ISSN: 2472-3703 DOI: 10.6911/WSRJ.202311_9(11).0014

Image Super-resolution Reconstruction Algorithm Based on Image Gradient and Generative Adversarial Network

Dongmei Ma[1, a], Jiahao Zhu[1, b, *] and Yu Li[2, c]

[1]School of Physics and Electronic Engineering Northwest Normal University, Gansu Lanzhou, China

[2]Engineering Research Center of Gansu Province for Intelligent Information Technology and Application, Gansu Lanzhou, China

[a]madongmei@nwnu.edu.cn, [b]2543231561@qq.com

Abstract

In order to solve the problem that existing image super-resolution reconstruction models are prone to poor visualization and structural distortion when generating images, we propose a Deep Gradient Guidance Network (DGGN). This model is optimized by means of adversarial training to improve the perceptual effect of the generated images.DGGN introduces a gradient branch to convey the features of the gradient image and fuses the gradient information with the image branch to prevent the distortion of the image edges. We also refer to networks such as MSRB and ResNext, and propose improved multi-scale residual units that are applied to the base module of the image branch and the gradient branch in order to enable the model to better access multi-scale information. The discriminative network uses Wasserstein distance with gradient penalty to improve the stability of network training. Experimental results show that our algorithm performs well in perceptual evaluation, and it can effectively prevent image structure distortion and improve the quality of the generated images compared to perceptually-driven algorithms such as SRGAN, ESRGAN, and NatSR. In addition, our model has a low computational complexity of 22.6 GFLPOs, which is about 1/4 and 1/10 of the computational complexity compared to ESRGAN and SPSR, respectively. This innovative DGGN model is expected to make a significant breakthrough in the field of image super-resolution by improving the perceptual effect of generated images and reducing the computational complexity.

Keywords

Image super-resolution reconstruction; Image gradient; Wasserstein distance; Gradient penalty.

1. INTRODUCTION

Single Image Super-Resolution Reconstruction (SISR) is one of the research areas that have received much attention in the field of computer vision. Its main goal is to generate corresponding high resolution images (High Resolution, HR) from single or multiple low resolution images (Low Resolution, LR) in order to improve the visual perception quality of the images and provide richer information about the image details. This problem is considered as one of the classical representatives of the underlying vision problem [1].

Image super-resolution reconstruction has a wide range of applications in many fields, including but not limited to video surveillance, remote sensing imaging, medical image analysis, etc. [2]. With the development of deep learning, significant progress has been made in image

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202311_9(11).0014

super-resolution reconstruction methods. Initially, Dong et al. proposed an image super-resolution reconstruction algorithm based on convolutional neural network, i.e., SRCNN (Image Super Resolution Convolution Neural Network) [3].SRCNN utilizes a convolutional neural network to learn the mapping relationship from low-resolution image to high-resolution image, but it only contains three convolutional layers which is not good at extracting deep feature information, and it needs to interpolate and enlarge the low-resolution image by a double-cubic interpolation algorithm before inputting it into the network for reconstruction, which increases the computational complexity[4]. Subsequently, researchers proposed more efficient, faster and better image super-resolution reconstruction algorithms, one of which is FSRCNN (Fast Image Super-Resolution Convolution Neural Network) [5], which was improved from SRCNN by Dong et al. Shi et al. [6] proposed a new up-sampling method, Effecent Subpixel Convolutional Layer (ESPCN), which utilizes pixel shifting to expand the image size without the zero complement operation of the inverse convolution, effectively avoiding the problems of image edge distortion and the image "checkerboard effect". Ledig et al [9] proposed Image Super-Resolution Generative Adversarial Network (SRGAN) based on Generative Adversarial Network, which optimizes the model by perceptual loss and still obtains perceptually better images at 4× high magnification factor[7] [8]. Lim et al [10] constructed an Enhanced Deep Residual Network for Image Super-Resolution (EDSR) to remove the BN layer from SRResNet to reduce the computational complexity of the model and at the same time to improve the model performance and to prevent artifacts in the image.

