Computer Systems Science & Engineering
DOI:10.32604/csse.2022.020176
Article

Tech Science Press

A New Method of Image Restoration Technology Based on WGAN

Wei Fang[1,2,*], Enming Gu[1], Weinan Yi[1], Weiqing Wang[1] and Victor S. Sheng[3]

[1]School of Computer & Software, Engineering Research Center of Digital Forensics, Mininstry of Education, Nanjing University of Information Science & Technology, Nanjing, 210044, China
[2]Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, Suzhou, 215325, China
[3]Texas Tech University, USA
*Corresponding Author: Wei Fang. Email: hsfangwei@sina.com
Received: 12 May 2021; Accepted: 16 July 2021

Abstract: With the development of image restoration technology based on deep learning, more complex problems are being solved, especially in image semantic inpainting based on context. Nowadays, image semantic inpainting techniques are becoming more mature. However, due to the limitations of memory, the instability of training, and the lack of sample diversity, the results of image restoration are still encountering difficult problems, such as repairing the content of glitches which cannot be well integrated with the original image. Therefore, we propose an image inpainting network based on Wasserstein generative adversarial network (WGAN) distance. With the corresponding technology having been adjusted and improved, we attempted to use the Adam algorithm to replace the traditional stochastic gradient descent, and another algorithm to optimize the training used in recent years. We evaluated our algorithm on the ImageNet dataset. We obtained high-quality restoration results, indicating that our algorithm improves the clarity and consistency of the image.

Keywords: Image restoration; WGAN; DCGAN; context semantic

1 Introduction

Image inpainting has been an important topic in the area of computer vision for many years. We often make reasonable assumptions and calculations about the missing pixels through some algorithms, and then use the calculated "fake" pixels to satisfy the visual and semantic rationality. When we deal with certain image restoration tasks in daily life, we will find ourselves doing the same.

Different degrees of damage and damage to different areas of the picture determine the difficulty of repair. Generally, the fewer pixels in an image, the more blurred the pixels, and the more difficult it is to restore.

Although image inpainting technology has been studied for decades, it is still a very challenging subject in the field of computer vision and graphics. Some of the widely used image restoration techniques are shown below:

1) Partial and variational changes are used to build *a priori* data model and then to fill in the missing information.

2) The image is obtained by sampling the image to find an image block that is similar to the missing portion. This method excels in solving image restoration problems with missing backgrounds. However, since this method requires looking for missing pixels in the background, it is not possible to patch the portion of the image that does not appear in the background. For example, when we try to restore images with missing faces, this method cannot repair the non-repetitive content because they cannot capture the overall structure of the image to be restored.

3) The learning network is built by deep learning to learn the characteristics of images, and we use the confrontation generation network of depth convolution for image restoration based on image semantics. We use deep learning to construct the nonlinear relationship of the sample space, and then use the features and nonlinear relationships to restore the incomplete image that is input.

At present, the application of deep learning in image restoration has received great attention. Deep learning is widely used in repair tasks, such as improving image resolution and image coloring. For example, we can use CNN for image recognition tasks, using DCGAN to classify identification tasks [1]. Deep learning is now used to accomplish image inpainting tasks through some trained image models that can be used to repair images for specific damage situations.

To some extent, deep networks can generate some missing pixel information for specific situations [2]. Let's take a rectangular fix area as an example. If a well-trained deep learning network can better repair the image of 32 × 32 missing area, but it may not show good results in image with 10 × 10 missing area.

Recent work done by Li et al. [3] shows that more effective results can be achieved by optimizing intermediate neural networks. The optimized intermediate neural network will generate style images in some low convolutional layers.

The main contributions of this paper can be summarized in three aspects:

- Compared with the current popular DCGAN-based image inpainting method, we use WGAN [4] technology. In this way, no matter what the shape and location of the lost area, we can perform image restoration based on the image context semantics.
- In the image inpainting experiment, we used ImageNet Stanford Vision Lab [5] as the data set and selected enough image samples to enhance the training of the image model.
- Our model can generate repair results that are beautiful and visually more similar to the original image.

