Image Semantic Inpainting with DCGAN GIt: https://github.com/ansarrice/576final

Ainiwaer Ansaerjiang[1], Han Jiang[2], Lulu Song[3], and Weihong Xie[4]

[1]**aa171@rice.edu, aa171, Rice University, Computer Science**
[2]**hj32@rice.edu, hj32, Rice University, Computer Science**
[3]**ls116@rice.edu,ls116, University of Texas Health Science Center at Houston, Biostatistics**
[4]**wx19@rice.edu, wx19, Rice University, Computer Science**

ABSTRACT

In recent years, the increasing demand for computer vision and the development of deep learning algorithms such as neural networks, image detection, image segmentation, and image restoration technologies based on deep learning methods have been widely used in the medical, military, and film industries. Image-inpainting refers to using the missing part of the image's neighbor information or the overall information of the image to generate missing areas or reduce image noise according to certain repair rules. So far, various image restoration algorithms have been proposed for various image-inpainting problems. Although these algorithms can repair the image to a certain extent, there are also problems such as loss of complete information of the original image, and the repair effect is not ideal. In this final project, image inpainting algorithms are dived deeper by building deep convolutional generative adversarial networks (DCGANs) algorithms.

In this project, DCGANs is optimized to make the repair effect more ideal. The idea of using DCGANs for image inpainting is somehow different. First, input the defect image and repeat training the generator, which is used to generate the best fake picture corresponding to the defect image. At the same time, train the discriminator by using the original image corresponding to the defect image. The discriminator in DCGAN is a classifier, which can be utilized to classify the restored images. With the discriminator loss, the restoring results is evaluated, the discriminator is penalized, and the discriminator's weights are updated continuously through backpropagation. In this way, the model generates the best-generated image and fills the missing part in the extraction location to achieve the purpose of restoration.

Keywords: Deep-learning, Image-inpainting, DCGAN

BACKGROUND AND MOTIVATION

With the penetration of electronic devices into people's lives, more and more information is transmitted and stored in the form of digital images. However, the acquired images can lose some information due to equipment failure, environmental interference, human factors, and storage container failure. This seriously affects the information contained in the images. Image restoration is the act of using the implicit information in an image to improve the viewing experience by adding missing parts, blurred parts, and other insufficient information in a way that is intuitive to the observer. The implicit information in the image mostly refers to the specific information obtained from the surrounding area to be repaired and the big picture information obtained from the overall perspective of the image. The ultimate purpose of this is to make the image that is visually better after restoration unnoticeable to the observer. Therefore, image restoration is an extremely important part of the image field, and it is necessary to restore images that are not clear enough or are defective to a certain extent.

Before the performance of the tools could meet and support the calculation of huge data, the traditional method of image restoration was only based on mathematics and human intuition. When there are enough equipments to support deep learning, more and more outstanding and excellent related technologies emerge and are applied in lots of areas including the image field. Because the implementation effect of deep learning far exceeds that of traditional methods, and the application cost is extremely considerable compared with traditional methods, the choice of deep learning has become the first choice in the field of vision. Actually, it becomes a trend to apply deep learning to the image field.

Since the convolutional neural network is applied to neural network-related technologies, it can be feasible and effective to find the best combination of network parameters to achieve better results according to the change of the corresponding network loss, which is mentioned by O'Shea and Nash (2015). Creswell et al. (2018) has mentioned that the realization of Generative Adversarial Network relies on a sufficiently powerful computing system and effective and clear evaluation conditions. Therefore, the network can gradually adjust its parameters according to the results, and the results are getting closer and closer to the target, which is an unsupervised learning process.

In the related field of deep learning, a fully convolutional image restoration network was proposed that is capable of repairing large area occlusions in images, and details has been discussed by Iizuka et al. (2017). However, it is difficult to extract effective features, which results in unnatural texture structure of the restored image. Li et al. (2017) proposed a deep generative network-based image restoration model that generates more consistent image semantic images. The generative network was used to synthesize the most probable defective parts to achieve the overall repair. However, the restoration is vague and has semantic incorrectness since it does not utilize the overall image information. Pathak et al. (2016) proposed deep convolutional adversarial generative networks combining the strengths of both CNN network and GAN. Specifically, a deep convolutional adversarial generative network (DCGAN) is a combination of a CNN network with supervised learning and a GAN with unsupervised learning. Combining the strengths of both, better restoration can be achieved with reasonable network construction and loss function.