Although existing image super-resolution reconstruction algorithms based on deep learning and generative adversarial networks have improved in terms of image perceptual quality, and objective metrics, some of the algorithms, e.g., ESRGAN [11], SRGAN [9], NatSR [12], etc., do not recover the high-frequency details well, and produce geometrical distortions in the reconstruction process. In this regard, Wang et al [13] proposed SFTGAN, which uses semantic segmentation probability maps as a priori information to guide the reconstruction, making the model easier to recover the real texture.Sun et al [14] used gradient distributions and gradient fields to guide the super-resolution reconstruction prior to representing the image gradients, whose statistical correlations are modeled based on the parameters observed in the LR image to estimate the HR edge-related parameters, and the modeling process is done point by point, which resulting in high algorithmic complexity and poor flexibility.Zhu et al [15] complete the modeling by collecting a dictionary of gradient patterns and combinations of deformable gradients.

The discriminative network uses WGAN-GP [17] containing gradient penalties to improve the stability of generative adversarial network training and to prevent the degradation of generative network performance caused by the disappearance of the gradient of the discriminative network. Comparing with the existing image super-resolution reconstruction models, the main contributions of this paper are:

a) Propose a gradient-guided image super-resolution reconstruction model, where the generative network adds gradient branches on top of the baseline image branches to prevent structural distortion of the generated image, and at the same time, utilizing the advantages of the network structure of ResNext [18] and Inception [19], an improved multi-scale residual unit is proposed as the base module to increase the network sensing field while reducing the The computational complexity of the model is reduced while increasing the network receptive field.

b) The discriminative network adopts Wassertein distance with gradient penalty to prevent the gradient of the discriminator from disappearing and improve the stability of the whole generative adversarial network training.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202311_9(11).0014

2. PROPERTIES

In order for the model to easily detect image features at different scales, Li et al [20] proposed the Multi-Scale Residual Block (MSRB), which consists of multi-scale feature fusion and local residual learning.The MSRB is a dual bypass structure, where the two bypasses use 3×3 and 5×5 convolution kernels, respectively, and the two bypasses share information between them .

Let and be the inputs and outputs of the MSRB, respectively, then the MSRB is defined as:

$$\begin{aligned} S_1 &= R(w_{3\times3}^1 * M_{n-1} + b^1) \\ P_1 &= R(w_{5\times5}^1 * M_{n-1} + b^1) \\ S_2 &= R(w_{3\times3}^2 * concat(S_1, P_1) + b^2) \\ P_2 &= R(w_{5\times5}^2 * concat(P_1, S_1) + b^2) \\ S' &= w_{1\times1}^3 * concat(S_2, P_2) + b^3 \end{aligned} \tag{1}$$

where w is the convolution weights, b is the bias, and R denotes the ReLU activation function. Finally the residual connection is used to sum the output and input:

$$M_n = S' + M_{n-1} \tag{2}$$

2.1. WGAN-GP

WGAN uses Wassertein distance instead of the traditional JS scatter [21], and all parameters of the WGAN discriminator training should satisfy the Lipschitz condition, i.e., for two inputs x1 and x2:

$$| D(x_1) - D(x_2) | \leq | x_1 - x_2 | \tag{3}$$

WGAN directly adopts the method of weight clipping to ensure that all training parameters of the discriminative network are bounded.WGAN effectively alleviates the problem of instability during the training of generative adversarial networks, but WGAN adopts the method of weight clipping to ensure that the parameters of the discriminative network satisfy the Lipschitz condition, which is susceptible to the problem of vanishing or exploding gradient.

WGAN-GP does not need to use weight clipping and is only effective for regions in the real and generated sample sets as well as regions between true and false samples, there is no problem of disappearing or exploding gradient, the gradient is controllable and easy to be adjusted to the appropriate size.

