Image Inpainting Research Based on Deep Learning

Zhao Ruixia
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 1364343954@qq.com

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

***Abstract*—With the rapid development of computer technology, image inpainting has become a research hotspot in the field of deep learning. Image inpainting belongs to the intersection of computer vision and computer graphics, and is an image processing technology between image editing and image generation. The proposal of generative adversarial network effectively improves the problems of poor image inpainting effect and large difference between the inpainting image and the target image, and promotes the development of image inpainting technology. In this paper, the image inpainting is based on the generation of confrontation networks. Its network structure establishes two repair paths, namely the reconstruction path and the generation path, and the two paths correspond to two groups of networks. The encoder and generator in the network respectively complete the encoding and decoding tasks based on the residual network. The discriminator also uses the patch block discriminator on the basis of the residual network to discriminate the authenticity of the image. This paper uses Places2 data set to verify the algorithm, and uses PSNR and SSIM two objective evaluation methods to evaluate the quality of the repaired image. Experiments show that the algorithm inpainting effect is better.**

Keywords-Image Inpainting; Generation Adversarial Networks; Residual Network; Patch

I. INTRODUCTION

With the development and popularization of computer technology, Internet technology and multimedia technology, digital image processing technology has also developed rapidly. In the process of storage, transmission and use of digital image information, the phenomenon of image information damaged and loss will occur. These damaged areas affect the visual effect of the picture and the integrity of the information, and have a certain impact on the application of the picture. People urgently need a technology and method that can automatically inpainting damaged digital images, so digital image inpainting technology is born.

Image inpainting is one of the most popular areas of deep learning. Its basic principle is to give an image of a damaged or corroded area, and try to use the intact information of the known area of the damaged image to inpainting the damaged area of the image[1-2]. Digital image inpainting methods can be divided into two major categories: traditional image inpainting methods and deep learning-based image repair methods. Traditional image repair methods can be divided into: structure-based image repair technology and texture synthesis-based image inpainting technology. Both image inpainting algorithms based on structure and texture can inpainting the loss of small areas such as folds. With the expansion of the missing areas, the inpainting effect gradually deteriorates. There are problems such as incomplete semantic information and blurred images in the inpainting results, which makes the image inpainting effect ineffective, ideal. The emergence of deep neural networks allows the model to obtain the understanding of image semantic information through multi-level feature extraction, and to a certain extent improves the repair effect of large-area damaged images.

As deep learning shows exciting prospects in the fields of image semantic inpainting and situational awareness, and image inpainting algorithms based on deep learning can capture more advanced features of images than traditional inpainting algorithms based on structure and texture, so often used for image inpainting. At present, image inpainting based on generative adversarial networks is a major research hotspot in the field of deep learning image inpainting, which lays a solid foundation for the development of image inpainting technology.

DOI: 10.21307/ijanmc-2020-013

A. The basic idea of generating adversarial networks

Generative adversarial network is undoubtedly one of the popular artificial intelligence technologies, and was rated as the "Top Ten Global Breakthrough Technologies" in 2018 by the MIT Technology Review. The generative adversarial network is composed of a generative network and a discriminant network. The purpose of the generative network is to estimate the distribution of data samples from a given noise and generate synthetic data. The purpose of the discriminant network is to distinguish the input data from the generated data or the real data. The generative network and the discriminant network are a set of confrontational relationships. The source of the confrontational ideas comes from the zero-sum game in game theory. The two sides of the game use each other's strategy to change their confrontation strategy in an equal game, so as to achieve the goal of winning[3]. It is extended to the generative antagonistic network, that is, the generative network and the discriminant network are the two sides of the game. The optimization goal is to achieve Nash equilibrium[4], the generative network tries to produce closer to real data. Accordingly, the discriminant network tries to distinguish more perfectly between real data and data generated by generators. As a result, the two networks progressed in confrontation, and continued to confront each other after the progress, the data obtained by the generating network became more and more perfect, approaching the real data.

