Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/5/20230540

Generative adversarial network based image inpainting

Jingjing Gao

Beijing University of Technology, No.100 Ping Le Yuan, Chaoyang District, Beijing City, PR China, 100124.

gaojingjing@emails.bjut.edu.cn

Abstract. Image inpainting, which is the repair of pixels in damaged areas of an image to make it look as much like the original image as possible. Deep learning-based image inpainting technology is a prominent area of current research interest. This paper focuses on a systematic and comprehensive study of GAN-based image inpainting and presents an analytical summary. Firstly, this paper introduces GAN, which includes the principle of GAN and its mathematical expression. Secondly, the recent GAN-based image inpainting algorithms are summarized, and the advantages and disadvantages of each algorithm are listed. After that, the evaluation metrics, and common datasets of deep learning-based image inpainting are listed. Finally, the existing image inpainting methods are summarized and summarized, and the ideas for future key research directions are presented and prospected.

Keywords: Computer vision, Image inpainting, Deep learning, Generative adversarial networks(GAN).

1. Introduction

Image Inpainting is a technology seeks to repair any damaged pixels in an incomplete image before reconstructing and producing a high-quality, in-depth semantic approximation of the original image. With the swift development of deep learning-related technologies and the application of artificial intelligence science, along with a significant rise in computer processing power, which has significantly advanced both science and technology as well as improved the quality of life for people. Many computer vision applications rely heavily on deep learning-based image inpainting techniques. It has grown to be a significant area of computer vision research.

The significance of studying image inpainting is not only to improve the current research method, but also to expand and improve its application in real life. The main applications are:

1) Object removal: removing unwanted objects from an image and repairing the occluded areas of the objects.

2) Image inpainting: repair pixel loss in images caused by improper processing, such as scratches.

3) Image retouching: Retouch photos of different people, such as getting rid of wrinkles, moles and other facial features.

4) Text removal: Remove unwanted text, watermark, etc. from images. Because of the wide application of image inpainting in real life, this field receives a lot of attention from researchers and has great development prospects.

Traditional image inpainting methods are unable to repair large broken areas or images with complex structures of blends and textures. Therefore, researchers have applied deep learning to computer vision tasks, and related image inpainting methods have emerged as a result.

Among the deep learning models with more prominent inpainting effects are the AutoEncoder (AE) [1], the U-Net [2], the Generative adversarial networks [3], transformer [4], and so on. They train deep models to obtain graphical image high-level semantic information, learning image structure texture information to repair large regionally broken images by training deep models to obtain high-level semantic These methods address the shortcomings of traditional image inpainting and achieve excellent inpainting results.

In this paper, a systematic and comprehensive study is conducted for GAN-based image inpainting methods. Various GAN-based image inpainting methods are analyzed and described, and common datasets and evaluation metrics are listed.

2. Related works

This section briefly introduces the basic principles of GAN and the mathematical representation of GAN.

2.1. Generative Adversarial Networks

GAN(Generative Adversarial Networks) is an unsupervised deep learning model for generating data by computer. The model generates excellent outputs by learning two modules in the framework: the Generative Model and the Discriminative Model. Generative adversarial networks are considered to be one of the most promising and active models currently used in the direction of sample data generation, image generation, image inpainting, image conversion, text generation, etc.

Generator: Generates data by machine, with the aim of "fooling" the discriminator as much as possible, and the generated data is recorded as $G(z)$.

The discriminator: determines whether the data is fake or not, with the aim of finding "fake data" created by the "generator" as much as possible.

Thus, G and D form a gaming process, and as training (adversarial) proceeds, G generates data closer and closer to the real data, and D discriminates the data at a higher and higher level.

2.2. Training Process of GAN

Stage 1: fixing "discriminator D" and training "generator G". Using a discriminator with good performance, G keeps generating "fake data" and then gives it to this D to judge. At the beginning, G is very weak, so it is easy to be discriminated. At this time, D is basically a "blind guess" and has a 1/2 probability of determining right or not.

Stage 2: Fix the "generator G" and train the "discriminator D" (After we will use A to refer to discriminator D and S to refer to generator G). It is meaningless to continue training B after passing the first stage. At this time, we fix B and start training A. At this time, we fix S and start training A. Through continuous training, A improves his discriminative ability, and eventually it can almost determine the fake data.

