Research on Super-resolution Image Based on Deep Learning

Tong Han
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 110955443@qq.com

Chuang Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 10423848513@qq.com

Li Zhao
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 332099732@qq.com

Abstract—Image super-resolution is a kind of important image processing technology in computer vision and image processing. It refers to the process of recovering high-resolution image from low-resolution image. It has a wide range of real-world applications, such as medical imaging, security and others. In addition to improving image perception quality, it also helps improve other computer vision tasks. Compared with traditional methods, deep learning methods show better recons-truction results in the field of image super-resolution reconstruction, and have gradually developed into the mainstream technology. This article will study the depth in the super resolution direction is important method of types of introduction, combed the main image super-resolution reconstruction method, expounds the depth study of several important super-resolution network model, the advantages and disadvantages of different algorithms and adaptive application scenarios are analyzed and compared, this paper expounds the different ways in the super resolution to liquidate, Finally, the potential problems of current image super-resolution reconstruction techniques are discussed, and the future development direction is prospected.

Keywords-Super Resolution;Neural Network; Convolu-tional Neural Network; Deep Learning

I. INTRODUCTION

Image resolution is a set of performance parameters used to evaluate the richness of detail information contained in an image, including time resolution, spatial resolution and color level resolution, which reflects the actual ability of imaging system to reflect object detail information [1]. High-resolution images usually contain greater pixel density, richer texture detail, and higher reliability than low-resolution images. However, in practice, due to the constraints of acquisition equipment and environment, network transmission medium and bandwidth, image degradation model itself and many other factors, we usually can not directly obtain the ideal high-resolution image with edge sharpening and block blur [2]. The most direct way to improve the image resolution is to improve the optical hardware in the acquisition system. However, it is difficult to improve the manufacturing process and the manufacturing cost is very high, so it is often too costly to solve the problem of low image resolution physically. Therefore, from the perspective of software and algorithm, the technology of image super-resolution recons-truction has become a hot research topic in many fields such as image processing and computer vision. Image super-resolution reconstruction technology refers to the restoration of a given low resolution image into a corresponding high resolution image by a specific algorithm. To be specific, image super-resolution reconstruction technology refers to the process of reconstructing high-resolution images from a

given low-resolution image by using relevant knowledge in digital image processing, computer vision and other fields through specific algorithms and processing processes [3]. It aims to overcome or compensate the problems of blurred image, low quality and insignificant region of interest caused by the limitations of image acquisition system or acquisition environment [4]. A simple way to understand super-resolution reconstruction is to change a small image into a large one to make the image more "sharp".

The existing super-resolution reconstruction algorithms are generally divided into three categories: interpolation based methods, which are simple but provide too smooth reconstruction images, lose some details and produce ringing effect; The modeling-based method has a better reconstruction effect than the interpolation method, but when faced with a large amount of calculation, the calculation process is time-consuming, difficult to solve and greatly affected by the amplification factor [5]. Based on the learning method, this kind of algorithm solves the problem sensitive to the scale scaling factor and has the best reconstruction effect, which is the mainstream direction of current research.

Currently studying mainstream of super-resolution reconstruction algorithm based on the depth of the main network model is divided into three categories: convolutional neural networks and generate network and another rise nearly new method is to explore the inverse transformation of GAN, the three kinds of models can be a good image of high frequency information, improve the resolution of the images, closer to the original image.

II. NETWORK MODEL OF SUPER-RESOLUTION RECONSTRUCTION METHOD

At present, superresolution networks based on deep learning can be divided into three categories: (1) supersegmentation methods based on convolutional neural network model; (2) superscore method based on generative adversarial network model; (3) it is explored that the inverse transform of GAN [6]; Each has advantages and disadvantages

A. Superdivision method based on convolutional neural network direction

1）SRCNN：

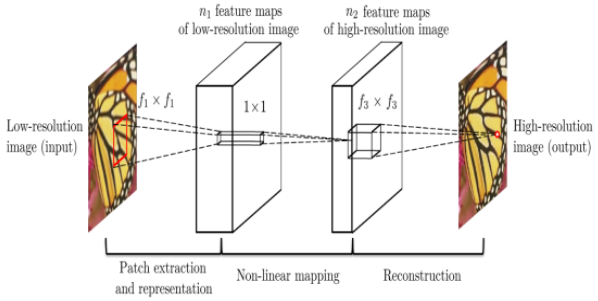

Figure 1. SRCNN network model.