2 Related Work

In this section, we will introduce the development of image restoration, and also explain some of the ideas and principles of GAN used in our work, so as to lay a theoretical foundation for our proposed algorithm in the following chapter.

2.1 Image Inpainting Issues to Be Solved

In the reign of image inpainting, the repair of smaller areas firstly takes the consistency of the boundary and region into account. Here are some basic ideas that diffusion repair must be based on the local information [6]. However, it is difficult to apply this kind of thinking to the repair of larger areas. That means, in the large area repair, we usually consider the high-level semantic information of the repair objects firstly. As a result, the process of inpainting area based on the massive priori accumulated information. We focus on solving the problem of inpainting the high-level semantic information in images.

Since the image repair method has been widely studied, we can divide the different types of methods into three categories according to the different technology used [7]. There is the calculus of variations based on partial derivative, sample-based image inpainting methods which originated from texture synthesis technology and mixed image inpainting methods. [8] The deep-learning method is a kind of new method which was proposed in recent years.

2.2 Deep Learning-Based Image Inpainting

The inpainting method based on deep learning can be divided into two types according to the network structure of deep learning: first of all, the deep-learning image inpainting method based on the convolutional self-coding network, and the second method is based on the generative adversarial network. It greatly improves the result of image inpainting. One of the more popular improvements is based on the DCGAN or LSTM method [9]. The following type of method is based on the structure of a recurrent neural network (RNN).

Since October of 2014, Goodfellow et al. invent a class of machine learning systems called generative adversarial network [10]. The GAN network has become one of the most promising methods in recent years in the complex distribution of unsupervised learning. In all, given a training set, this technique learns to generate new data with the same statistics as the training set. For example, a GAN trained on photographs can generate new photographs that look at least superficially authentic to human observers. In image inpainting, it can generate the loss zone of the image. As a result, the use of the GAN network in the image inpainting becomes a direction for researchers to explore.

2.3 The Basic Process of Generative Adversarial Networks

The basic structure of the GAN network consists of a generative network G and a judgment model D. In essence, discriminative model D is a classifier, whose input images are real images from the training data set in the training process. In the training of generative network G, it distinguishes candidates produced by the generator from the true data distribution. It is usually a binary classifier in the form of CNN. Generative network G is an anti-convolution neural network that converts random input values into images.

In the training stage, we don't consider the damage of images. It means that we use undamaged data for GAN network training for the first step. When the training is complete, generative network G has the ability to generate new images from a sample distributed by a noise signal z (distributed as Pz). For the repair of a broken image, we can generate a new sample that is similar enough to the known part of Image I by generating network G. As a conclusion, the image inpainting method based on GAN structure is to generate the image by the generator. The input of the generator can be random noise. Thus, the image inpainting method based on GAN is different from the image inpainting method based on self-coding structure. The image inpainting method based on the self-coding structure is to generate the inpainting area by the whole broken image.

2.4 The Application and Problems of GAN

As a generation network with self-supervised learning ability, GAN enables the generative network G to generate images of a specified type for the confrontation training of a certain type of image. The most successful application is to generate a realistic face image. At present, the most successful method [11] is based on the DCGAN network. It trains the face generation model on the CelebA data set. Through the trained generation network G, we can generate a face image with $64 \times 64 \times 3$ pixels by the 100-dimensional vector. At the same time, Guilin Liu proposed a method called partial convolution [12] to achieve the inpainting of irregular image regions. Image restoration can achieve good results in low-resolution images, especially for specific types of image restoration. However, for the target of high-resolution image restoration, GAN training is very difficult, especially for images that are different in

content [13]. The situation is more difficult when we face the big hole of the image. We will make further improvements to this.

3 The New Approach of Image Restoration

The basic idea of this article is to get the optimal forgery function by combining the function of context loss and perceived loss. We defined a loss function to find the best fake image from the collection of fake images. This loss function consists of two parts: context semantic loss and perceived loss, which ensures the similarity of the image to be repaired and the content of the original image, and the semantic and visual authenticity of the restored image. We map the image to a smaller potential space by modifying the reverse propagation algorithm of the loss function. And then we input the mapped vector into the WGAN network to generate the best-forged image of the image to be repaired. Finally, we use the Poisson fusion algorithm to make the final results more realistic and effective.