2.2. Gradient image computation

The gradient of an image refers to the rate of change of a pixel point in an image along both x and y directions. The gradient of an image describes the rate of change of the image and the edge information of the image, the part with larger gradient value and large change of gray value is the edge part of the image; the part with smaller gradient value and small change of gray value is the smooth part of the image.

$$\nabla f(x,y) = \begin{pmatrix} g_x \\ g_y \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x} f(x,y) \\ \frac{\partial}{\partial y} f(x,y) \end{pmatrix} = \begin{pmatrix} f(x+1,y) - f(x-1,y) \\ f(x,y+1) - f(x,y-1) \end{pmatrix} \tag{4}$$

$$\theta = \arctan \frac{f(x+1,y) - f(x-1,y)}{f(x,y+1) - f(x,y-1)}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202311_9(11).0014

Equation (4) shows that the variation of any pixel (x,y) in the image along the x-axis is the difference between the pixel value on the right side of the pixel and the pixel value on the left side of the pixel; and the variation along the y-axis is the difference between the pixel value on the lower side of the pixel and the pixel value on the upper side of the pixel, and the two components form a two-dimensional vector, which is taken to be the tangent of the angle θ of the gradient of the image.

3. IMAGE GRADIENT AND GENERATIVE ADVERSARIAL NETWORK IMAGE SUPER RESOLUTION RECONSTRUCTION ALGORITHMS

3.1. Generation Network

Aiming at the current image super-resolution reconstruction model is prone to structural distortion, excessive image smoothing, and poor perception, this paper proposes a depth gradient guidance network DGGN, whose generative network structure is shown in Figure 1:

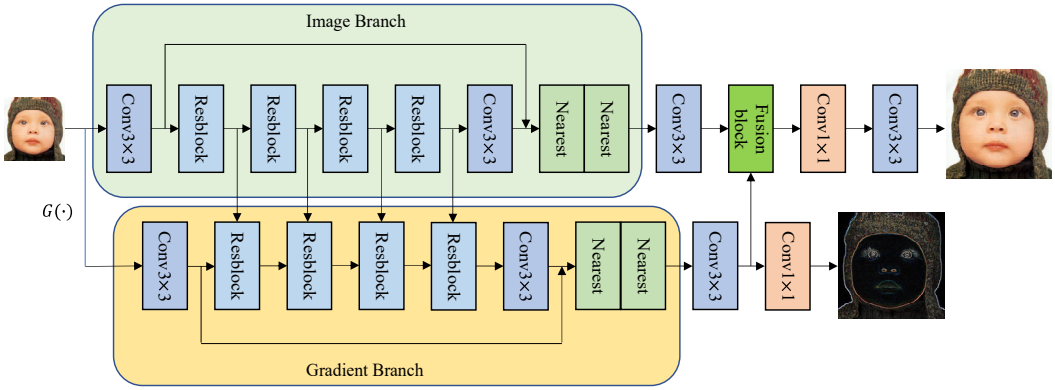

Figure 1. Deep Gradient Guidance Network

DGGN consists of an image branch and a gradient branch. The image branch has basically the same structure as the existing deep learning-based image super-resolution reconstruction model, which consists of one or more convolutions to complete the shallow feature extraction, and then uses multiple base modules to complete the deep feature extraction, sends the extracted deep features to the upsampling module to complete the amplification, and finally undergoes a layer of convolution to obtain the final output.

For the computation of gradient in Fig. 1, let A=[1,0,-1] be a 1×3 convolution kernel, AT be the transpose of A, ∇S_x and ∇S_y be the gradient of the image S in the x and y directions, respectively, and "*" denotes the convolution operation, the magnitude of the gradient of the image S can be obtained in the following way:

$$\nabla S_x = A * S$$
$$\nabla S_y = A^T * S \tag{5}$$

In Eq. (5), A is used to calculate the gradient in the x-direction and AT is used to calculate the gradient in the y-direction. Since the magnitude of the gradient is sufficient to represent the degree of sharpening of the image edges, the direction of the computed gradient is not considered in this paper.

Inspired by MSRN [20] and combining the advantages of ResNext [18] and Inception [19] networks, this paper proposes an improved multiscale residual unit, the structure of which is shown in Fig. 2. The DGGN multiscale residual unit extends the double-bypass structure of MSRB into four branches, uses a 1 × 1 convolution before and after each of the 3 × 3 and 5 × 5 convolutions to improve its nonlinear mapping capability, and finally the extracted multiscale

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202311_9(11).0014

features are fused and the inter-channel dependency is enhanced by introducing the SEblock [22] channel attention module.