SRCNN network consists of three modules: feature extraction, nonlinear mapping and final reconstruction [7]. These three modules correspond to three convolution operations. The first layer of CNN: feature extraction of input images. The purpose of block precipitation and representation is to obtain a series of feature maps from the input image Y. The second layer of CNN: nonlinear mapping of features extracted from the first layer. The nonlinear mapping corresponds to "convolution activation" operation [8]. The third layer of CNN: reconstruct the mapped features to generate high-resolution images. The Reconst-ruction process is also a convolution operation, but there is no activation function here [9].

SRCNN proposes to use conventional conv-olution collocation activation function to simulate the feature encoding process of the traditional SR method. From feature extraction to the final reconstruction, the convolution operation is used, which is very concise and efficient.

Loss function: SRCNN uses the mean squared error as a loss function to evaluate the difference between the network output and the true label.

The author conducted comparative experiments on 91images and mageNet datasets respectively. With the increase of iteration times, SRCNN obtained higher PSNR on ImageNet, which indicates that the increase of data volume can improve that the performance of the network [10].SRCNN first proposed to simulate the feature encoding process of the traditional SR method by

using a series of conventional convolution collocation activation functions. Compared with the traditional SR method, SRCNN is a simple end-to-end learning method with better network performance and inference speed.

2）ESPCN

The core concept of ESPCN is the subpixel convolutional layer. As shown in the figure above, the input of the network is the original low-resolution image. After passing through two convolutional layers, the feature image obtained has the same size as the input image, but the feature channel is r2(r is the target magnification of the image). The r2 channels of each pixel are rearranged into an r*r region, corresponding to an r*r size subblock in the high-resolution image, so that the feature image of size R 2*H*W is rearranged into a high-resolution image of size 1*rH*rW. Although this transformation is called sub-pixel convolution, there is no convolution operation. Through the use of sub-pixel convolution, in the process of image amplification from low resolution to high resolution, the interpolation function is implicitly included in the previous convolutional layer and can be automatically learned [11]. The image size is transformed only at the last layer, and the previous convolution operation is carried out on low-resolution images, so the efficiency is higher.

B. Based on generative adversarial network direction

1）SRGAN:

In this paper, generative adversarial network was used to solve the super resolution problem [12]. It is mentioned in this paper that the mean square deviation is used as the loss function when training the network. Although high peak signal-to-noise ratio can be obtained, the recovered images usually lose high-frequency details, which makes people unable to have good visual perception. SRGAN uses perceptual loss and adversarial loss to enhance the realism of the recovered images. Perceptual loss is the feature extracted by using convolutional neural network. By comparing the features of the generated image after convolutional neural network and the features of the target image after convolutional neural network, the generated image and the target image are more similar in semantic and style [13]. The original GAN text gives an example: the generation network G is the person who prints counterfeit banknotes, and the discrimination network D is the person who detects counterfeit banknotes. G's job is to make the counterfeit money he prints as much as possible to fool D, and D is to distinguish as best as possible whether the money he gets is the real money in the bank or the fake money printed by G. In the beginning, G was not good enough, and D was able to point out what was wrong with the bill. G After each failure to carefully summarize experience, strive to improve themselves, progress every time. Until the end, D could not judge the authenticity of the banknotes [14]. The work of SRGAN is that G network generates high resolution images from low resolution images, and D network determines whether the obtained image is generated by G network or the original image in the database. When the G-net can successfully fool the D-net, then we can complete the super resolution with this GAN.

In this paper, the SRResNet (the generating network part of SRGAN) is optimized by mean square error, and the results with high peak signal-to-noise ratio can be obtained. By calculating the perception loss on the high-level features of the trained VGG model to optimize SRGAN, and combining the discriminant network of SRGAN, the results with realistic visual effects can be obtained, although the peak signal-to-noise ratio is not the highest.