B. Development of deep learning models

GeneratingSince the input of the GAN generation model is random noise data, in actual applications, there are generally clear variables to control the category or other information for the data to be generated, such as generating specific numbers from 1 to 9 numbers. In order to solve the problem of generating labeled data, Conditional Generative Adversarial Networks are proposed, and information such as category labels and pictures are added to the input to make the image more in line with the target[5]. The foundation of image inpainting technology based on deep learning is the convolutional neural network, which uses the convolutional neural network to extract high-dimensional features and information prediction, which makes the image inpainting technology develop rapidly[6-7]. Because the network of generating model and discriminating model in GAN is too simple, there will be image blur when generating large-size images. In order to generate clear images, Radford A et al.[8] proposed deep convolutional generation adversarial networks. With the emergence of several unsupervised image conversion models, such as CycleGAN[9], DualGAN[10], DiscoGAN[11], it provides better ideas for image inpainting technology.

II. NETWORK STRUCTURE

Image inpainting not only requires that the results conform to human visual habits, making it difficult for the human eye to detect the traces of inpainting (undetected)[12], meanwhile inpainting the information contained in the missing pictures as much as possible, so that the restored image can be as much as possible Same as the image before the damage. Based on this restoration goal, this paper builds an image inpainting network framework suitable for this article by studying and analyzing the structure principles of GAN.

Using the neural network's ability to extract high-dimensional features of images, the structural framework of this paper is built. In this paper, a parallel dual-path framework based on GAN is used: one is to reconstruct the path, and use the given real image and masked image to obtain its complementary image to reconstruct the original image; the other is to generate the path and use the given masked image to inpainting. The input image of the generated path and the input image of the reconstructed path are complementary images of each other. The network structure is built on the basis of the residual network. Its structure includes three parts: encoder, generating network and discriminating network. The image inpainting process in this paper is: (1)Input the masked image and the complement image (the masked image and the supplementary image are the real image) into the encoders E1 and E2 of the reconstruction path and the generation path to encode; (2)The extracted two image features were fused and input into generator G1 and G2; (3)The generator reconstructed image and the real image are input into the discriminator D1 for discrimination; (4)The generated image, the fused image and the real image are input into the discriminator D2 for discrimination; (5)The discriminators D1 and D2 output the discriminant results and feed them back to the encoder, generator and discriminator through the back propagation algorithm to update the network parameters and train the network. The overall structure of the network is shown in Figure 1.

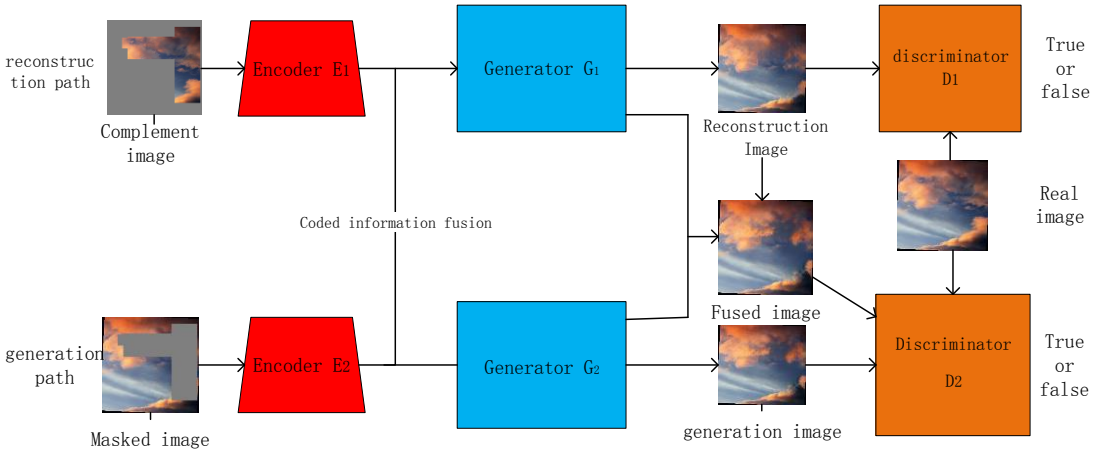


Figure 1. Data flow diagram of GAN

A. *Encoder*

The encoder extracts the features of the image based on the residual network. The inputs of encoders E1 and E2 are three-channel images of 256×256 pixels. The residual block is composed of two layers of convolution and one layer of skip link. The first layer uses a convolution kernel of size 3×3. The length is 1 and the padding size is 1. The second layer uses a 3×3 convolution kernel with a sliding step size of 1 and no padding. The residual structure of the encoder is shown in Figure 2.