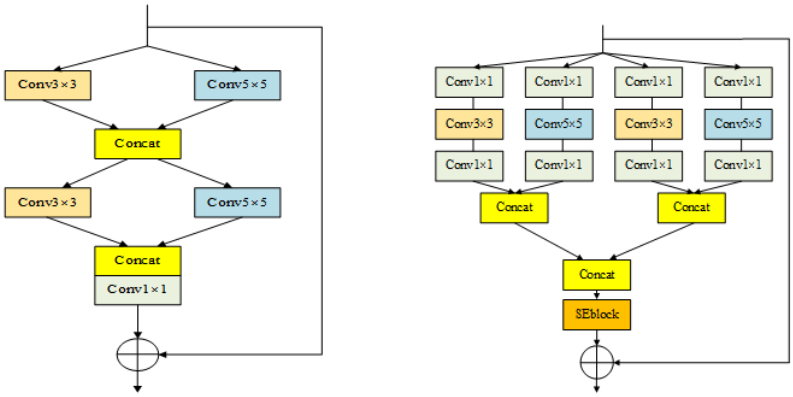

Figure 2. Comparsion of multi scale residual unit

The multi-scale residual cell of DGGN is similar to Resnext, but differs from Resnext in the following ways:

(1) DGGN is used to accomplish the task of image super-resolution reconstruction, so the residual unit of DGGN does not contain the BN layer, meanwhile, DGGN uses the LeakyReLU activation function, which ensures that the activated features can still contain rich detail information.

(2) The 1×1 convolution in the residual unit of DGGN has the same number of input and output feature channels, and the main role of the 1×1 convolution in Resnext is to map high-dimensional features to low dimensions.

(3) DGGN can capture multi-scale information by 3×3 convolution and 5×5 convolution, and Resnext mainly consists of 3×3 convolution

3.2. discriminant network

The structure of the discriminant network is shown in Fig. 3. The discriminant network is roughly similar to VGG, but uses a convolutional kernel of 4 × 4 with a step size of 2 for downsampling. Same as WGAN, the last layer of the discriminative network in this paper does not use Sigmoid, but the discriminative network in this paper uses gradient penalty to penalize the gradient paradigm of the discriminative network for each input independently, and the use of the BN layer modifies the gradient along with the batch information, so the discriminative network in this paper does not use the BN layer.

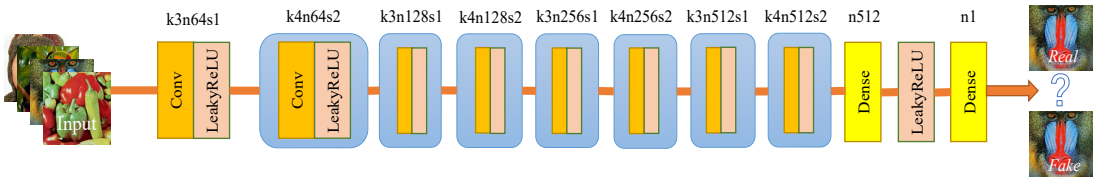

Figure 3. Architecture of discriminator

3.3. loss function

The objective function of this paper can be expressed as:

$$L = \theta L_{percept} + \alpha L_{IB}^{pixel} + \mu L_{IB}^{grad} + \beta L_{IB}^{adv} + \varepsilon L_{IB}^{adv_{grad}} + \eta L_{GB}^{grad} \quad (6)$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202311_9(11).0014

where L_percept is the perceptual loss [23], using VGG as the feature extraction network. Noting that φ_i (·) is the feature map output from the ith convolutional layer of the VGG network, the perceptual loss is denoted as

$$L_{percept} = \frac{1}{n}(\sum_{i=1}^{n} | \varphi(\hat{x}) - \varphi(x) |_1^1) \quad (7)$$

That is, the L1 distance is calculated between the generated image and the original image after VGG extraction of features.

