A Survey of Image Inpainting Research Based on Generative Adversarial Network

Youyu Sun, Baoshan Sun

Abstract— **Purpose The concept of image restoration originated from the manual restoration of murals and other artworks during the European Renaissance. Image restoration technology is different from other processing technologies in the field of computer vision. It has high requirements for image feature extraction. Therefore, restoration techniques based solely on convolution and other methods cannot achieve human visual recognition in terms of restoration effects. The proposal of Generative Adversarial Network (GAN) provides a new idea for the field of image restoration. It adopts the idea of image generation model and discriminant model for adversarial training, and the repair effect is more in line with the characteristics of visual perception. Method Generative confrontation network has more powerful feature learning and feature expression capabilities than traditional machine learning algorithms when performing image processing. In the early stage of a large number of literature research work, it is found that the conditional generative confrontation network CGAN, the deep convolution-based generative confrontation network DCGAN and the Wasserstein generative confrontation network are more widely used. This article mainly introduces the basic ideas and methods of GAN, CGAN, DCGAN and WGAN, and analyzes and summarizes their advantages and disadvantages in image restoration. Conclusion The current research on GAN-based image restoration methods has made a certain degree of progress, but GAN as a new type of network model still needs further research in theory.**

Index Terms— **Generative confrontation network, image restoration, CGANs, DCGANs, WGANs**

I. INTRODUCTION

With the continuous development of computer technology and the emergence of cloud computing, GPU and other technologies, research and applications based on deep learning have received widespread attention from all walks of life. For the field of image restoration, traditional image restoration methods mainly focus on two aspects: structure-based image restoration and texture-based image restoration. On the one hand, structure-based image restoration is based on diffusion-based methods, mainly including BSCB model, TV model, and CDD model. This method mainly uses the edge information of the area to be repaired, determines the direction of diffusion, and diffuses the known information into the edge. On the other hand, the most classic texture-based image restoration method is the Criminisi algorithm based on sample blocks. This method first divides the image into a set, designs a matching principle, and finds the block with the highest similarity to fill the missing area. The traditional repair method has better processing effect in the case of small area damage and the semantic information of the damaged area is not lost, but for large area damage or lack of semantic information, the repair effect presents a large area of fuzzy state

(a) (b)

(c) (d)

(a) Small area is damaged and semantics are not missing;
(b) Small area is damaged and semantics are missing;
(c) large area is damaged and semantics are not missing;
(d) large area is damaged and semantics are missing.

The research method based on deep learning was first proposed in 1943, but it was not popular due to the social development level at that time. Until 2012, in the ILSVRC image recognition competition held by ImageNet, ALexNet based on the deep learning system won the championship, and deep learning returned to people's field of vision. Based on the continuous development and improvement of deep learning, this method is gradually applied to the field of image restoration. The early stage of the image restoration model based on deep learning is mainly based on convolutional neural networks. It uses the unique advantages of convolutional neural networks in training data to generate efficient image structure prediction maps, which overcomes the shortcomings of traditional restoration methods in structural restoration . However, the convolution-based repair method still has obvious artifacts in the texture part, and the repair result is not satisfactory. Taking into account the limitations of convolutional neural networks, in 2014 Goodfellow proposed a new type of network model generative confrontation network. The network consists of two models, one is a generative model and the other is a discriminative model. The confrontation training of the game is continuously optimized until it is mature. Compared with

Manuscript received September 16, 2020

the pure generative model repair results, it shows more powerful feature learning and expression capabilities.

II. IMAGE INPAINTING TECHNOLOGY BASED ON GANs

In this part, we mainly analyze the principles and methods of GANs itself and its derivative models CGANs, WGANs, DCGANs and the advantages and disadvantages of image restoration.