In this paper, there are two parallel paths for image inpainting: reconstruction path and generation path. The network structure of the encoder is the same, and the combination of residual modules is used. The network structure contains 7 residual modules. The network structure of the encoder is shown in Figure 3.

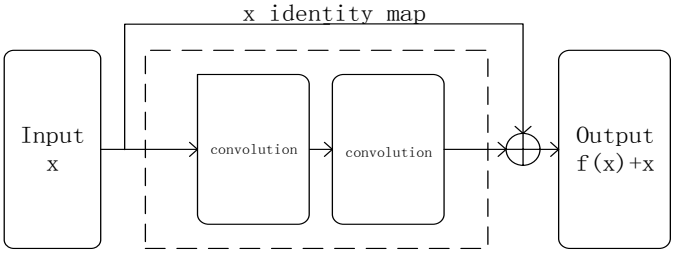


Figure 2. Residual structure of the encoder

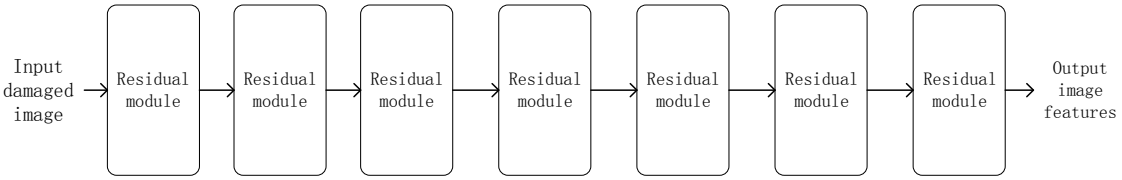


Figure 3. Encoder network structure

B. *Generate network*

The generating network adopts Res-Net network structure, and uses the residual decoding block to decode the features extracted in the encoding stage. In the generation network, the residual block is used in the decoding stage. The residual block in the decoding stage is composed of three parts: a convolution layer, a deconvolution layer, and a skip link layer. The convolutional layer uses a convolution kernel with a

size of 3×3, a sliding step size of 1, and a padding of 1. The deconvolution layer uses a 3×3 convolution kernel with a sliding step size of 2 and a padding of 1. After the deconvolution operation, the padding of the output image is 1. The skip link layer performs a deconvolution operation, using a convolution kernel with a size of 3×3, a sliding step size of 2, and a fill of 1. After the deconvolution operation, the output image has a fill of 1. The generated network uses the Spectral Normalization method to normalize the output data. The network structure of the residual block in the decoding stage is shown in Figure 4.

A self-attention mechanism has been added to the network. The self-attention mechanism uses residual blocks and uses Short+Long Term to ensure the consistency of the appearance of the generated image. The network structure of the generated network is shown in Figure 5.

C. The training principleDiscrimination Network

The discrimination network adopts the structure of PatchGAN. The difference between PatchGAN and ordinary GAN is that the output of ordinary GAN is the evaluation of the entire image, and the output of PatchGAN is an N×N matrix. Each element of the N×N matrix represents the original image. The larger receptive field in the map corresponds to a patch in the original picture. This paper runs a patch discriminator on the image in a convolution mode. The discriminator outputs a patch block of 70×70 size, and each element represents the probability value of the real image. This paper judges that the input of the network is a picture, the target picture is used as a positive example, and the inpainting picture is used as a negative example, so as to judge whether the inpainting picture is true. The discriminators D1 and D2 in this paper have the same network structure and use five-layer convolution. The first three layers use a 4×4 convolution kernel with a sliding step size of 1 and a padding of 1; the last two layers use a 4×4 convolution kernel with a sliding step size of 2 and a padding of 1. The discriminant network first extracts the features of the input image, and then analyzes and compares the extracted features. The network structure of the discrimination network is shown in Figure 6.