To further expand the acceptance domain and the stability of training, we introduced a two-stage network structure from coarse to fine. The first network makes initial predictions, and the second network uses the output of the previous networks as input to predict more detailed and effective results. The former network is trained with reconstruction losses, and the second network is trained using a combination of reconstruction loss and GAN loss. The two-stage network structure is essentially more similar to the learning method of residual learning or deep supervision. In the implementation of the network layer, we used the mirror fill method for all convolutional layers, and deleted the batch normalization layer, because we found that it will destroy the color consistency of the image. We give up using the tanh and sigmoid functions. In addition, we also found that it is more efficient that to separate the global feature representation of the GAN training from the local feature representation during the experiment. The schematic diagram of the context encoder is shown in Fig. 1.

Figure 1: In the context encoder, we generally reduce the number of parameters in the network by using a channel-wise fully connected layer which employs convolution to pass information between channels

3.1 Improved Network Structure in Image Inpainting

Before anything else, we used a multi-layer perceptron to process the MNIST handwritten data set. After this, we used a deep convolutional network to construct the generator to implement more complex image generation tasks. The intermediate hidden layer of the generator network structure changes as the image processing complexity changes. Fig. 2 shows the network structure of the image generation network and its associated parameter settings for image restoration based on contextual semantics. The network will perform n loop operations after receiving the input image, and then will generate and output an image too. The loop operation can extract the feature information of the image more effectively, which replaces the design of the deep network [14]. This will effectively reduce the number of parameters of the network model, making the network model more comprehensive in learning the image texture and spatial structure and with high efficiency.

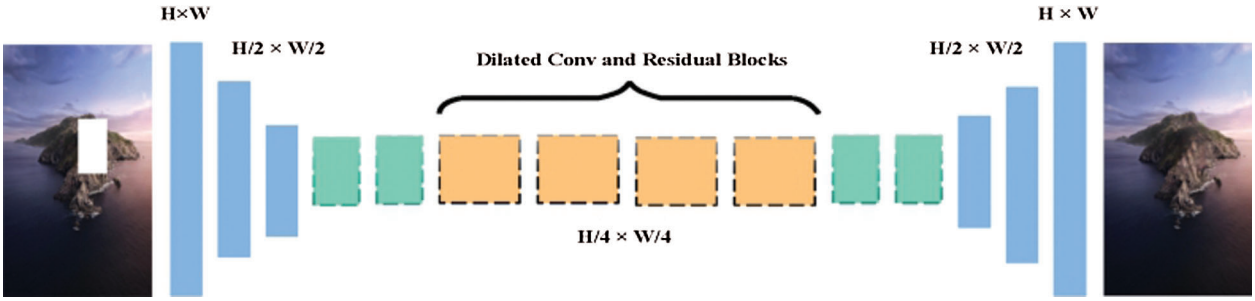

Figure 2: Overview of our generative inpainting framework. We use ELU as the activation function and remove the sigmoid activation function from the last layer in the network structure

3.2 Generator with Batch-Norm

The main function of the discriminator is to continuously improve the accuracy of identifying samples by confrontation and learning with the generator. In our article, the discriminator is composed of a series of convolutional layers. Its structure is shown in the figure below. Conv in the figure represents the convolutional layer, and full connection represents the fully connected layer. Generally speaking, the discriminator has two types of inputs, namely the image generated by the generator and the real image. In the network structure shown in the figure below, the real sample is input. At the same time, we also adopt the idea of a conditional confrontation network. We use the broken image as a condition to splice the condition with the original input as a final input to the discriminator. The network model will output a scalar to represent the discriminator's evaluation of the input data. The purpose of adding a broken image to the judger as a condition is that we will give the generated image a high score only if the generated sample is sufficiently realistic and matches the broken image. In the words, we want to fix the images as closely as possible with the original image. Because we use the Wasserstein distance which was first proposed in 2017 to measure the difference between the two distributions, we need to perform independent Lipschitz constraints on each input sample of the discriminator. Then we can avoid the interdependence between different samples in the same batch. And we remove the sigmoid activation function from the last layer in the network structure.