This project focuses on the application of DCGAN to image restoration, discuss the issues that should be considered in the process of application, and the specific steps for implementation. Then experiments are performed. And the network and the restoration structure are improved based on the results. Work includes introduction the advantages and disadvantages of DCGAN, explaining the principles of its image restoration, building and training the DCGAN network for image restoration, introduction the reasons for the selection of steps in the restoration process, and evaluating the results of the construction of the model in terms of the final restoration results and the model parameters curve.

DEEP LEARNING APPLIED TO IMAGE INPAINTING

The inpainting process will be done in three steps as follows.

1. Interpreting the image as a sample in a probability distribution.
2. Build a neural network based on the interpretation step.
3. Make the network generate the most fake image based on the information of the original image

Interpreting images as probability distributions

In the computer's view, images and characters and even audio. There is no difference in their storage contents, they are all different arrangements of numbers in their corresponding storage space. Therefore, it does not matter to the computer whether the images are intact or broken. They all look like numbers. So suppose a bit of storage space is determined, what is the value of the number y at this location? It is naturally straightforward to know for a determined set of rows of images. When uncertain, one can assume all possible values of y for this place and evaluate that most likely value according to the individual results, i.e., get the probability of each possible y. It is only necessary to convert the generated restoration image into, the problem of finding the maximum probability of all possible missing values, then the most probable of the results found is the best generated image. This is shown in Figure 1.

Image Inpainting with Convolutional Neural Networks

Convolutional neural networks now excel in deep learning problems about image classification on Li et al. (2017). The main principle of convolutional neural network is to extract features, all the features that are determined to be a class of things, to form a model. When using the model to determine a new image, the new image is looked at to see if there are any features that have been identified. The more features that match the model, the more the model is judged to be the same thing. If there are fewer matches, a small value is given, which means that it is very unlikely to be the same type of thing. Which of the extracted features is the more important and effective one? For example, dog teeth is a more useful feature than the color of the dog's fur in determining whether the object is a dog or not. Then each feature is given a feature weight. The higher the weight, the more important the feature is. As for how to determine which

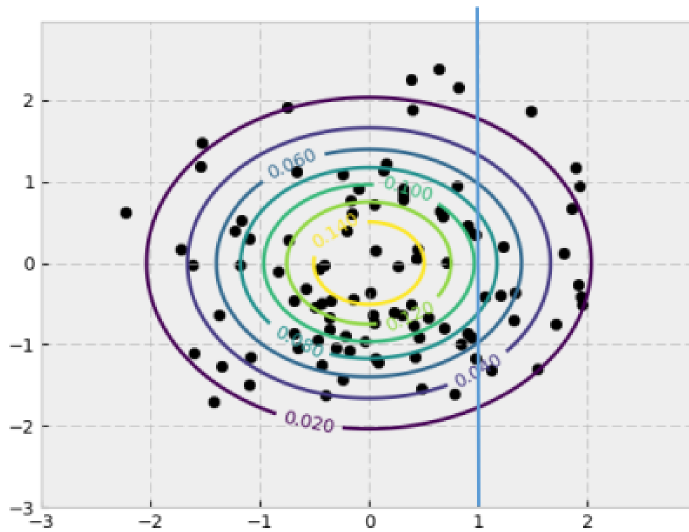


Figure 1. Probability distribution of images.

feature is more important, naturally, to try it. With a large enough data training, when using the feature to make his weight greater to determine the return results more accurate let him a little larger, and vice versa. And in order to facilitate the manipulation of weights, they are gathered together as a new data type, called weight set.

The weight set is actually a filter that is used to eliminate the interference data and determine the core data. The more features that match the filter, the more likely it is that the thing corresponding to the image matrix and the thing corresponding to the filter are one and the same. This operation is the convolution operation, as shown in Figure 2.

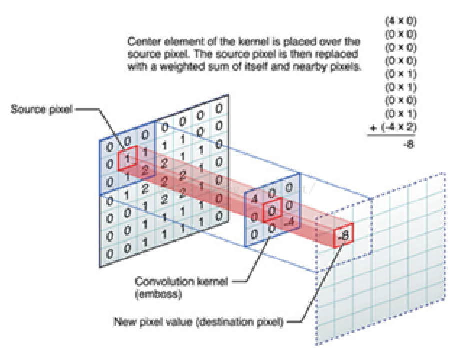


Figure 2. Convolution operation

The outermost circle of zeros of the matrix is complemented to facilitate the corresponding convolutional mapping of the elements in the corners and edges. Transpose convolution, the operation is similar to the inverse process of convolution, but the content is different. Transpose convolution needs to fill the output with empty elements to continuously expand the output in order to output a larger result. As shown in Figure 3.