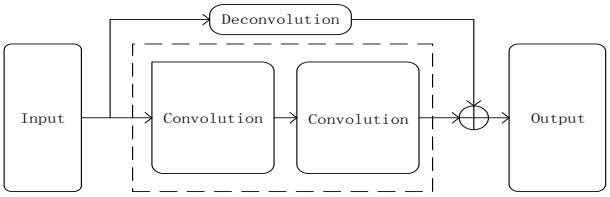


Figure 4. Decoding residual block network structure

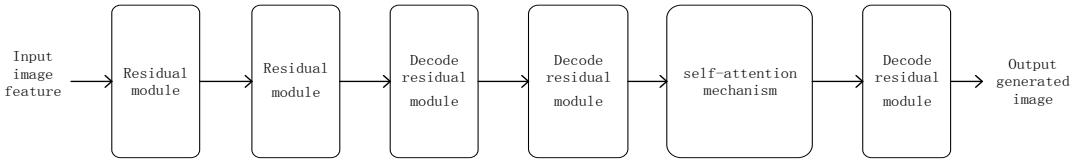


Figure 5. Generate network structure diagram

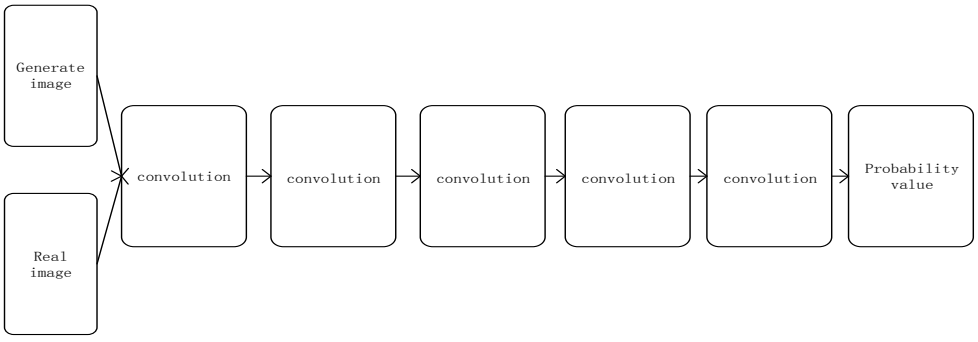


Figure 6. Discriminant network structure diagram

III. NETWORK TRAINING

In this paper, WGAN-GP loss is used to optimize the network structure. WGAN-GP is an improvement of WGAN. A gradient penalty method is proposed to improve the continuity constraint conditions, making GAN convergence more stable. The loss function of WGAN-GP is composed of the loss LG of the generator and the loss LD of the discriminator. The calculation formula of generator loss can be written as

$$\begin{aligned} L_D^{WGAN} &= E\left[D(x)\right] - E\left[D\left(G(z)\right)\right] + L_{gp} \\ L_{gp} &= \lambda E\left[\left(\left|\nabla D\left(\alpha x - \left(1-\alpha G(z)\right)\right)\right| - 1\right)^2\right] \\ L_D &= L_D^{WGAN} + L_{gp} \end{aligned} \tag{1}$$

Where x represents a randomly selected sample in the data set and $D(x)$ represents the result output when the input of the discriminant model is a real sample. L_D^{WGAN} Represents the loss function corresponding to the WGAN discriminator, Lgp represents the gradient penalty loss function newly added in WGAN-GP, and λ represents the penalty coefficient.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental environment

In order to verify the effectiveness of the algorithm proposed in this article, on the Ubuntu platform, the Python language and the PyTorch deep learning framework are used. Experiment with 5000 images of Place2, a public data set. The image size is 256×256 pixels, and the ratio of 8: 2 is used for training and testing.

B. Experimental results

Since the image inpainting task is to repair the incomplete part of the image, the data set should be mask processed before the inpainting task. In this paper, the image preprocessing is divided into two methods: random masked and intermediate masked. After the data processing is completed, the image inpainting task is performed.