Repeat the first stage and second stage. Through the continuous cycle, both S and A become stronger and stronger. Eventually we get a S that works very well, and we can use it to generate data.

2.3. Mathematical Expression of GAN

2.3.1. Mathematical expression of GAN. The generative model maps the data into the generative space from an input space (i.e., through the input data, the output data is generated under the action of a

DOI: 10.54254/2755-2721/5/20230540

function). In order to make the generated data distribution closer to the real one, the generating function G typically models a variety of completely different distribution types in the form of a neural network.

Here we use A to refer to discriminator D and S to refer to generator G.

The following is the cost function in the generative adversarial network, taking the discriminator D as an example, the cost function is written J(A)J^{(A)}J(A), and the form is shown below:

$$J^{(A)}\left(\theta^{(A)},\theta^{(S)}\right)=-\frac{1}{2}E_{x\sim P_{data}}\log A(x)-\frac{1}{2}E_{x\sim P_{z}}\log\left(1-A(S(z))\right) \quad (1)$$

The generators and discriminators are closely related and they both can be considered as a zero-sum game, so their combined cost should be zero.

3. Deep learning based image inpainting methods

There are many inpainting methods based on deep learning, here we mainly summarize and list the GAN-based image inpainting methods and summarize the advantages and disadvantages of each method.

Table 1. The characteristics of GAN class image inpainting methods.

Method	Resolution	Loss Function	Form	Advantages	Limitation
DGM [5]	64×64	Guided L1 Confrontation	End-to-end semantic repair	A priori knowledge Context loss	Model instability
GFC	128×128	L2 Contrast pixel softmax	End-to-end semantic repair	Semantic parsing network	Texture blurring
PGN [6]	128×128	L1 vs. TV	End-to-end semantic repair	Course learning Progressive GAN	Irregular region repair
NEO	256×256	L2 against L1	Two-stage flag-guided repair	U-Net flag generator	Boundary apparent
DE-GAN [7]	256×256	Domain embedding multi-model adversarial	End-to-end face inpainting	Domain embedding GAN	Multiple face inpainting
SPG-Net	256×256	L1 vs. TV	Two-stage segmentation-guided repair	Segmentation prediction Segmentation guidance	Texture reimaging
CR-Fill [8]	256×256	Context reconstruction L1 Contrast	Two-stage context repair	Context reconstruction loss	Lower resolution
HR [9]	512×512	TV content L2 Confrontation	Two-stage content texture repair	Image content texture constraint	Texture artifacts
HRII	512×512	L1 hinge contrast Confidence prediction	Two-stage sample iterative inpainting	Feedback mechanism iterative inpainting	Large computational resources
AOTGAN [10]	512×512	Confrontation Reconstruction Style Perception	End-to-end semantic repair	Aggregate context transformation	GAN High computational cost

3.1. GAN class image inpainting methods

The GAN structure-based image inpainting method is to generate the image to be restored directly by a generator, and the input can be random noise. Compared with patch method and Encoder-Decoder

DOI: 10.54254/2755-2721/5/20230540

method, GAN has more advantages in face inpainting, so it is more promising to be used in human daily life.

4. Evaluation index and datasets

4.1. Evaluation index

So as to evaluate the performance of image inpainting methods, researchers have developed different evaluation metrics to evaluate the restored images they generate. Objective evaluation metrics and subjective evaluation metrics constitute the evaluation metrics. Subjective evaluation metrics mainly rely on human subjective judgment ability, so in most cases, we use objective evaluation metrics for quantitative evaluation. Table 2 lists some major complete reference of image evaluation metrics and their characteristics.

Table 2. The characteristics of mage evaluation index.

Evaluation index	Numerical size	Function	Advantages	Limitation
MAE	↓	Measure of image error	Model robustness	Not conducive to model convergence
MSE	↓	Measures image similarity	Accelerates model convergence	Subject to large errors
UQI	↑	Measure of image quality	Determine correlation structure distortion	Difficult to capture correlation
PSNR	↑	Measure of image distortion	Simple and fast	Dependent on pixel point error
SSIM	↑	Measuring structural similarity of images	Introducing structural judgments	Non-structural errors are difficult to determine

Table 3. The characteristics of dataset.

