Image Inpainting Using Double Discriminator Generative Adversarial Networks

Chao Zhang¹ and Tong Wang¹,*

¹Engineering Research Center of Digitized Textile & Apparel Technology, Ministry of Education, College of Information Science and Technology, Donghua University, Shanghai 201620, P. R. China

*Corresponding author e-mail: wangtong@dhu.edu.cn

Abstract. The image inpainting algorithms based on GAN have problems such as color inconsistencies of the inpainted region with the surrounding part and the mode collapse during the training. This paper proposes an inpainting algorithm based on double discriminator generative adversarial networks (GANs). This algorithm introduces the loss function of WGAN-GP into a global discriminator and local discriminator independently. It combines the Mean-Square Error (MSE) and feature reconstruction loss function to train the inpainting model and repair the missing area. The algorithm uses more dilated convolutions instead of standard convolutions to obtain a larger receptive field and the skip connection to enhance the structural prediction ability of the generator. The experiment results on the labeled faces in the wild dataset show that our algorithm improves the quality of image inpainting and the stability of the training process.

1. Introduction

With the rapid development of the Internet and the popularization of smartphones, images and videos are playing an increasingly important role in information transmission. The quality of images and videos has become a topic of great concern. This makes the image inpainting attract more and more attention from researchers. Image inpainting refers to repairing the missing area in an image or video using known information, which ultimately makes it impossible for the observer to distinguish whether the image has been inpainted or not[1].

Traditional image inpainting methods can be divided into two categories: texture-based and structure-based image inpainting methods. The structure-based inpainting algorithm is applied to repair small-sized missing areas. The representative one is the Bertalmio Sapiro Caselles Ballester (BSCB) inpainting model proposed by Bertalmio et al. [2]. Texture based inpainting algorithm is good at repairing large-sized missing areas, for example, the texture synthesis algorithm based on sample blocks proposed by Criminisi et al. [3].

In recent years, with the rapid development of deep learning, convolutional neural networks(CNN) have been widely used in various fields due to the ability to extract deep features, combined with deep learning of image inpainting gradually become the mainstream research. However, image inpainting based on convolutional neural networks has the problems of blurred repair results and inconsistent context semantics.
The inpainting algorithm proposed in this paper is based on the double discriminator GANs. We aim to ensure the inconsistencies of the repair area with the surrounding part and to solve the instability of the training of GANs. The main contributions of this paper are as follows:

1. The image inpainting algorithm is based on double discriminator GANs together with a WGAN-GP loss function to improve training stability and prevent mode collapse.

2. More dilated convolutions instead of standard convolutions are used to increase the receptive field without changing the number of parameters, and also the skip connection to fuse the feature information of the bottom layer and the high layer to improve the prediction accuracy of the generator.

3. The feature reconstruction loss is added to MSE loss to get the structural loss. The structural loss can strengthen the structural information in the missing area.

2. Related work

2.1. Generative Adversarial Networks

In 2014, Goodfellow proposed Generative Adversarial Networks (GANs), which consists of a generator and a discriminator [4]. The model of GANs is shown in Figure 1. Martin Arjovsky et al. proposed to introduce the Wassertein distance into the GANs model[5]. Gulrajani proposed an improved algorithm WGAN-GP for WGAN[6].

![Figure 1. The model of GANs.](image1)

2.2. Image inpainting based on GAN

As shown in Figure 2, an encoder-decoder is used as the generator, the encoder extracts depth features from the image to be repaired, and the decoder restores the extracted features to the repaired image.

![Figure 2. Image inpainting based on GANs.](image2)

Pathak et al. proposed a combination of encoder-decoder and generative adversarial networks for image inpainting[7]. Iizuka et al. proposed to use double discriminators to ensure the continuity of pixel values in the boundary region of the repaired image[8]. Yu et al. proposed a two-stage inpainting network and used two WGAN-GP losses and MSE loss to train the model, and the inpainting network converges faster than Iizuka's network, and the results are more accurate[9].
3. Proposed algorithm

3.1. The structure of the network

In this paper, a double discriminator is used to GANs as the image inpainting model. The overall network architecture is shown in Figure 3.

![Figure 3. Overall network architecture.](image)

3.2. Generator network

The generator network uses the encoder-decoder based on the full convolution network (FCN). Five-layer dilated convolutions are used in the encoder, which can increase the receptive field and increase the authenticity of the output pixel when the parameter amount is unchanged. Besides, skip connection is introduced into the generator network. The skip connection can merge the features of the bottom layer and the high layer, improve the structural prediction ability of the generator, and accelerate the convergence of the network.

The specific parameters of the generator network are shown in Figure 4. Conv is the standard convolution, Dconv is the dilated convolution, and Deconv is the deconvolution. Except that the activation function of the output layer is Tanh, LeakyRelu is used as the activation function of the other layers.