The inpainting result of occlusion in the image is shown in Figure 7. Where (a) represents the damaged image, (b) represents the inpainting image, and (c) represents the real image.

The inpainting result of random masked in the image is shown in Figure 8. Where (a) represents the damaged image, (b) represents the inpainting image, and (c) represents the real image.

C. Experimental analysis

At this stage, there are mainly two kinds of image evaluation methods: subjective evaluation method and objective evaluation method. This article combines the subjective evaluation method and the objective evaluation method to evaluate the repaired image.

1) Subjective evaluation

From the experimental results of 4.2, it can be seen that the content of the image inpainting by this method is basically the same as the target image, the color is very similar to the target image, and direct visual observation of the image is real and natural. The inpainting of texture is natural and continuous.

2) Objective evaluation

The objective evaluation method uses peak signal-to-noise ratio measurement (PSNR) and structural similarity (SSIM) to evaluate the repaired image. The higher the PSNR, the less distortion in the picture inpainting process, and the better the inpainting picture. SSIM measures the similarity of the two images. A higher value indicates that the two images are more similar. The maximum value is 1. The definition of peak signal-to-noise ratio, the expression is:

$$\begin{aligned} \mathrm{MSE} &= \frac{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(I_0(i,j) - I(i,j))^2}{M \times N} \\ \mathrm{PSNR} &= 10\log\left(\frac{G_f^{\ 2}}{MSE}\right) \end{aligned} \tag{2}$$

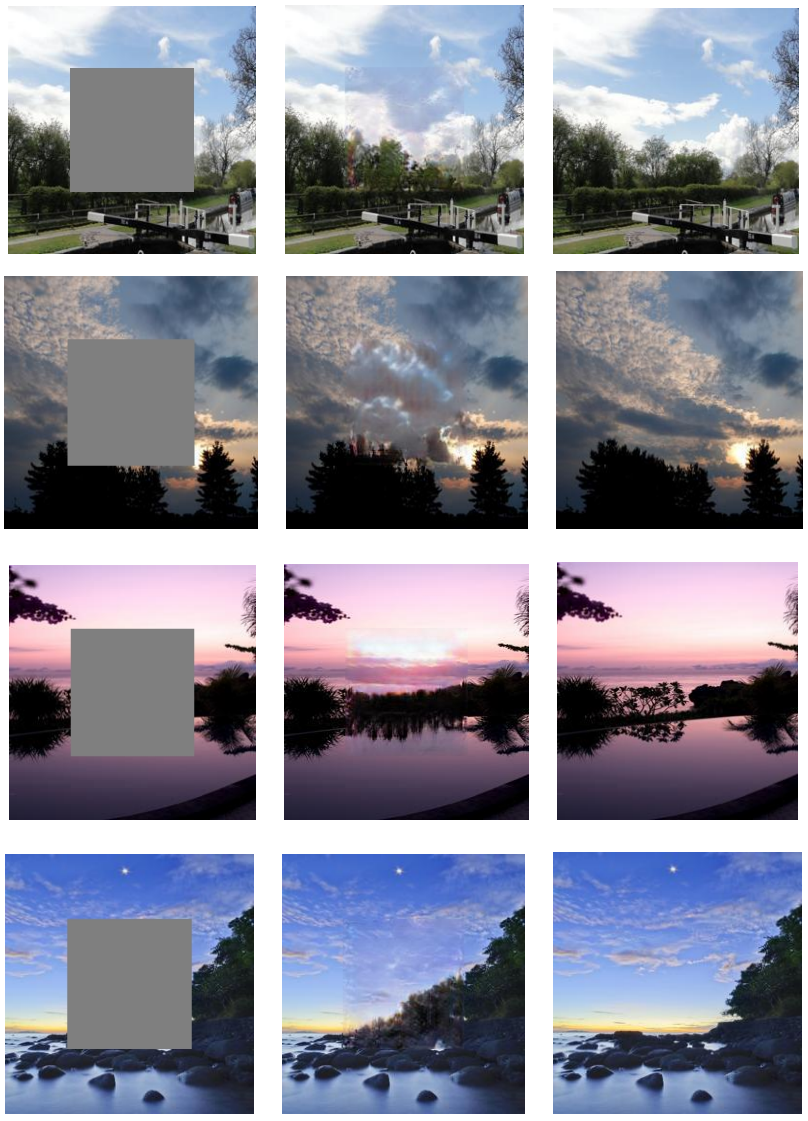
MSE is the mean square error. The default value is 255, $I_0(i,j)$ represents the pixel value at (i,j) in the real image, $I(i,j)$ represents the pixel value at (i,j) in the inpainting image, and M * N represents the area size of the inpainting image.

