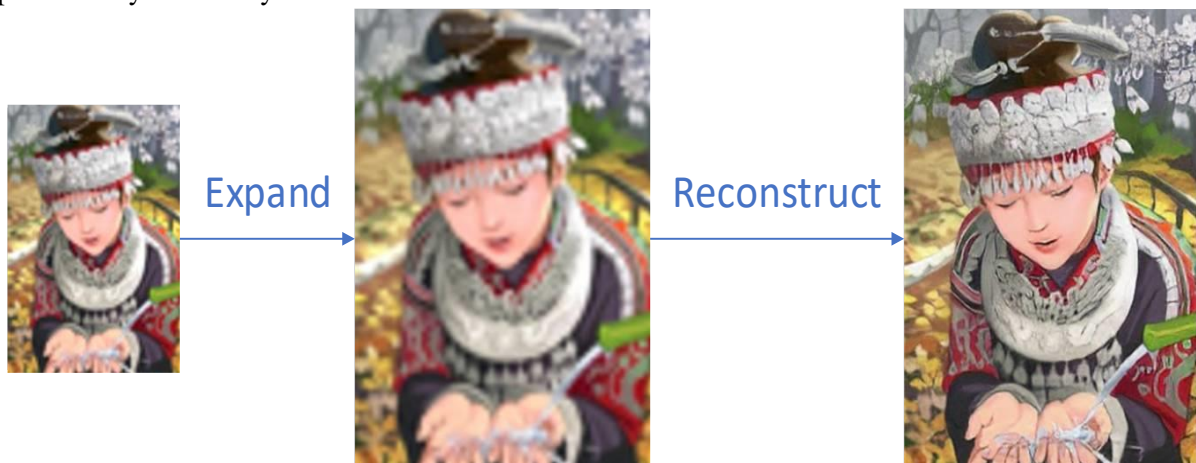

Figure 1. Structure of SRCNN.

Although the rational application of deep learning technology makes SRCNN algorithm outperform many traditional image processing algorithms at that time in performance, it also has an inevitable defect,

that is, it adopts MSE loss function as the final objective function. Misled by the MSE function, despite that the reconstructed image performs well on PSNR and SSIM, in fact, the reconstructed image does not well conform to the subjective feelings of human eyes. The author of the later SRGAN algorithm pointed out that this is because the objective function only pays attention to the differences between image pixels and ignores the overall structure of the image. Using this loss function to reconstruct the image will make the reconstructed image lose a lot of details and texture information, resulting in the actual image quality not as good as expected[7].

The calculation formula of MSE loss function is as follows:

$$MSEloss(x_i, y_i) = (x_i - y_i)^2 \tag{1}$$

With the purpose of avoiding the inherent defects of MSE loss function adopted by SRCNN, SRGAN algorithm uses the Generative Adversarial Networks(GAN) to create the extra information after the image resolution is improved, and defines a new perception objective function to calculate the loss.

In that paper[7], the author also proposed a comparison algorithm named SRResNet. Although it still choose the MSE as its loss function, SRResNet uses a deep enough residual convolution network model, Which makes it different from the past algorithm and can also achieve better reconstructed images in contrast with other image super-resolution reconstruction algorithms using MSE function.