4. EXPERIMENTS

4.1. Experimental environment setup

The experimental environment of this paper is as follows: the CPU uses i5 13400F; the GPU is NVIDIA RTX3060 with 12GB of video memory and 32GB of RAM. all the experiments are conducted in Ubuntu22.04 environment, and the deep learning frameworks are PyTorch2.0 and CUDA12.0.

4.2. Data set

All experiments in this paper use the DIV2K dataset, a total of 1000 images with 2K resolution, of which 800 are in the training set, and 100 are in the validation set and test set. 800 images in the training set are cropped into sub-images of size 480×480 in step 240, and after cropping, a total of 32,592 images are included in the training set, and random flipping is used to enhance data with a probability of 0.5 to the images. Horizontal flipping is performed for data enhancement, the input low resolution image size is 48×48 and the corresponding high resolution image size is 192×192.

4.3. Analysis of experimental results

In order to prove the effectiveness of gradient branching, this paper verifies for gradient branching, removes the gradient branching and only retains the image branching, meanwhile, in Eq. (12), takes the values of μ, ε and η as 0, and calculates the PI values for the two cases, and the results are shown in Table 1:

Table 1. Gradient branch validity verification

	Set5	Set14	BSD100	Urban100
gradient-free branching	3.5716	3.0795	2.7136	3.9815
branch with gradient	3.3543	2.7565	2.4097	3.5936

The results in Table 1 show that when the generative network does not contain gradient branches, the performance is significantly lower than the model when it contains gradient branches. When the generative network contains gradient branches, the PI values on the four test sets decrease by 0.2173, 0.3230, 0.3039, 0.3879 respectively compared to no gradient branches, indicating that the model in this paper produces images with higher perceptual effects when the generative network contains gradient branches.

In this paper, some of the classical PSNR driving algorithms EDSR [10], RCAN [26], RDN [27], DBPN [28] and some of the perceptual driving algorithms SRGAN [9], ESRGAN [11], NatSR [12], and SPSR [16] are selected to compare with the algorithms in this paper, and the results are shown in Tables 2 and 3, respectively:

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202311_9(11).0014

Table 2. Objective evaluation index of different algorithms on test detasets

	Set5	Set14	BSD100	Urban100
EDSR[10]	32.46/0.8968	28.80/0.7876	27.71/0.7420	26.64/0.8033
RCAN[26]	**32.63/0.9002**	**28.87/0.7889**	**27.77/0.7436**	**26.82/0.8087**
RDN[27]	32.47/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028
DBPN[28]	32.47/0.8980	28.82/0.7860	27.72/0.7400	26.38/0.7946
ESRGAN[11]	30.33/0.8521	26.10/0.6996	24.53/0.6327	22.74/0.6824
SRGAN[9]	29.87/0.8494	26.44/0.7132	24.97/0.6445	23.02/0.6875
NatSR[12]	30.92/0.8626	27.23/0.7365	25.66/0.6715	23.72/0.7132
SPSR[16]	30.30/0.8439	26.44/0.7141	24.80/0.6428	23.10/0.6942
DGGN(Ours)	30.38/0.8452	26.09/0.7029	24.49/0.6273	23.12/0.6864

Table 3. Perceptual index of different algorithms on test datasets

	Set5	Set14	BSD100	Urban100
EDSR[10]	5.9857/0.2088	5.2630/0.2963	5.2610/0.3249	4.9844/0.2727
RCAN[26]	6.3691/0.2164	5.7150/0.3106	5.7581/0.3320	5.4178/0.3015
RDN[27]	6.0092/0.2135	5.4633/0.3040	5.5412/0.3300	5.2502/0.2905
DBPN[28]	6.1324/0.2109	5.4677/0.2986	5.4896/0.3259	5.1363/0.2838
ESRGAN[11]	3.7522/0.0748	2.9261/0.1329	2.4793/0.1614	3.7704/0.1229
SRGAN[9]	3.9820/0.0882	3.0851/0.1663	2.5459/0.1980	3.6980/0.1551
NatSR[12]	4.1648/0.0939	3.1094/0.1758	2.7801/0.2114	3.6523/0.1500
SPSR[16]	**3.2743/0.0644**	2.9036/**0.1318**	**2.3510/0.1611**	**3.5511**/0.1184
DGGN(Ours)	3.3543/0.0721	**2.7565**/0.1454	2.4097/0.1645	3.5936/**0.1107**