Type	Datasets	Year	Total number	Resolution
Architecture	Facade	2013	606	---
Textures	DTD	2014	5640	---
Street Scene	SVHN	2011	>60000	32×32
	Paris StreetView	2012	15000	936×537
	Cityscapes	2016	25000	2048×1024
Scene	ImageNet	2015	14197122	---
	Places2	2017	1000000	256×256
Face	Helen Face	2012	2000	---
	CelebA	2015	202599	178×218
	FFHQ	2019	70000	1024×1024

DOI: 10.54254/2755-2721/5/20230540

4.2. Datasets
Deep learning-based image inpainting methods need experiments on a large amount of images to evaluate the effect of the method and to learn image feature information by training a large amount of images. However, it is very difficult to collect images and corresponding broken images by oneself. Therefore, researchers usually use public image datasets for training and testing. Some of the datasets frequently used by researchers are given in Table 3.

5. Conclusion
Image inpainting is an irreplaceable part in the field of computer vision. With the lightning speed development of computers and frequent use of digital tools in recent years, the field of image inpainting has also received more attention from researchers. The research on image inpainting based on deep learning is relatively short, but the progress flies by optimizing in terms of model structure, loss function, and prior information to obtain better inpainting results. There we mainly summarize Gan-based image inpainting, and briefly summarize the common datasets and evaluation metrics for image inpainting. The following descriptions are given for the shortcomings of existing image inpainting methods to advance future research work.

5.1. High-resolution image inpainting
Studying low-computational cost high-resolution image inpainting models is one of the most urgent tasks today. More current image inpainting methods still focus on the study of low-resolution image inpainting. However, with the development of the data era, it is obvious that low-resolution images can no longer meet the demand for commercial use.

5.2. Create a dataset based on Asian face images
How to create a dataset based on Asian face images is one of the key directions for future research. Currently, deep learning-based inpainting methods have achieved better inpainting results on face datasets. However, the face datasets that are heavily used at present are all datasets of European and American face images. The models trained with these datasets will show inaccurate and even wrong inpainting results when restoring Asian faces. Therefore, collecting face datasets with Asian facial features is one of the priorities of the current research.

5.3. Enables face image inpainting in different tasks and scenarios
How to implement face image inpainting in different tasks and scenarios is a challenge that needs to be solved in the future. In daily life, face image inpainting is used in many applications, such as in public safety and face recognition. However, face image inpainting in different tasks and scenarios will have many challenges that cannot be predetermined in advance, such as face image inpainting when people wear masks under epidemic normalization. All these problems will enhance the difficulty of face image inpainting, so collecting and organizing broken face image datasets for different tasks and scenarios is one of the research hotspots in the future perhaps.

References
[1] RUMELHART D E, HINTON G E, WILLIAMS R J. (1985). Learning internal representations by error propagation[R]. California Univ San Diego La Jolla Inst for Cognitive Science.
[2] RONNEBERGER O, FISCHER P, BROX T. (2015). U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, Munich, Germany, 234-241.
[3] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. (2014). Generative adversarial nets[J]. Advances in neural information processing systems, 27.
[4] VASWANI A, SHAZEER N, PARMAR N, et al. (2017). Attention is all you need[J]. Advances in neural information processing systems, 30.

DOI: 10.54254/2755-2721/5/20230540

[5] YEH R A, CHEN C, YIAN LIM T, et al. (2017). Semantic image inpainting with deep generative models[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu, HI, USA: 5485-5493.
[6] ZHANG H, HU Z, LUO C, et al. (2018). Semantic image inpainting with progressive generative networks[C]//Proceedings of the 26th ACM international conference on Multimedia. Seoul, South Korea, 1939-1947.
[7] ZHANG X, WANG X, SHI C, et al. (2022). De-gan: Domain embedded gan for high quality face image inpainting[J]. Pattern Recognition, 124: 108415.
[8] ZENG Y, LIN Z, LU H, et al. (2021). Cr-fill: Generative image inpainting with auxiliary contextual reconstruction[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Canada: 14164- 14173.
[9] YANG C, LU X, LIN Z, et al. (2017). High-resolution image inpainting using multi-scale neural patch synthesis[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu, HI, USA: 6721-6729.
[10] ZENG Y, FU J, CHAO H, et al. (2022). Aggregated contextual transformations for high-resolution image inpainting[J]. IEEE Transactions on Visualization and Computer Graphics.