3. Introduction of SRGAN

3.1. Generative adversarial network
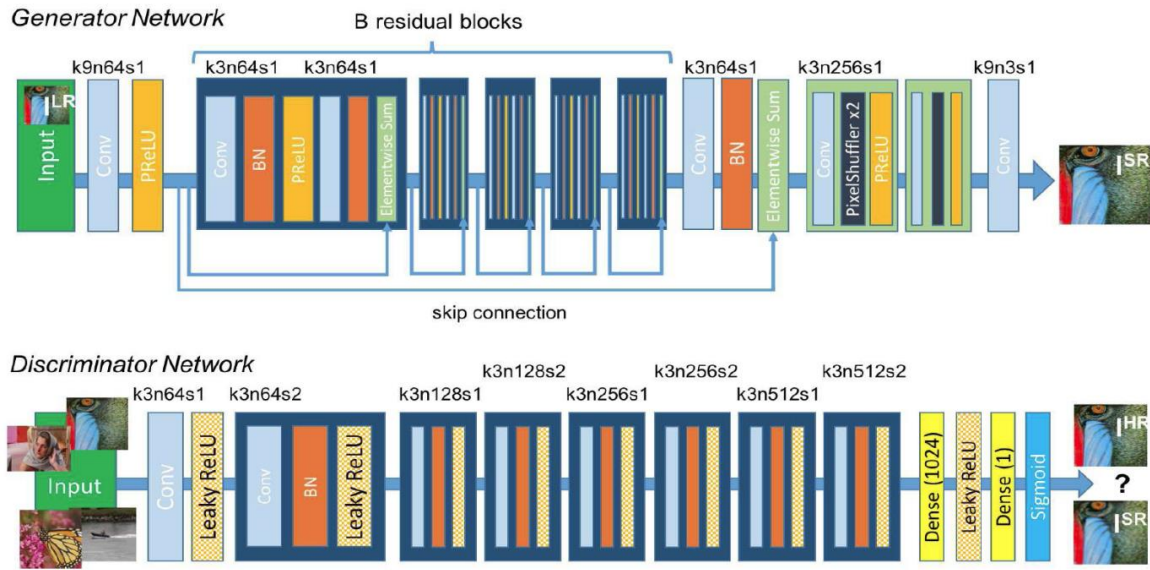The following image depicts the basic principle of GAN network:

Figure 2. The principle of GAN.

Compared with SRCNN algorithm, the biggest difference of SRGAN algorithm is that it adopts generative adversarial network to fit the extra information after the image resolution is improved. GAN[8] is mainly enlightened by the ideology of the game theory. GAN constructs two deep learning models named generator and discriminator. Firstly, generator network produces images through a specific convolution neural network; Then discriminator network determines whether the image is the original one or is reconstructed by generator network. The task of generator network is to produce realistic images as realistic as it can to confuse discriminator network and make it fail to judge; Similarly, discriminator network holds the task to check if the image is fake image generated by generator network.

3.2. The Loss function of SRGAN

SRGAN proposed a new loss function with more superior performance.It is the sum of two loss functions: content loss and adversarial loss.

The calculation formula of SRGAN's loss function is shown as below:

$$L_{VGG_{i,j}}^{SR} = \frac{1}{X_{i,j}Y_{i,j}}\sum_{m=1}^{X_{i,j}}\sum_{n=1}^{Y_{i,j}}(\phi_{i,j}(G_{\theta_G}(I^{LR}))_{m,n} - \phi_{i,j}(I^{HR})_{m,n})^2$$

(2)

$$L_{Gen}^{SR} = \sum_{n=1}^{N} -logD_{\theta_D}(G_{\theta_G}(I^{LR}))$$

(3)

$$L^{SR} = L_X^{SR} + 10^{-3}L_{Gen}^{SR}$$

(4)

Formula (2) Formula (3) and Formula (4) refer to the content loss, adversarial loss and the loss of SRGAN, respectively.

SRGAN defines content loss as the feature map of high-frequency features of VGG-54, while adversarial loss is the deficiency generated from adversarial process of GAN network, which mainly focuses on texture details[7].

4. Improvement of SRGAN algorithm

In the SRGAN algorithm, the SRResNet algorithm is adopted by the generator of GAN network achieved better results than other traditional image superscore algorithms at that time. However, the ResNet network structure was initially designed oriented to the high-level computer vision problem, while the image superresolution belongs to the low-level computer vision problem. It is likely not the best choice to apply the structure of ResNet network to image super-resolution reconstruction[9] without any modifications. In detailed the key problem of doing so is that the Batch normalization operation(BN) does not perform well in image super-resolution reconstruction.

BN[10] is an important technology in deep learning, which can not only simplify the process to train deeper networks and reduce the time of model convergence, but also prevent model become overfitting during training. However, BN does not perform well in terms of image super-resolution and image generation[11] Roughly speaking, BN is similar to a contrast stretch for an image. After the image is operated by BN, the image's color distribution will be normalized, which will destroy the original contrast information of the image. Therefore, BN will make the model training speed unstable, and even the loss function diverges at last. When BN is added to the neural network used in image super-resolution[12], the images output produced by the network are required not to do any modification with the color, contrast and brightness of input images, but only the resolution and some details are changed. Therefore, the existence of BN impairs the quality of the reconstructed images of convolutional neural network.