Table 2 Objective evaluation metrics of each algorithm, Table 2 analysis shows that the residual channel attention based algorithm RCAN obtained the highest PSNR and SSIM values in the four test sets. Algorithms using perception-driven algorithms, such as SPSR and ESRGAN, have lower PSNR and SSIM values than algorithms driven by PSNR, and since the optimization objective of SRGAN consists of only perceptual and antagonistic losses, and does not include the L1 or L2 losses, SRGAN has the lowest PSNR and SSIM values.The DGGN has the highest PSNR and SSIM values in the four test sets, which are similar to the SPSR are close to each other, the PSNR values in Set5 and Urban100 test sets are improved by 0.08 and 0.02 than SPSR respectively, and the SSIM values are not much different.

Table 3 Perceptual evaluation metrics of each algorithm, synthesizing the results in Table 2, it can be found that algorithms with higher PSNR and SSIM values also have higher values of the perceptual metrics PI and LPIPS, generating smoother images. While algorithms with perception-driven algorithms have lower PSNR and SSIM values, and lower PI and LPIPS values, but generate images with richer detail information, which are more satisfying to human visual perception effects. Table 3 analyzes found that overall, the PI and LPIPS values of the DGGN algorithm are close to the SPSR, with a decrease of 0.1488 in the PI value over the SPSR on the Set14 test set and a decrease of 0.0077 in the LPIPS value over the SPSR on the Urban100 test set.

4.4. Performance Comparison of Classical Perception-Driven Algorithms

In order to further verify the effectiveness of the algorithms in this paper, the number of parameters and the computational amount of five classical perception algorithms, namely SFTGAN, ESRGAN, NatSR, SPSR, and SRGAN, are selected for comparison, and the results are shown in Table 4:

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202311_9(11).0014

Table 4. Params and FLOPs of different perception-driven algorithms

	SFTGAN[13]	NatSR[12]	ESRGAN[11]	SRGAN[9]	SPSR[16]	DGGN(ours)
Params	1.8M	5.0M	16.7M	0.7M	24.8M	14.6M
FLOPs	0.8G	12.5G	89.7G	7.0G	265.1G	23.7G

The analysis of the results in Table 4 shows that the number of parameters of DGGN is 14.6M, which is lower than the 24.8M of SPSR and the 16.7M of ESRGAN, and the floating-point computation of DGGN is 23.7GFLOPs, which is 1/10 and 1/4 of that of SPSR and ESRGAN, respectively, which proves that this paper's algorithm still obtains results comparable to that of SPSR with the number of parameters and computation being smaller than that of SPSR. SPSR comparable results.