3.3 The Goal of GAN's Optimization

The difference from other GANs that use ReLU as an activation function is that we use ELU as the activation function. Because compared to ReLU, ELU does not make the derivative abrupt at any point, and it can also produce a negative output. Fig. 3 shows the difference between ReLu and ELU. This function tends to converge results to zero at a faster rate and often achieves more accurate results. Its function definition and image are as follows:

$$R(z) = \begin{cases} z & z > 0 \\ \alpha(e^{Z} - 1) & z \leq 0 \end{cases} \tag{1}$$

Different from the previous image restoration network that relies only on DCGAN for surveillance, we use the WGAN-GP method and use WGAN-GP loss as the loss function of the image restoration task. In general, this method performs well when used in conjunction with l_1 distance.

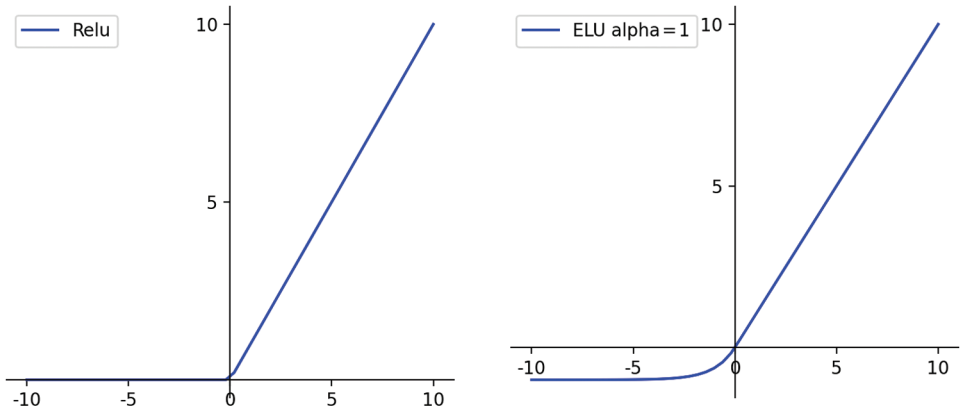

Figure 3: The graph of ELU and relu. ELU is different from relu. obviously, it increases in x ∈ R

Specifically, this method compares the generated data distribution with the actual data distribution by using the Earth-Mover distance. The learning objective function we built by applying Kantorovich-Rubinstein is as follows:

$$\min_{G}\ (\max_{D\in D}\ (E_{x\sim P_r})[D(x)] - E_{\tilde{x}\sim P_g})[D(\tilde{x})])) \tag{2}$$

where D* is a set of functions that satisfy Lipschitz's continuous conditions Stanford Vision Lab. P_g is a distribution defined by implicit $\tilde{x} = G(Z)$. Z is the input to the generator. This is in the neural network as follows:

$$max\left(\mathrm{E}_{\mathrm{x}\sim\mathrm{P_r}}[\mathrm{f_w}(\mathrm{x})] - \mathrm{E}_{\mathrm{z}\sim\mathrm{p(z)}}[\mathrm{f_w}(\mathrm{g}_\theta(\mathrm{z}))]\right) \tag{3}$$

This neural network is very similar to the Discriminator in GAN. There are only a few subtle differences. The last layer of Critic discards sigmoid. Critic's objective function has no log entries. Critic has to truncate the parameters in a certain range after each update. Gulrajani et al. [15] improved the WGAN approach by introducing gradient penalties $\lambda E_{\hat{x}\sim P_{\hat{x}}}(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2$. Among them, $\hat{x}$ is the point on the straight line through the sampling of the P_g and P_r. The gradient of D^* in the line $\hat{x} = (1-t)x + t\tilde{x}$ will point to the current $\hat{x}$, so $\nabla_{\hat{x}} D^*(\hat{x}) = \frac{\tilde{x}-\hat{x}}{\|\tilde{X}-\hat{X}\|}$. For image inpainting tasks, we only introduce a gradient penalty for missing pixels.