This paper proposes an improved generator model of SRGAN, which replaces the original SRResNet algorithm with its modified version: EDSR algorithm[13]. Compared with SRResNet, the most meaningful modification of EDSR[14] is to remove the superfluous module BN, so that the resources previously taken by BN can be used to other valuable parts[14]. After deleting the BN, EDSR can utilize the resources to deepen the convolution neural network layers and can extract more features of input images with the same computing resources, so that EDSR will have better performance than SRResNet. In the training period, EDSR trains the low-fold upsampling model first, so as to obtain the parameters extracted from the training of that model, which can be used to initialize and can accelerate the training period of the high-fold upsampling model, which usually will be time-consuming, and also after training the model will be competent to produce images with higher quality.

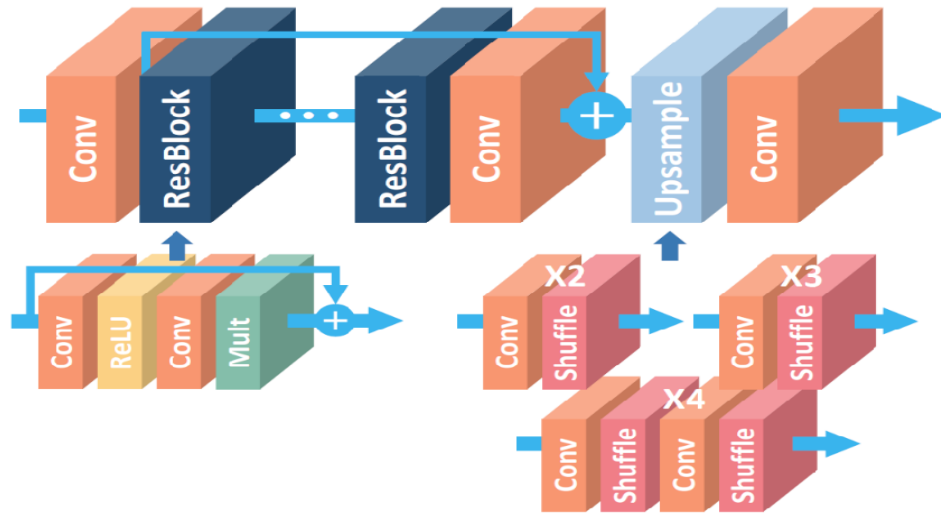The following is the network structure diagram of EDSR algorithm:

Figure 3. The structure of EDSR.

5. Result and analysis of experiment

5.1. Environment of experiment

The operating system of this experiment is Windows10, and the programming language is Python. The software PyCharm is used as the platform, and PyTorch deep learning framework is selected as the development model, which is implemented by Python 3.7 compilation.

The evaluation indexes adopted in this experiment are SSIM[5], PSNR and MOS. The closer the SSIM value is to 1, the more extraordinary the feature retention degree of the reconstructed HD output image is. The higher the PSNR value, the less the proportion of noise generated by HD output image reconstruction[9]. However, as mentioned above, we cannot judge the effect of image reconstruction only by SSIM and PSNR, so we introduce MOS as a complement. MOS is to ask the raters to rate the images from a minimum of 1 (low visual quality) to a maximum of 5 (high visual quality). Therefore, MOS depends on human eyes to judge the effect of image reconstruction. Generally speaking, MOS is more capable of reflecting photo-realistic images than SSIM and PSNR[15].

5.2. Introduction of dataset

For the training data, Urban100 is selected as the data set in this paper. First, 100 original high-resolution images in the data set are downsampled, after that the obtained low-resolution images are selected and preserved as the original data for the training of the model.

For the test data, Set5, Set14 and DIV2K datasets were selected as samples to judge the performance of each network models in this experiment.

5.3. Result of experiment

5.3.1. Experimental data of SRCNN, SRGAN and Ours. The Comparison of SRCNN, SRGAN and ours is shown in the following table:

Table 1. Data of each models.