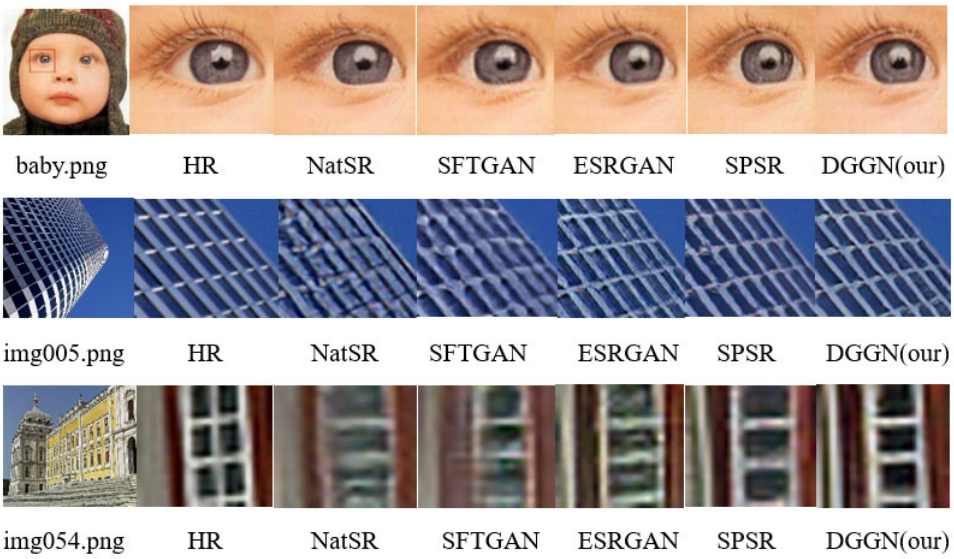

Figure 4. Visualize results of different algorithms

The visualization results of some selected classical perception-driven algorithms are shown in Figure 4. The human eye brow details and building details are selected for local zoom. The algorithms using perception-driven all use generative adversarial network and perceptual loss optimization, which can generate images with more realistic effect and better perceptual quality, for the human eye brow details, the algorithms using perception-driven all can generate images with higher visual effect, SFTGAN, SPSR and DGGN are three methods using a priori information to guide the reconstruction, and DGGN generates more rich details and generates more realistic image effect than SFTGAN and SPSR. are richer and produce more realistic image results. For building details, some of the perceptual algorithms, such as SRGAN and NatSR, have structural distortion, SFTGAN uses semantic segmentation probability map to guide the reconstruction, and its guiding effect is greatly affected by the semantic segmentation model, and it cannot guide the reconstruction well, DGGN and SPSR algorithms both use gradient information to guide the reconstruction, and DGGN adopts multi-scale residual unit to extract features, the DGGN uses multi-scale residual unit to extract features, the feature information obtained is richer, and the visual effect of the final generated image is better than the algorithms such as SPSR and ESRGAN.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202311_9(11).0014

5. CONCULUSION

In this paper, we propose an image super-resolution reconstruction model, DGGN, based on gradient-guided generative adversarial networks, which introduces improved multiscale residual units by borrowing the existing ResNext, Inception, and MSRB structures and applying them to the image branch and gradient branch of generative networks. This is done to make it easier for the model to acquire multi-scale feature information, thus solving the problem that existing image super-resolution reconstruction models are prone to structural distortion when generating visual effects. In order to improve the stability of training, the discriminative network adopts the WGAN-GP method containing gradient penalty. Compared with the perception-driven based method SPSR, DGGN not only generates images with higher visual quality, but also reduces the number of model parameters and computational complexity.

REFERENCES

[1] Tang Yanqiu, Pan Hong, Zhu Yaping, et al. A survey of image super-resolution reconstruction [J]. Joural of electronic, 2020, 48(7):1407-1419.

[2] Nan Fangzhe, Qian Yurong, Xing Yanni, et al.Survey of single image super-resolution based on deep learning[J]. Application Research of Computers, 2020, 37(2):321-326.

[3] Dong Chao, Loy C C, He Kaiming, et al. Learning a deep convolutional network for image super-resolution[C] // Proc of the European Conference on Computer Vision(ECCV). Germany: Springer, 2014:184-199.

[4] Dong Chao, Loy C C, Tang Xiaoou. Accelerating the super-resolution convolutional ceural network[C]//Proc of the European Conference on Computer Vision(ECCV). Germany: Springer, 2016: 391-407.

[5] Li Xu, Jimmy S R, Ce L,et al. Deep convolutional neural network for image deconvolution[C]// Proc of the Conference and Workshop on Neural Information Processing Systems(NIPS). 2014:1790-1798.

[6] Shi Wenzhe , Caballero J , Huszár F , et al. Real-Time single image and video super-resolution using an efficient sub-pixel convolution neural network[C]//Proc of the Conference on Computer Vision and Pattern Recognition(CVPR). USA:IEEE, 2016:1874-1883.

[7] Kim J, Lee J K, Lee K M, et al. Accurate image super-resolution using very deep convolutional networks[C] // Proc of the Conference on Computer Vision and Pattern Recognition(CVPR). USA: IEEE，2016: 1646－1654.