3.4 Our Proposed Algorithm

Algorithm 1: Training of our proposed framework

while G has not converged **do**

for i = 1, …, n critic **do**

Sample batch images x from training data;

Construct $t \sim U[0,1]$;

Select random masks m for x;

$\mathrm{z} \leftarrow \mathrm{x}^*\mathrm{m}$;

$\hat{x} \leftarrow \mathrm{z} + \mathrm{G}(\mathrm{z},\mathrm{m}) \odot (1-\mathrm{m})$;

$\hat{x} \leftarrow (1-t)x + t\hat{x}$;

Update two critics with x, $\hat{x}$ and $\tilde{\boldsymbol{x}}$;

(Continued)

Algorithm 1 (continued)

end for

Sample a batch image x from training data;

Generate random masks m for x;

Update inpainting network G;

Counted $l_1$loss and two adversarial critic losses;

end while

4 Classification Model

We used the ImageNet dataset Stanford Vision Lab to evaluate our inpainting model. This data set contains more than 30,000 manually labeled training images, which are divided into plants, geological structures, natural objects, sports, artificial objects, fungi, human states, and animals. There are at least a hundred pictures in each category. Some categories contain more than 6,000 images. We divided it into training set, test set, and validation set in the ratio of 50:9:1. Firstly, our model learned on the training set and then tested the results on the validation set.

We have cropped and scaled the image. It is guaranteed that the size of each image input is 256 × 256. We used a part of the real image to test it. In the process of training, we used Adam as the optimization algorithm.

We set the learning rate as 5×10^{-4} and set as 0.5. We set the batch size to 2. We stop training after 40 epochs. We spent a total of nearly a week on the NVIDIA GPU.

4.1 Qualitative Comparison

For the case where the middle rectangle of the image is missing, the CA [16] solution produces distorted structural and color problems. SH performs well at this point, but it is blurred and lacks some details [17]. We put the results in the figure below, we can see that our model can effectively guess the content lost by the image.

Our model also excels when we want to remove the obstacles in the image (uncentered). We randomly selected 300 images from the ImageNet test data set and generated a regular rectangular mask for each image. Then we ran different methods on the damaged image to get the final result. During training, we used a maximum hole size of 128 × 128 with a resolution of 256 × 256. Both methods are based on a fully convolutional neural network. All reported results are direct outputs of the trained model without any post-processing. We used the common evaluation metrics (i.e., mean l_1 loss, PSNR and TV loss) (calculated using the complete image in the pixel space and the ground real image) to quantify the performance of the model. Tab. 1 shows the results of the evaluation. In the deep learning-based approach, our model is superior to all other methods in two indicators. The result can be explained by the fact that the existing method only considers the texture of the completed image to be realistic while ignoring the structure of the image. Figs. 4 and 5 show the results of our experiment.

Table 1: Comparison of PSNR, TV loss and mean $\boldsymbol{l_1}$ loss on ImageNet dataset

Decimal	Mean l_1 loss (%)	PSNR	TV loss (%)
CA	13.57	19.22	19.55
PICN [17]	12.91	20.10	12.18
Our method	8.2	19.32	25.6

Figure 4: Some image inpainting examples generated by our network model have achieved good effect visually in practice

Figure 5: When we use a mask to cover some pixels with isolated features, we can use the fake image generated by the network model to remove some objects like obstacles in the image (uncentered)

4.2 Ablation Study

Loss Function We tried to replace the perceptual loss with only l_2 loss. But in the subsequent experiments, we found that this method would cause that the inpainting results tend to be blurred. At the same time, we also tried to extract the features of VGG19 using different activation layers. The results show that the results of ELU are better than ReLU.

Reconstruction Loss We conducted the following training: experiments were not performed using the reconstructed l_1 loss during training. Our conclusion is that this will produce some strangely repaired edges. This tends to make the edges of the repair results repeat. They generate artifacts which make the images unrealistic. Although this is an essential component of image restoration. But this does not improve the semantics of the image. Therefore, we should use l_1 loss in training.

With and Without Regularization We train a complete model on the part of image-net Stanford Vision Lab without Regularization. Experimental results show that this will result in a lack of structured texture. Some examples of failure cases are, when the animal image is partially obscured, the model will prioritize placing the background in the foreground. For example, when the building is partially covered, the model places the sky and the ground in the missing place, which leads to our failure.