Dataset	Evaluation index		HR	SRCNN	SRGAN	Ours
	SSIM	1	0.8617	0.8013	0.8412	
Set5	PSNR		∞	30.19	27.98	28.93
	MOS	4.42	2.99	3.83	4.04	
	SSIM	1	0.8188	0.7962	0.8093	
Set14	PSNR		∞	28.91	25.27	26.65
	MOS	4.38	2.86	3.64	3.85	
	SSIM	1	0.8388	0.8206	0.8293	
DIV2K	PSNR		∞	29.87	26.76	27.08
	MOS	4.46	2.73	3.54	3.63	

According to the table data, as mentioned above, under the influence of MSE loss function, SRResNet[7] has the highest PSNR and SSIM, but lower MOS, while SRGAN utilizes generative adversarial network. And loss function with better performance to improve the realism of reconstructed images. Although SSIM and PSNR are not as good as SRResNet, MOS is significantly better than SRResNet, indicating that SRGAN performs better than SRRestNet.

The image super-resolution reconstruction algorithm based on EDSR and improved SRGAN proposed in this paper performs similar to SRGAN with good visual sensory effects in terms of evaluation indexes. While keeping the subjective evaluation index MOS high, the objective evaluation indexes PSNR and SSIM values are improved, which proves that the proposed algorithm has better comprehensive performance.
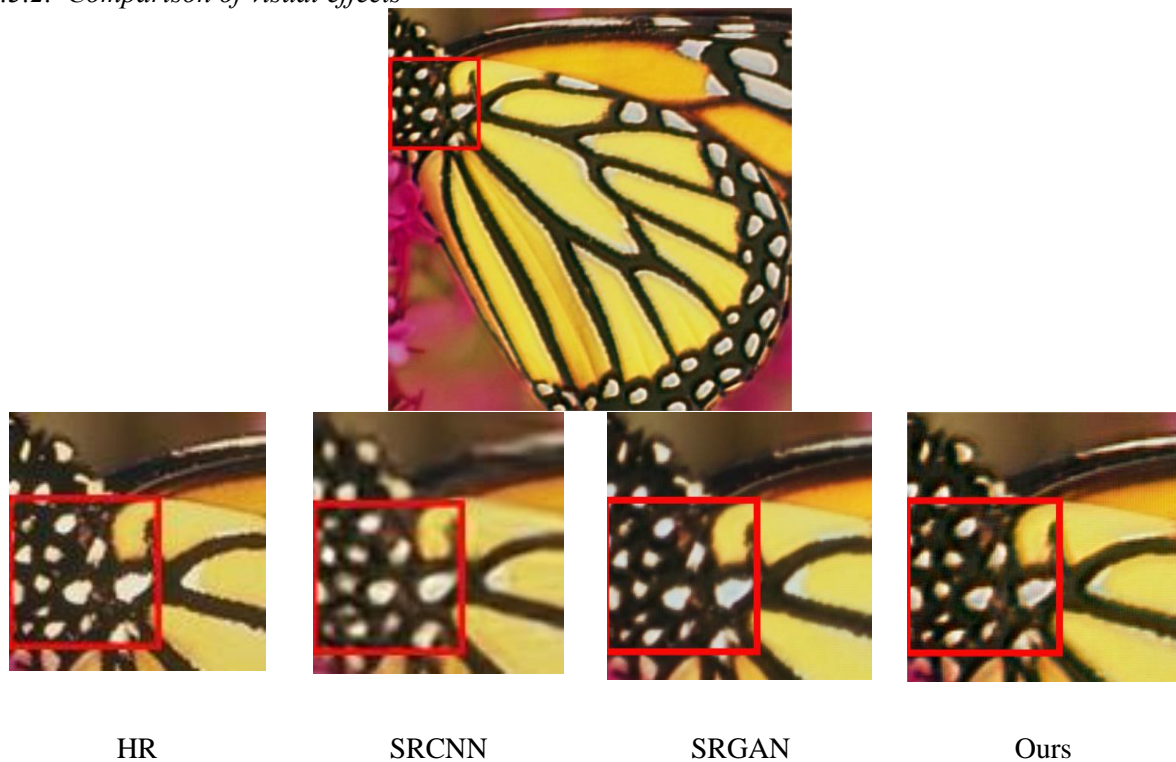
5.3.2. Comparison of visual effects

| HR | SRCNN | SRGAN | Ours |

Figure 4. Original high resolution image compared with reconstruct result of img_003 of set5 using SRCNN, SRGAN, and Our algorithm.

| HR | SRCNN | SRGAN | Ours |

Figure 5. Original High resolution image compared with reconstruct result of img_007 of set14 using SRCNN, SRGAN, and Our algorithm.