The definition of structural similaritycan be written as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^{\ 2} + \mu_y^{\ 2} + C_1)(\sigma_x^{\ 2} + \sigma_y^{\ 2} + C_2)} \tag{3}$$

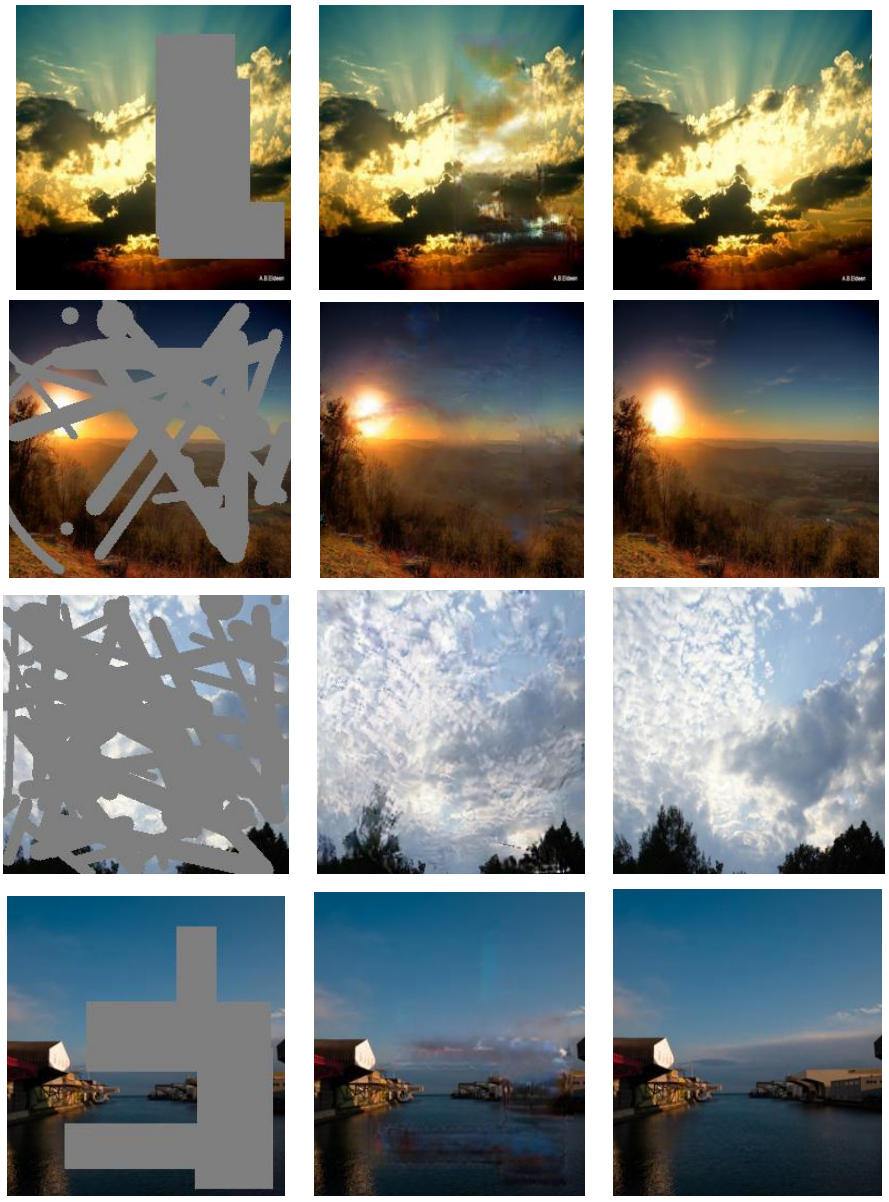
x and y represent the two input images, where μ_x is the average of x, μ_y is the average of y, σ_y^2 is the variance of x, σ_y^2 is the variance of y, σ_{xy} is the covariance of x and y, and C_1, C_2 are Used to maintain a stable constant. L is the dynamic range of pixel values, generally taken as 255.