5 Conclusion

We propose a new method based on WGAN for image inpainting and implement it. The experimental results show that using the WGAN method to learn image features can significantly improve image

inpainting and make the results more semantic. As future work, we intend to expand the use of other GAN methods and experiment with image restoration for higher resolution images. We plan to apply it to some areas where we need to quickly fix images. Such as real-time processing of photographic photos. We will also explore the effect of the location of the missing image on the image restoration.

Funding Statement: This work was supported by the National Natural Science Foundation of China (Grant No. 42075007), the Open Project of Provincial Key Laboratory for Computer Information Processing Technology under Grant KJS1935, Soochow University and the Priority Academic Program Development of Jiangsu Higher Education Institutions, Graduate Scientific Research Innovation Program of Jiangsu Province under Grant no. KYCX21_1015.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] W. Fang, F. Zhang and V. S. Sheng, "A method for improving CNN-based image recognition using DC-gan," *Computers, Materials & Continua*, vol. 57, no. 1, pp. 167–178, 2018.

[2] W. Fang, Y. Ding, F. Zhang and V. S. Sheng, "Gesture recognition based on cnn and dcgan for calculation and text output," *IEEE Access*, vol. 7, pp. 28230–28237, 2019.

[3] C. Li and M. Wand, "Combining markov random fields and convolutional neural networks for image synthesis," in *Proc. CVPR*, Las Vegas, USA, pp. 2479–2486, 2016.

[4] M. Arjovsky, S. Chintala and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. ICML*, Sydney, Australia, pp. 214–223, 2017.

[5] Stanford Vision Lab PU, Stanford University, ImageNet, 2016. [Online]. Available: http://www.image-net.org.

[6] R. A. Yeh, C. Chen, T. Yian, A. G. Schwing, M. Hasegawa *et al.,* "Semantic image inpainting with deep generative models," in *Proc. CVPR*, Havaii, USA, pp. 5485–5493, 2017.

[7] W. Fang, F. Zhang, Y. Ding and V. S. Sheng, "A new sequential image prediction method based on lstm and dcgan," *Computers, Materials & Continua*, vol. 64, no. 1, pp. 217–231, 2020.

[8] L. Pan, J. Qin, H. Chen, X. Xiang, C. Li *et al.,* "Image augmentation-based food recognition with convolutional neural networks," *Computers, Materials & Continua*, vol. 59, no. 1, pp. 297–313, 2019.

[9] W. Fang, Y. Ding, F. Zhang and V. S. Sheng, "Dog: A new background removal for object recognition from images," *Neurocomputing*, vol. 361, no. 9, pp. 85–91, 2019.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. WardeFarley *et al.,* "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, no. 1, pp. 2672–2680, 2014.

[11] R. Gao and K. Grauman, "On-demand learning for deep image restoration," in *Proc. CVPR*, Havaii, USA, pp. 2223–2232, 2017.

[12] G. Liu, F. A. Reda and K. J. Shi, "Image inpainting for irregular holes using partial convolutions," in *Proc. ECCV*, Munich, Germany, pp. 85–100, 2018.

[13] Y. Tu, Y. Lin, J. Wang and J. U. Kim, "Semi-supervised learning with generative adversarial networks on digital signal modulation classification," *Computers, Materials & Continua*, vol. 55, no. 2, pp. 243–254, 2018.

[14] S. Wang, J. He, C. Wang and X. Li, "The definition and numerical method of final value problem and arbitrary value problem," *Computer Systems Science and Engineering*, vol. 33, no. 5, pp. 379–387, 2018.

[15] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu *et al.,* "Generative image inpainting with contextual attention," in *Proc. CVPR*, Salt Lake City, USA, pp. 5505–5514, 2018.

[16] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. C. Courville, "Improved training of wasserstein gans," *Advances in Neural Information Processing Systems*, vol. 28, no. 5, pp. 5767–5777, 2017.

[17] X. Li, Y. Liang, M. Zhao, C. Wang and Y. Jiang, "Few-shot learning with generative adversarial networks based on WOA13 data," *Computers, Materials & Continua*, vol. 60, no. 3, pp. 1073–1085, 2018.