![Figure 4. Architecture of the generator network.](image)

3.3. Discriminator network

Both the global discriminator and local discriminator networks use convolutions to extract the features of the input image.

Figure 5 (a, b, c) shows the specific parameters of the local discriminator, the global discriminator, and the fully connected layer independently. Both the local discriminator and the global discriminator use LeakyReLU as the activation function.

![Figure 5. The architecture of the discriminator network.](image)
3.4. Loss function

The loss function used in this paper includes three parts: MSE loss of the repaired image and the real image, WGAN-GP loss, and feature reconstruction loss.

The MSE loss used in this paper is defined as follows:

$$L_{mse} = \|M_c \odot (G(x \odot M_c) - x)\|^2$$  \hspace{1cm} (#1)

where $M_c$ is a binary mask, with 0 representing the missing position, and $x$ is the real image.

The WGAN-GP loss is defined as follows:

$$L_d = E[D(G(x \odot M_c))] - E[D(x)] + \lambda E[\|\nabla_x D(x)\|^2 - 1]^2$$ \hspace{1cm} (#2)

$$L_{gan} = -E[D(G(x \odot M_c))]$$ \hspace{1cm} (#3)

The feature reconstruction loss is defined as follows:

$$L_{feat} = \|C(G(x \odot M_c)) - C(x)\|^2$$ \hspace{1cm} (#4)

where $C$ is a feature extractor, using a pre-trained VGG-16 network.

Feature reconstruction loss and MSE loss form a structural loss, which is used to train the generator network separately:

$$L_{struct} = L_{mse} + L_{feat}$$ \hspace{1cm} (#5)

The structural loss and the WGAN-GP loss form a joint loss, which is used to train the generator network:

$$L_g = L_{struct} + \alpha L_{gan}$$ \hspace{1cm} (#6)

where $\alpha$ is a hyperparameter, $\alpha$ is set to 4e-4 in this paper.

3.5. Algorithm flow

This paper chooses to use the ADAM optimizer to train the generator network, the learning rate is 2e-4, and the RMSProp optimizer is used to train the discriminator network, and the momentum is 0.5.

An overview of the training process is stated in Algorithm 1. Use Eq.(1) to calculate the MSE loss $L_{mse}$. Use Eq.(4) to calculate the feature reconstruction loss $L_{feat}$. Using the Eq.(5), add $L_{mse}$ and $L_{feat}$ to get $L_{struct}$. Use Eq.(2) calculate the adversarial loss $L_d$. Use Eq.(3) to calculate generator adversarial loss $L_{gan}$. Use Eq.(6) to calculate $L_g$.

Algorithm 1 Training procedure of the network

1: while epoch<200 do
2:     Sample a minibatch of images $x$ from the training set;
3:     Generate masks $M_c$, with 0 representing the missing area for each image in $x$;
4:     $\bar{x} = x \odot M_c$, $\bar{x}$ is the image to be repaired, send $\bar{x}$ into the network $G$ to get the generated image $\bar{x}$;
5:     $x_{completion} = \bar{x} \odot (I-M_c)$, $x_{completion}$ is the repaired image;
6:     if epoch <20 then
7:         Calculate $L_{struct}$ to train and update the generator network $G$;
8:     else
9:         Randomly generate a mask image $M_d$, which represents the local area of each image in $x$;
10:        $x_{local\_fake} = x_{completion} \odot M_d$, $x_{local\_real} = x \odot M_d$
11:        Send $x_{completion}$ and $x_{local\_fake}$ and $x_{local\_real}$ to $D_g$ and $D_l$, respectively, and calculate the $L_d$;
12:        Use $L_d$ to train and update the discriminators $D_g$ and $D_l$;
13:        if epoch > 40 then
14:            Calculate $L_g$ to update generator network $G$;
15:        end if
16:     end if
17:     end if
18: end while

4. Experiment and results

4.1. Experimental details

The experiment in this paper uses the LFW dataset, which consists of 13233 face images of 128×128×3 size[10]. The platform of this experiment is Ubuntu 16.04.1, and the graphics card is NVIDIA 2080Ti. Use the programming language Python 3.6.8 and the deep learning framework Tensorflow-GPU 1.12.0 in the experiments.
4.2. Analysis of experimental results

The comparison between the MSE loss training generator and the structural loss training generator is shown in Figure 6 independently. The addition of feature reconstruction loss greatly enhances the generator’s structure generation ability.