This paper compares four different image inpainting models, using PSNR and SSIM methods to evaluate.



(a)damaged image (b)inpainting image (c)real image

Figure 7. Inpainting result of intermediate masked



(a)damaged image (b)inpainting image (c)real image

Figure 8. Inpainting result of random masked

TABLE I. EVALUATION RESULTS OF PSNR AND SSIM METHODS

Image inpainting model	PSNR	SSIM
CE[13]	18.72	0.843
GL[14]	19.90	0.836
GntIpt[15]	20.38	0.855
GMCNN[16]	20.62	0.851
Ours	24.06	0.857

V. CONCLUDE AND PROSPECT

In this paper, the image inpainting network structure is built based on GAN. The residual network is used in the encoding and decoding process to reduce the gradient disappearance and gradient explosion problems. Using the loss function of WGAN-GP to update the network parameters to inpainting the image, not only the similarity of the inpainting image structure, but also the matching degree of the image texture. The Place2 dataset is used for network training and testing. The subjective evaluation method and the objective evaluation method are used to evaluate the inpainting image. The objective evaluation method selects SSIM and PSNR to make an objective evaluation of the inpainting image. The comparison between the image inpainting model and the inpainting model of other papers verifies the effectiveness of the algorithm in this paper.

REFERENCES

[1] Bertalmio M, Sapiro G, Caselles V, et al. Image inpainting[C]. international conference on computer graphics and interactive techniques, 2000: 417-424.

[2] Guillemot C, Meur O L. Image Inpainting : Overview and Recent Advances[J]. IEEE Signal Processing Magazine, 2014, 31(1): 127-144.

[3] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.

[4] Ratliff L J, Burden S A, Sastry S, et al. Characterization and computation of local Nash equilibria in continuous games[C]. allerton conference on communication, control, and computing, 2013: 917-924.

[5] Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014:2672-2680.

[6] Pathak D, Krahenbuhl P, Donahue J, et al. Context Encoders: Feature Learning by Inpainting[C]. computer vision and pattern recognition, 2016: 2536-2544.

[7] Yang C, Lu X, Lin Z, et al. High-Resolution Image Inpainting Using Multi-scale Neural Patch Synthesis[C]. computer vision and pattern recognition, 2017: 4076-4084.

[8] Radford A, Metz L, Chintala S, et al. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. arXiv: Learning, 2015.

[9] Zhu J, Park T, Isola P, et al. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks[C]. international conference on computer vision, 2017: 2242-2251.

[10] Kim T, Cha M, Kim H, et al. Learning to Discover Cross-Domain Relations with Generative Adversarial Networks[J]. arXiv: Computer Vision and Pattern Recognition, 2017.

[11] Yi Z, Zhang H, Tan P, et al. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation[C]. international conference on computer vision, 2017: 2868-2876.

[12] Efros A A, Freeman W T. Image quilting for texture synthesis and transfer[C]. international conference on computer graphics and interactive techniques, 2001: 341-346.

[13] Pathak D, Krahenbuhl P, Donahue J, et al. Context Encoders: Feature Learning by Inpainting[C]. computer vision and pattern recognition, 2016: 2536-2544.

[14] Iizuka S, Simoserra E, Ishikawa H, et al. Globally and locally consistent image completion[J]. ACM Transactions on Graphics, 2017, 36(4).

[15] Yu J, Lin Z, Yang J, et al. Generative Image Inpainting with Contextual Attention[C]. computer vision and pattern recognition, 2018: 5505-5514.

[16] Wang Y, Tao X, Qi X, et al. Image Inpainting via Generative Multi-column Convolutional Neural Networks[C]. neural information processing systems, 2018: 329-338.