A. GAN principle and current situation

The generative confrontation network model consists of two parts of work, one is a generator that generates fake samples, and the other is a discriminator that is used to determine whether the samples generated by the generator are true. The purpose of the generator is to try to deceive the discriminator so that the discriminator cannot distinguish whether the input data is a real sample or a generated sample; the purpose of the discriminator is to try to distinguish between the real sample and the generated sample, and feed back to the generator a discriminant score to promote the generation of the network Continue to learn and optimize. Therefore, the objective function that the GAN network continuously learns is to solve the maximum and minimum optimization problems, as shown in Equation 1:

$$\min_{G}\max_{D} E[\log D(x)+\log(1-D(G(z)))] \quad ....(1)$$

In Equation 1, D(x) means the real sample G(z) means the generated sample. When the first input is real data, the discriminator makes the value of the objective function as large as possible; when the latter input is generated data, G(z) is as small as possible, and the objective function is as large as possible to deceive the discriminator. The generator discriminator continuously fights against training, and finally reaches the Nash equilibrium. The GAN network is similar to two black boxes, and the trainer can only focus on the input and output. The generator learns the distribution characteristics of the real sample data, and generates a data sample that is as similar as possible to the real training data for the noise z of the input (uniform distribution, Gaussian distribution, etc.); the discriminator can be understood as a binary classifier in essence, responsible for estimation The possibility that a sample comes from the real data set or from the generator. The model structure diagram of GAN is shown as in Fig. 1

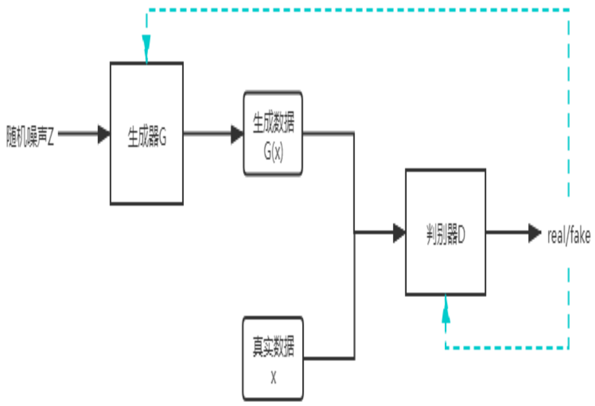

Fig.1 GAN model structure flow chart

GAN-based image repair does not consider the damaged image during training, but uses the generator to directly generate the image to be repaired. In the training phase, the pixel value penalty and the previously trained classification error penalty are used to classify the image repair effect distinguish. The image restoration method trained by GAN has been effectively improved in terms of image restoration details. Figure 2 shows the GAN-based image restoration process.

B. Derivative model of GAN

This section analyzes the improved algorithms for GAN model training instability and easy model collapse. The main ones are CGAN, DCGAN and WGAN. The effects of these three improved methods have been widely cited, so we briefly analyze this method.

i. CGAN

The original GAN does not need to distribute the hypothetical data during data training, but directly samples the method, which really achieves a theoretical approach to the real data. However, this method that does not require pre-modeling is too free, and the effect is not ideal for larger images and more pixels. In order to solve the problem of GAN being too free, a natural idea is to add constraints to GAN so that it can have some pertinence in its training and develop towards the training goal, that is, the core idea of CGAN. Compared with GAN, the objective function of CGAN only adds the condition variable Y:

$$\min_{G}\max_{D} E[\log D(x\,|\,Y)+\log(1-D(G(z\,|\,Y)))] \quad ...(2)$$

Both the generator G and the discriminator D add a condition variable Y, add a condition variable to the model through the Y variable, use the data to control the training of the model, and finally guide the data generation. The condition can represent any information, such as category information, or other modal data, and different scenarios have different meanings.

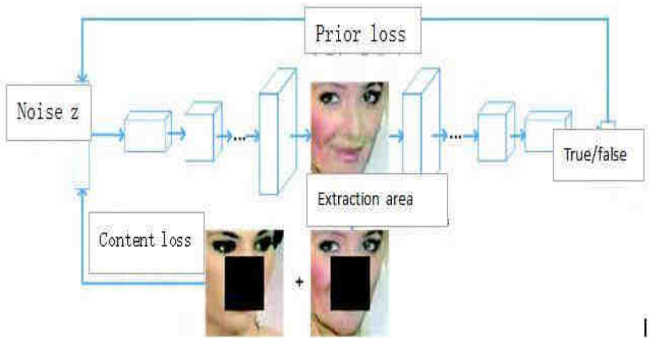

Fig. 2 GAN-based image restoration structure diagram

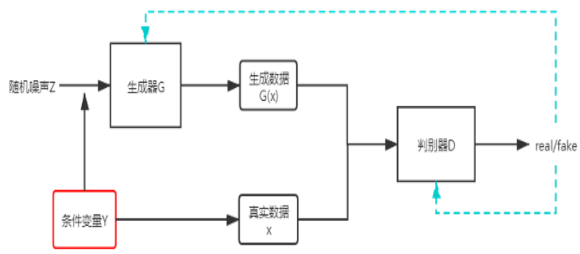

Fig.3 CGAN model structure flow chart

CGAN mainly overcomes the excessive freedom of the GAN network when generating images. It can guide the image to be repaired according to a priori conditions, and to a large extent improve the visual perception of image repair. CGAN can be understood as a transition from a purely unsupervised generative confrontation network to a supervised one. CGAN is mainly used in image restoration for a given draft image