SRGAN provides a new loss function. In previous SR, MSE loss function is used to teach the network how to realize LR to HR, but this will smooth the details of the image. Although the PSNR is very high, the human eye does not have a good visual sense. GAN objective function can be defined as shown in Equation (1):

$$\min_{\theta_G}\max_{\theta_D} V(D,G) = E_{x\sim Ddata(x)}[logD(x)] + E_{x\sim D_Z(Z)}[log(1-D(G(x)))]. \quad (1)$$

Fixed, to learn adjustments, that is, in order to train a discriminator network. Then, we fix the discriminant network parameters to learn to adjust the parameters of the generating network. Its purpose is to make the parameters of the generating network as large as possible by adjusting them. That is to say, the training of the generator network is to make the output result of the discriminant network output a high score, so as to deceive the discriminator. Therefore, we can see that after the generator becomes stronger, the next discriminant network will continue to become stronger, increasing the ability to distinguish between true and false. The generator, in turn, will continue to increase the score of the fake (the output of the input after G) in the discriminator, and then the discriminator will continue to improve and iterate, and the two will fight and grow each other, and finally the trained generator network will be the network we want.

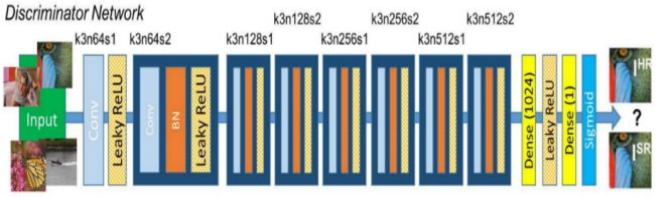

Figure 2. SRGAN generator and discriminator network structure diagram

SRGAN loss consists of two parts: content loss and adversarial loss, which are weighted and summed with a certain weight, as shown in Equation (2).

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR}$$
$$l_{MSE}^{SR} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta G}(I^{LR})_{x,y})^2 \quad (2)$$

In this paper, we define the VGG loss based on the ReLU activation layer of the pre-trained 19-layer VGG network to obtain the Euclidean distance between the feature representation of the image and the reference image. The feature map of a certain layer is proposed on the trained vgg, and this feature map of the generated image is compared with that of the real image.

$$l_{VGG/i,j}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\varphi_{i,j}(I^{HR})_{x,y} \frac{n!}{r!(n-r)!} - \varphi_{i,j}(G_{\theta G}(I^{LR}))_{x,y})^2 \quad (3)$$

Equation (3) generates adversarial loss: generates a data distribution that the discriminator cannot distinguish. $G_{\theta G}(I^{LR})$ represents the probability that the image generated by the generator will be a natural image by the discriminator.

$$l_{Gen}^{SR} = \sum_{n=1}^{N} -logD_{\theta D}\left(G_{\theta G}\left(I^{LR}\right)\right) \quad (4)$$

Equation (4) is an improved generator loss function proposed in this paper. To minimize this expression is to maximize the probability that the generated image given to the generator by the discriminator is true. The goal is to fool the discriminant network by producing a high discriminant value.

2）ESRGAN:

In this paper, some improvements are made on the basis of SRGAN, including improving the structure of the network, the decision form of the adjudicator, and replacing a pre-trained network for calculating the perceptual loss. Three key parts of SRGAN are studied in detail: network structure; adversarial loss; Perceptual domain loss. And improve each item to get ESRGAN. ESRGAN makes several major improvements to SRGAN:

a) Introduce changes to the generator architecture.

b) The improvement of adversarial loss is mainly the use of elativistic GAN to make relative realness instead of the absolute value.

c) Perceptual loss is composed of features before activation (previously activated features).

d) Pre-train the network to optimize PSNR first, and then use GAN to fine-tune it.

To be specific, the paper proposed a resist-in-residual Dense Block (RRDB) network unit [15], in which the BN (Batch Norm) layer was removed.

In addition, let the discriminator predict the truth of the image rather than whether the image "is a fake image". Finally, the perceptual domain loss is improved by using pre-activation features, which can provide stronger supervision for luminance consistency and texture recovery. With the help of these improvements, ESRGAN gets better visual quality as well as more realistic and natural textures.