[8] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution[C] // Proc of the Conference on Conputer Vision and Pattern Recognition(CVPR). USA:IEEE, 2016:1637-1645.

[9] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C] // Proc of the Conference on Computer Vision and Pattern Recognition(CVPR). USA:IEEE, 2017:105-114.

[10] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C] // Proc of the Conference on Computer Vision and Pattern Recognition Workshops(CVPRW). USA:IEEE, 2017:1132-1140.

[11] Wang Xintao, Yu Ke, Wu Shixiang, et al. ESRGAN:Enhanced super-resolution generative adversarial network[C] // Proc of the European Conference on Computer Vision(ECCV). Germany: Springer, 2018:63-79.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202311_9(11).0014

[12] Soh J W, Park G Y, Jo J, et al. Natural and realistic single image super-resolution with explicit natural maniford discrimination[C] // Proc of the Conference on Computer Vision and Pattern Recognition(CVPR).USA:IEEE, 2019:8114-8123.

[13] Wang Xintao, Yu Ke, Dong Chao, et al. Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform[C] // Proc of the Conference on Computer Vision and Pattern Recognition (CVPR).USA:IEEE, 2018:606-615.

[14] Sun Jian, Xu Zongben, Shum Heung-Yeung. Gradient profile prior and its applications in image super-resolution and enhancement[J]. IEEE Transactions on Image Processing. 2011,20(6):1529-1542.

[15] Zhu Yu, Zhang Yanning, Bonev B, et al. Modeling Deformable Gradient Compositions for Single-Image Super-resolution[C] // Proc of the Conference on Computer Vision and Pattern Recognition (CVPR).USA:IEEE, 2015:5417-5425.

[16] Ma Cheng, Rao Yongming, Cheng Yean, et al. Structure-preserving super-resolution with gradient guidance[C] // Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR).USA:IEEE, 2020:7769-7778.

[17] Ishaan G, Faruk A, Martin A, et al. Improved training of Wasserstein GANs[EB/OL].(2017-12-25)..

[18] Xie Saining, Girshick R, Dollár P, et al. Aggregated Residual Transformations for Deep Neural Networks[EB/OL].(2017-4-11).[2022-07-23].

[19] Szegedy C, Ioffe S, Vanhoucke V. Inception-v4,inception-resnet,and the impactof residual connections on learning[EB/OL].(2016-08-23)..

[20] Li Juncheng, Fang Faming, Mei Kangfu, et al. Multi-scale residual network for image super-resolution[C]//Proc of the European Conference on Computer Vision(ECCV).Germany: Springer, 2018:527-542

[21] Martin A, Soumith C, Léon B. Wasserstein GAN[EB/OL].(2017-12-06).

[22] Hu Jie, Shen Li, Albanie S, et al. Squeeze-and-excitation networks[EB/OL].(2019-05-16).

[23] Johnson J , Alahi A ,Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution[C] // Proc of the European Conference on Computer Vision(ECCV).Germany: Springer, 2016:694-711.

[24] Blau Y, Mechrez R, Timofte R, et al. The 2018 pirm challenge on perceptual image super-resolution[C] // proc of the European Conference on Computer Vision(ECCV). Germany: Springer, 2018:334-355.

[25] Zhang Richard, Isola P, Alexei A. Efros., et al. The unreasonable effectiveness of deep features as a perceptual metric[C] // Conference on Computer Vision and Pattern Recognition (CVPR). USA:IEEE,2018:586-595.

[26] Zhang Yulun, Li Kunpeng, Li Kai, et al. Image super-resolution using very deep residual channel attention networks[C] // Proc of the European Conference on Computer Vision(ECCV). Germany: Springer, 2018:294-310.

[27] Zhang Yunlun, Tian Yapeng, Kong Yu, et al. Residual dense network for image super-resolution[EB/OL]. (2018-05-27).

[28] Haris M, Shakhnarovich G, Ukita N. Deep back-projecti networks for single image super-resolution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(12): 4323-4337.