ii. DCGAN

For the original GAN network, there are no hard and fast rules for the network structure of the generator and the discriminator and the original GAN network is a completely unsupervised network model, but in fact, most of the existing problems are supervised . In view of the fact that convolutional neural networks have good applications in some tasks in the field of supervised learning, the effect in the field of unsupervised learning is not obvious. Therefore, hoping to help bridge the gap between CNNs in supervised learning and unsupervised learning, a class called Deep Convolution Generative Adversarial Networks (DCGAN) is proposed. The principles of DCGAN and the original GAN are the same, except that the two structures of GAN, G and D, are adjusted accordingly. The adjustment part is as follows:

1. Cancel all pooling layers in the generator, use transposed convolution, and add stride convolution to the discriminator instead of pooling;
2. Batch regularization: In addition to the output layer of the generative model and the input layer of the discriminant model, the other network layers have joined the BatchNorm layer to stabilize learning;
3. Remove the fully connected layer and use a fully convolutional network;
4. The generator uses RuLU as the activation function, and the last layer uses the Tanh function;
5. Use LeakyReLU as the activation function in the discriminator.

DCGAN is a better improvement after GAN. Its main improvement is in the network structure. Using deep convolution as a generator can obtain the semantic information of the image in a deeper level, providing a good network topology for GAN training. The structure shows that the generated features have the computational properties of vectors. But the stability of GAN training has not been fundamentally solved. The following is the structure of the generator model of the original text in the LSUN experiment as shown in Figure 4.

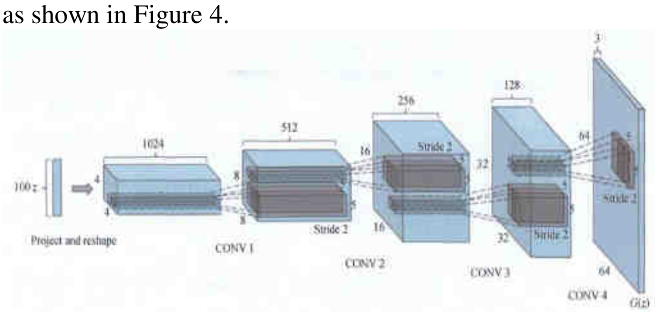

Fig.4 DCGAN model structure diagram

iii. WGAN

The original GAN has the problems of unstable training and mode collapse. The main reason for the unstable training of GAN is that cross entropy (JS divergence) is not suitable for measuring the distance between disjoint parts of the distribution, so the WGAN paper proposes to turn The method of using wassertein distance to measure the distance between the generated data distribution and the real data distribution theoretically solves the problem of unstable training, difficult training, and the loss function of the generator and the discriminator cannot indicate the training process.

WGAN mainly improves GAN from the perspective of loss function, and its main contributions are:

1. Remove the sigmoid function of the last layer of the discriminator;
2. The loss of generator and discriminator does not take log;
3. Force the updated weight to be truncated to a certain range, such as [-0.01, 0.01], to meet the lipschitz continuity condition mentioned in the paper;
4. It is recommended to use optimizers such as SGD and RMSprop.

WGAN mainly improves GANs at the level of loss function, so that the model no longer focuses on the balance problem of generating and discriminating the model, and to a certain extent solves the problem of model training instability. At the same time, cross entropy is used as the training standard to solve the model collapse problem while ensuring the diversity of generated samples.

III. ACKNOWLEDGMENT

In this paper, the structure and principle of GAN model and its different derivative models are studied and analyzed. Through analysis, it is found that image restoration based on GAN network is theoretically feasible. At present, GAN-based image restoration technology performs well for low-resolution and specific environment restoration effects, but for high-resolution images, especially image restoration with more lack of semantic content, it needs to be further improved.