In the generator part, the author makes several changes to the generator G by referring to the SRResNet structure as the whole network structure: remove all BN layers; Turn the original block into residual-in-residual Dense Block (RRDB) which combines multi-layer Residual network and dense connections. Removing the BN layer has been shown to improve performance and reduce computational complexity [16].

The discriminator tries to estimate the probability that the real image is relatively more realistic than the fake image. In this paper, the loss function of the discriminator is defined as Equation (5):

$$L_D^{Ra} = -E_{Xr}\left[log\left(D_{Ra}\left(x_r, x_f\right)\right)\right] - E_{Xr}\left[log\left(1 - D_{Ra}\left(x_f, x_r\right)\right)\right] \quad (5)$$

The adversarial loss function of the corresponding generator is shown in Equation (6) :

$$L_G^{Ra} = -E_{Xr}\left[log\left(1 - D_{Ra}\left(x_r, x_f\right)\right)\right] - E_{Xf}\left[log\left(D_{Ra}\left(x_f, x_r\right)\right)\right] \quad (6)$$

Xf Is the image generated by the generator, Xr Is the input low resolution image.

The generator benefits from the gradient between the generated data in adversarial training and the actual data, and this adjustment enables the network to learn sharper edges and more detailed textures [17].

For generator G, its loss function is shown in Equation (7):

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1 \quad (7)$$

For the discriminator, its loss function is shown in Equation (8):

$$L_D = L_D^{Ra} = -E_{x_r}[log(D_{Ra}(x_r, x_f))] - E_{x_f}[1 - log(D_{Ra}(x_r, x_f))]. \quad (8)$$

The ESRGAN proposed in this paper makes improvements on the basis of SRGAN, including removing BN layer, replacing the basic structure with RDDB, improving the discriminator discrimination target in GAN, and using the features before activation to form the perceptual domain loss function. Experiments have proved that these improvements are effective in improving the visual effect of the output image [18].

In addition, the authors also use some techniques to improve the performance of the network, including scaling of the residual information and smaller initializations. Finally, the authors use a network interpolation method to balance the visual effect of the output image with the PSNR and other index values.

3）REAL-ESRGAN

In the problem of single image super-resolution, many methods use the traditional Bicubic method to achieve downsampling, but this is different from the real world downsampling situation, which is too single.

Blind super resolution is designed to recover unknown and complex degraded low resolution images. According to the different down-sampling methods used, they can be divided into explicit modeling and implicit modeling. Explicit modeling: the classical degradation model consists of blur, down-sampling, noise and JPEG compression. However, the downsampling model in the real world is too complex to achieve the ideal effect through the simple combination of these methods.

Implicit modeling: It relies on learning the data distribution and using GAN to learn the degradation model, but this method is limited by the data set and cannot generalize well to the images distributed outside the data set. In the real world, image resolution degradation is usually a complex combination of many different degeneracies. Therefore, the network extends the CLASSICAL first-ORDER degradation model to the REAL-WORLD higher-order degradation model by using multiple repeated degradation processes, each of which is a CLASSICAL degradation model. However, in order to balance simplicity and effectiveness, the second-order degradation model is actually used in the code. However, because the high-order degradation model is adopted, the degradation space is much larger than that of ESRGAN, so the training is more challenging. So the network made two changes based on ESRGAN:

Use U-Net discriminator to replace VGG discriminator used in ESRGAN;

spectral normalization was introduced to make training more stable and reduce artifacts.

The authors of the article made the following study, A higher-order degradation model and sinc filters were proposed to model common ringing and overshoot artifacts.

Some basic changes (for example, the U-Net discriminator with spectral normalization) have been adopted to increase the discriminator's capabilities and training stability.

Real-esrgan trained on purely synthetic data is able to recover most real-world images, with better visual performance than previous works, and is more practical in the Real world. Disadvantages of Gan: Superresolution neural network based on gan has been a relatively mature scheme. However, such a method requires the generator not only to extract and retain the structure information of LR images, but also to generate as realistic and high-resolution texture information as possible, which is difficult for the generator, especially in high-magnification tasks. The resulting problems include but are not limited to: The generated image is blurred and the generated image has false texture information [19].