It can be easily found that on the above figures, Compared with our image super-resolution reconstruction algorithm based on EDSR and improved SRGAN, the original SRGAN has more serious ringing phenomenon.

Ringing effect[15] is one of the common factors influencing the recovery image quality, its[16] impact on the quality of restored image is serious, is must avoid factors in the process of image restoration, but in fact ringing effect in the process of recovery is almost inevitable, especially when the noise appears in the image, it will destroy the frequency characteristic of image, so that ringing effect is more significant, Exacerbating the degradation of image quality.

Batch Norm operation performs similarly to the noise, it will diverge the training model and normalize the color distribution on image, which destroys the original contrast information on image. In image super-resolution reconstruction , it makes ringing phenomenon in the image more serious and affects the reconstruction effect of the image. In this paper, ENSR algorithm is applied to the improvement of SRGAN algorithm, and this problem is better solved, so that the quality of reconstructed images is significantly improved.

6. Summary

In this paper, we propose our algorithm based on EDSR and improved SRGAN. The main change is that the generator is modified by referring to the EDSR algorithm. The main change is that the BN

operation in the SRGAN generation network is deleted. The improved algorithm can significantly improve the quality and running speed of reconstructed images. The result from experiment indicates that the reconstructed image of the modified algorithm is superior to the traditional SRGAN on PSNR, SSIM and MOS.

References

[1] Yongxin Liu;Tiantian Duan, Research on image super-resolution reconstruction technology based on deep learning; -《Technology and Innovation》- 2018-12-05, doi: 10.15913/j.cnki.kjycx. 2018.23.040

[2] Ruijuan Mu, Image super-resolution reconstruction based on Low-rank sparse Decomposition and dictionary learning [D]. Jinan: Shandong Normal University, 2017.

[3] Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh: A fast learning algorithm for deep belief nets. Neural computation 18.7 (2006): 1527-1554. [doi: 10.1162/neco.2006.18.7.1527]

[4] LI Z P. Super-resolution reconstruction based on deep learning [D]. Nanchang: Nanchang University, 2020.

[5] SUN Zhi-jun, XUE Lei, XU Yang-ming, WANG Zheng. Overview of deep learning, 1001-3695(2012) 08-2806-05), [doi: 10.3969 /j.issn.1001-3695.2012.08.002]

[6] Dong C, Loy CC, He KM, et al. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2):295-307, [doi: 10.1109/TPAMI.2015.24 39281]

[7] Qian Bin, Super resolution image reconstruction (algorithm principle, Pytorch implementation) - including complete code and data, https://blog.csdn.net/qianbin3200896/article/details/ 104181552?, 2020.

[8] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Proceedings of the 3rd International Conference on LearningRepresentations. San Diego, 2014.

[9] Tai Y, Yang J, Liu XM. Image super-resolution via deep recursive residual network. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu : IEEE , 2017 . 2790-2798.

[10] Aipiano, Another understanding of Batch Normalization, https://blog.csdn.net/aichipmunk/article /details/ 54234646, 2017

[11] Wang XT, et al. ESRGAN: Enhanced super resolution generative adversarial networks. Proceedings of European Conference on Computer Vision. Munich: Springer, 2018. 63–79.

[12] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, Kyoung Mu Lee: Enhanced Deep Residual Networks for Single Image Super-Resolution, arXiv:1707.02921 [cs.CV]

[13] ABlueMouse, From SRCNN to EDSR, this paper summarizes the development of end-to-end super-resolution methods in deep learning.https://blog.csdn.net/aBlueMouse/article/details/ 78710553, 2017.

[14] Yao Yutong; Tan Quan dog; JiGuangKai; Low Resolution Image Reconstruction Method Based on SRGAN, Network, Electronic Technology and Software Engineering,https://kns.cnki.net /kcms/detail/detail.aspx?filename=DZRU202202039&dbname=cjfdtotal&dbcode=cjfd&v=, 2022-01-15

[15] GAO Ju: Super-Resolution Reconstruction based on Ringing Effect Suppression [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2019.

[16] Chfe910, Ringing Effect, https://blog.csdn.net/chfe007/article/details/40785883, 2014.