REFERENCES

[1] Goodfellow I J , Pouget-Abadie J , Mirza M , et al . Generative adversarial networks [J] . arXiv preprint arXiv: 1406 . 2661 , 2014 .

[2] Kingma D P , Welling M . Auto-encoding variational Bayes [EB /OL] . 2014-05-01 [2017-11-19] . https: / /arxiv . org / pdf /1312 . 6114 . pdf .

[3] Pu Y C , Gan Z , Henao R , et al . Variational autoencoder for deep learning of images , labels and captions[EB /OL] . 2016-09-28[2017-11-13] . https: / /arxiv . org / pdf /1609 . 08976 . pdf .

[4] Goodfellow I , Bengio Y , Courville A . Deep Learning [M] . Cambridge: The MIT Press , 2016: 26-29 .

[5] Goodfellow I , NIPS 2016 tutorial: generative adversarial networks[EB /OL] . 2017-04-03[2018-03-01] . https: / /arxiv . org / pdf / 1701 . 00160 . pdf .

[6] Salimans T , Goodfellow I , Zaremba W , et al . Improved techniques for training GANs [EB/OL] .

2016-06-10 [2017-12-05] . https: / /arxiv . org / pdf /1606 . 03498 . pdf .

[7] Arjovsky M , Chintala S , Bottou L . Wasserstein GAN [EB /OL] . 2017-12-06 [2018-02-23] . https://arxiv . org / pdf /1701 . 07875 . pdf .

[8] Gulrajani I , Ahmed F , Arjovsky M , et al . Improved training of Wasserstein GANs [EB/OL] . 2017-12-25 [2018-01-19] . https: / /arxiv . org / pdf /1704 . 00028 . pdf .

[9] Xu B , Wang N Y , Chen T Q , et al . Empirical evaluation of rectified activations in convolutional network[EB /OL] . 2015-11-27[2018-0-1-28] . https: / /arxiv . org / pdf /1505 . 00853 . pdf .

[10] Shen J H , Chan T F . Mathematical models for local nontexture inpaintings [J] . SIAM Journal on Applied Mathematics , 2002 , 62(3) : 1019-1043 . [DOI: 10 . 1137 / S0036139900368844]

[11] Tsai A , Yezzi A , Willsky A S . Curve evolution implementation of the Mumford-Shah functional for image segmentation , denoising , interpolation , and magnification [J] . IEEE Transactions on Image Processing , 2001 , 10(8) : 1169-1186 . [DOI: 10 . 1109 /83 . 935033]

[12] Li C , Wand M . Precomputed real-time texture synthesis with Markovian generative adversarial networks [C] / /Proceedings of the 14th European Conference on Computer Vision . Amsterdam , The Netherlands: Springer , 2016: 702-716 . [DOI: 10 . 1007 / 978-3-319-46487-9_43]

[13] Qiang Z P , He L B , Xu D . Exemplar-based pixel by pixel inpainting based on patch shift[C]/ /Proceedings of the 2nd CCF Chinese Conference on Computer Vision . Tianjin , China: Springer , 2017: 370-382 . [DOI: 10 . 1007 /978-981-10-7302-1 _31]

[14] Gatys L A , Ecker A S , Bethge M . Texture synthesis using convolutional neural networks[C]/ /Proceedings of the 28th International Conference on Neural Information Processing Systems . Montreal , Canada: ACM , 2015: 262-270 .

[15] Gatys L A , Ecker A S , Bethge M . Image style transfer using convolutional neural networks[C]/ /Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition . Las Vegas , NV , USA: IEEE , 2016: 2414-2423 . [DOI: 10 . 1109 /CVPR . 2016 . 265

[16] Demir U , Unal G . Inpainting by deep autoencoders using an advisor network [C] / /Proceedings of the 25th Signal Processing and Communications Applications Conference . Antalya , Turkey: IEEE , 2017: 1-4 . [DOI: 10 . 1109 / SIU . 2017 . 7960317]

[17] Wang K F , Gou C , Duan Y J , et al . Gener ative adversarial networks: the state of the art and beyond [J] . Acta Automatica Sinica , 2017 , 43(3) : 321-332 . [王坤峰，苟超，段艳杰，等． 生成式对抗网络 GAN 的研究进展与展望 [J] ． 自动化学报，2017，43 (3) : 321-332 .] [DOI: 10 . 16383 /j . aas . 2017 . y000003