C. GAN Inversion

1) PULSE

The main purpose of single image superresolution is to construct a high resolution (HR) image from the corresponding low resolution (LR) input. In previous approaches, which are usually supervised, the training target is usually measured for the pixel-oriented average distance between superresolution (SR) and HR images. Disadvantages: Optimizing such metrics often leads to ambiguity, especially in areas of high variance (detail). We propose another scheme to simulate the super-resolution problem based on creating realistic SR images with the correct reduction in scale. In this paper, we propose a novel super-resolution algorithm to solve this problem, PULSE (photo upsampling through latent space exploration), which can generate high-resolution, realistic images that have not been seen in the literature before. It does this entirely in a self-supervised manner and is not limited to the specific degradation operators used during training [20], which is different from previous approaches that require training of the database of LR-HR image pairs for supervised learni-ng.Instead of traversing the LR image and slowly adding details, PULSE traverses the high-resolution natural image manifold, searching for images narrowed down to the original LR image [21]. This is formalized by a "scaled-down loss" that guides exploration of the latent space of the generative model. By exploiting the properties of high-dimensional Gaussians, we restrict the search space to ensure that our output is realistic. As a result, PULSE generates super-resolution images that are both realistic and correctly scaled down. We show extensive experimental results that demonstrate the effectiveness of our approach in the field of facial superresolution, also known as facial hallucinations. This paper also introduces the limitations and biases of the current implementation methods using adjoint model cards with relevant metrics. The proposed method outperforms the latest methods in perceptual quality, and has a higher resolution and scaling factor than before. Advantages and disadvantages: Firstly, a generative network is pre-trained, and then given an LR image with fixed parameters, the

model tries to explore a latent code z, which should be down-sampled from the HR image generated by the generative network to LR image. Then, by adjusting the z, the LR obtained from the generated HR downsampling is the closest to the real LR, and it can be approximated that the network generates the target SR image. Compared with the first model, the trained generative network can generate richer and more realistic texture information. This method sounds nice, but at high multiples, it is difficult to retain the structure information of the image just by code z, so the generated image will suffer Identity distortion. At the same time, in the process of image generation, for each SR image generated, it needs to be iteratively estimated several times, which is very time-consuming, so this method can not be applied to real-time tasks obviously.

III. SUMMARIZATION AND PROSPECT

Future improvement directions in the field of supersegmentation may include proposing more complex loss functions; Implement arbitrary super-resolution construction; While improving the performance, the pursuit of lightweight; Effective combination of various network modules; How to reduce the image quality of data sets, such as blind over segmentation technology to solve the problem of unknown degradation model.

In addition, the training data are difficult to obtain, at present most of the model using the simulation data, the process is difficult to imitate the real image is reduced, the real image degradation is not only a lower resolution, but also in the process, will introduce all kinds of image noise, so based on the sampling of the trained model easy to fitting, generalization ability is bad. It is difficult to generalize the model. For specific types of images, such as facial super segmentation, it is necessary to train the face-related model specially, and the general super segmentation model is often difficult to obtain good results.

Although the performance of the existing deep learning image super-resolution reconstruction algorithms has been greatly improved compared with the previous ones, far surpassing the traditional algorithms, there is still a lot of room for improvement. Looking into the future, the research on super-resolution can be carried out from the following aspects:

(1) Improve network performance. Improving the image effect after reconstruction has always been a hot issue for researchers, but for different use needs, the performance requirements of the network are also different. For example, in video surveillance images, it is necessary to reconstruct the image with good visual perception effect and high reconstruction efficiency. In medical image reconstruction, it is necessary to reconstruct the image with better texture details and ensure authenticity. Therefore, the reconstruction efficiency is improved and better visual perception is obtained

Fruit, better texture details, higher magnif-ication and other aspects are the focus of future research to continue to improve the performance of super resolution network.

(2) Application of image super-resolution in various fields. For example, in the aspect of video, we will continue to optimize video enhancement algorithms including super-score algorithm, create industry-leading image restoration and image enhancement technology, help customers improve video quality, reduce video playback cost, provide lower consumption, lower power consumption, better subjective quality, With more models and algorithms that save more bit rates, users can enjoy UHD video experience on different mobile phones and networks.