![Figure 6. Effects of two loss training generators.](image)

To analyze the repair ability of our algorithm, two representative algorithms based on GAN in recent years (reference [7, 8]) were selected and compared under the same conditions. The repaired results are shown in Figure 7. From the perspective of visual experience, the repair effect of the algorithm in this paper is better in the case of medium area missing. The algorithm in reference [7] has no global discriminator and uses the original GAN loss function, which leads to distortion and blurring in the repair area. Compared with the algorithm in reference [7], the algorithm of reference [8] produces better quality images. Still, it has a blurry part in the repaired area due to the selection of the loss function and some shortcomings in the details of the network model. Our algorithm uses global and local discriminators to ensure that the repaired image maintains the consistency of semantics and structure globally and locally. The introduction of skip connection and feature reconstruction loss makes the generator's structural prediction ability enhanced.

![Figure 7. Comparisons with existing algorithms](image)

The values of PSNR and SSIM of the repaired results in three algorithms are shown in Table 1. SSIM is a full-reference image quality evaluation index, which measures image similarity from three aspects of brightness, contrast, and structure. PSNR is based on the error between corresponding pixels, the higher the PSNR, the closer the image, and the original image. So combining the above two indicators can be seen, for a medium-area missing image, the repair quality of the algorithm proposed in this paper is higher than that of the other two algorithms.
Table 1. PSNR and SSIM of the repaired results.

| Index | reference [7] | reference [8] | Ours |
|------|--------------|--------------|-----|
| PSNR | 29.4         | 31.2         | 32.3 |
| SSIM | 0.807        | 0.822        | 0.847 |

5. Conclusion
In this paper, the image inpainting algorithm is proposed based on double discriminator GANs. The WGAN-GP loss makes the training process more stable and solves the problem of mode collapse. The skip connection enhances the structural prediction ability of the generator. The feature reconstruction loss makes the details of the repair results more realistic. At last, the experimental results show that our algorithm is effective in repairing small and medium missing regions. However, it still has some problems in processing complex details, such as image blurring, which will be emphasized in future research.

6. Acknowledgments
This work was financially supported by "the Fundamental Research Funds for the Central Universities".

References
[1] Marcelo Bertalmio, Guillermo Sapiro, Vicent Caselles, et al. Image inpainting[C]. Conference on Computer Graphics & Interactive Techniques. DBLP, 2000:417-424.
[2] Xiaobao Lu, Weilan Wang, Duojie Zhuoma. A Fast Image Inpainting Algorithm Based on TV Model[J]. Lecture Notes in Engineering & Computer science, 2010, 2181(1):1457-1460.
[3] Telea, Alexandru. An image inpainting technique based on the fast marching method. Journal of graphics tools, 2004, 9.1:23-34.
[4] Goodfellow Ian J., Pouget-Abadie Jean, Mirza Mehdi, et al. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.
[5] Yue Cao, Bin Liu, Mingsheng Long, et al. HashGAN: Deep Learning to Hash with Pair Conditional Wasserstein GAN[C]. IEEE/CVF Conference on Computer Vision & Pattern Recognition. IEEE, 2018:1287-1296.
[6] Gulrajani Ishaan, Ahmed Faruk, Arjovsky Martin, et al. Improved Training of Wasserstein GANS[J]. Advances in Neural Information Processing Systems, 2017:5767-5777.
[7] Pathak Deepak, Krahenbuhl Philipp, Donahue Jeff, et al. Context Encoders: Feature Learning by Inpainting[J]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016:2536-2544.
[8] Satoshi Iizuka, Edgar Simo-Serra, Hiroshi Ishikawa. Globally and locally consistent image completion[J]. ACM Transactions on Graphics, 2017, 36(4):1-14.
[9] Jiahui Yu, Zhe Lin, Jimei Yang, et al. Generative Image Inpainting with Contextual Attention[J]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018:5505-5514.
[10] Erik Learned-Miller, Gary Huang, Aruni RoyChowdhury, et al. Labeled faces in the wild: A survey. Advances in face detection and facial image analysis. Springer, Cham, 2016:189-248.
[11] Fazil Altinel, Mete Ozay, Takayuki Okatani. Deep structured energy-based image inpainting. International Conference and Pattern Recognition (ICPR). IEEE, 2018: 423-428.
[12] Yijun Li, Sifei Liu, Jimei Yang, et al. Generative Face Completion[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017:3911-3919.
[13] Qianru Sun, Bernt Schiele. Natural and Effective Obfuscation by Head Inpainting[J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:5050-5059.
[14] Fei Q, Junyu Han, Pengjin Wang, et al. Structure guided fusion for depth map inpainting[J]. Pattern Recognition Letters, 2013:70-76.
[15] Alain Horé, Djemel Ziou. Image Quality Metrics: PSNR vs. SSIM[C]. International Conference on Pattern Recognition. IEEE Computer Society, 2010:2366-2369.