REFERENCES

[1] Murshed H,Wei Z,Ahmed M . Perfect Single Image (SR)Super-Resolutionwith,Deep Super Resolution Convolutional Neural NetworkandOpenCV Method [J]. IOSR journal of computer engineering, 2020(3):22.

[2] Ward C M, Harguess J, Crabb. Image quality assessment for determining efficacy and limitations of Super-Resolution Convolutional Neural Network (SRCNN)[C]//Applications of Digital Image Processing XL. SPIE, 2017, 10396: 19-30.

[3] Wang Lie, Yin Jin-wei. Small Object Detection Method Based on SRCNN and SSD network [J].Computer simulation, 2020, 37(3):5.

[4] Shen H F, Li P X, Zhang L P. Overview of image super-resolution reconstruction techniques and methods [J]. Optical Technology, 2009(2):7.

[5] Pu Jian, Zhang Junping, Huang Hua. Review of super-resolution algorithms [J]. Journal of Shandong University: Engineering Science Edition, 2009(1):6.

[6] ANWARS, KHANS, BARNESN. A deep journey into super-resolution: a survey[J]. ACM Computing Surveys (CSUR), 2020, 53(3):1-34.

[7] Kim J, Lee J K, Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2016

[8] Wang Jiaming, LU Tao. Satellite image super-resolution algorithm based on multi-scale residual deep neural network [J]. Journal of Wuhan Institute of Technology, 2018, 40(04):440-445.

[9] Wan Xuefen, Cui Jian, WANG Guanjun. Research on Image Super-resolution Reconstruction Processing Algorithm[C]// National Conference on Optoelectronics and Quantum Electronics Technology. Chinese Institute of Electronics, 2011.

[10] TIAN Yan, TIAN Jinwen, LIU Jian. Implementation for Super Resolution--An Improved Image Interpolation Based on Wavelet Implementation of Super-Resolution Technology -- An Improved Wavelet Interpolation Method [J]. Journal of Image and Graphics, 2003, 45(12):1422-1426.

[11] Jiang Hao, Wang Bofu, Zhuang Qiliang. Reconstruction of turbulent flow field based on super-resolution reconstruction method [J]. Experimental Fluid Mechanics, 2022(036-003).

[12] Wang Rong, Zhang Yonghui, Zhang Jian. Image super-resolution Reconstruction Method based on CNN [J]. Computer Engineering and Design, 2019, 40(6):6.

[13] Zhong Z, Chen Y, Hou S. Super-resolution reconstruction method of infrared images of composite insulators with abnormal heating based on improved SRGAN [J]. IET generation, transmission & distribution, 2022(10):16.

[14] Zou Penghui, Zeng Yijie, Duan Zhenghong. Research on image super-resolution Reconstruction based on SRGAN technology [J]. Science and Technology Trends, 2019(18):1.

[15] Liu Yiwen. Research on Low resolution face Detection Algorithm Based on Deep Learning [D]. University of Electronic Science and Technology of China, 2019.

[16] Hu Lei, Wang Zugen, Chen Tian, et al. An Improved super-resolution Reconstruction algorithm for SRGAN infrared Image [J]. Journal of System Simulation, 2021(033-009).

[17] Nagano Y, Kikuta Y. SRGAN for super-resolving low-resolution food images[C]//Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management. 2018: 33-37.

[18] Li J, Wu L, Wang S. Super resolution image reconstruction of textile based on SRGAN[C]//2019 IEEE International Conference on Smart Internet of Things (SmartIoT). IEEE, 2019: 436-439.

[19] Wang X, Yu K, Wu S. Esrgan: Enhanced super-resolution generative adversarial networks[C]//Proceedings of the European conference on computer vision (ECCV) workshops. 2018: 0-0.

[20] Wang X, Xie L, Dong C. Real-esrgan: Training real-world blind super-resolution with pure synthetic data[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 1905-1914.

[21] Menon S, Damian A, Hu S. Pulse: Self-supervised photo upsampling via latent space exploration of generative models[C]//Proceedings of the ieee/cvf conference on computer vision and pattern recognition. 2020: 2437-2445.