Convolutional networks are applied to image restoration mainly by inputting defective images and then using contextual information for convolutional self-encoder training and finally outputting a restoration map for the broken part, except that the images generated in this way are generally blurred because looking at the continuity only does not give accurate information. For this reason, local adversarial loss is added to determine whether the repair map is from the original image or from the generated image of the generator, and the more accurate the result is, the smaller the local adversarial loss is. This way the generator generates a repair image that is closer and closer to the real image. However, this does not

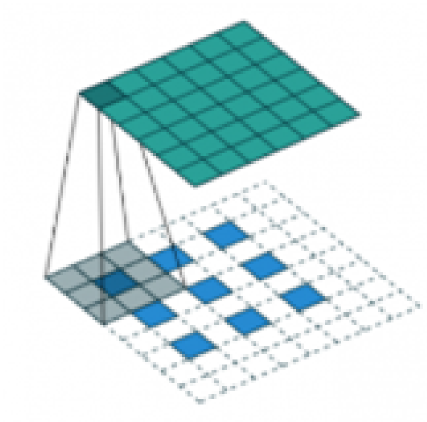


Figure 3. Transpose convolution 3*3 filter to continuously expand the input

guarantee the consistency of the generated repair area with the original image, e.g., the stitching edges look abrupt and not like a normal image. Therefore, a global confrontation loss is added to discriminate the authenticity of the repaired image, to determine whether it is the stitched image or the original image, and if it is certain that it is not the original image, then the generator need some major adjustments.

Image Inpainting with GAN

GAN is Generative Adversarial Network having two neural networks: Generator and Discriminator that are pitted against each other and are simultaneously trained by an adversarial process.Yeh et al. (2017) The principle of Generative Adversarial Network is that two networks optimize themselves by learning from each other with opposite purposes, i.e., the two networks form a competitive relationship. One network is a generator network G with the input image to be repaired and the output is the repaired image, and the other network is a discriminator network D with the input image being a pair of images and the output being the probability that it could be the original image. The output will be close to 1 if the original image is determined, and close to 0 if the generated image is determined. Ideally, the equilibrium is eventually reached where the discriminator cannot determine whether it is the fake or the original image.Zhang et al. (2016) It is easy to think of the question why the goal is not for the judge to discern that the input image is the original image? Because if the judge thinks that the generated fake image has a high probability to be the original image. Inevitably, it proves that the judge itself is not accurate enough, and the fake image that is considered to be the original image may not work well, and the judgment result will not be credible.Figure 4. shows the principle diagram of adversarial generation of GAN network.

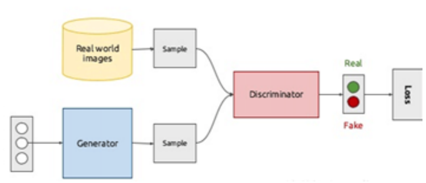


Figure 4. GAN network to generate images

To perform image-inpainting, the input is a high-dimensional feature vector that is tensed according to the broken image. The first step is to use the incomplete image and the corresponding complete original image input to the generative and discriminative networks, and train both the generative and discriminative networks based on the feedback. In this process, an appropriate repair error penalty is set so that the generator has the ability to generate fake images close to the original image, while the discriminator has the ability to separate the original image from the generated image. Ideally, the generated image directly output by the trained model is consistent with the original image to be repaired, and a simple cropping and stitching is performed to the original image.

Image Inpainting with DCGAN

The principle of DCGAN is to implement the generator and the discriminator of GAN network with convolutional neural network. It is a combination of two networks.Xia et al. (2019) For the experiments, considering the model training time, output effect, etc. Some modifications to the structure of the convolutional network need to be implemented.

(1) The pooling layer is eliminated, and the pooling layer eliminated in the judge network is replaced with a convolutional layer, and the pooling layer eliminated in the generator network is replaced with a transposed convolutional layer. Because the pooling layer will lose a lot of valuable information while, generating image information is required a lot of information.

(2) Use batch normalization for each layer after convolution before activation to prevent overfitting due to oversized parameters.

(3) Delete the dense layer and make the network a full convolutional network. This can accelerate the training speed.

(4) Delete the flatten layer and use vector blocks as the extracted high-dimensional feature vectors. The main goal is to reduce the computational effort.

(5) The relu function is used for activation except for the output layer when training the judge network. The sigmoid function is used for the output layer to ensure that the output is two outputs.

(6) The modified ReLU function, LeakyReLU function, is used for activation of the generator network training except for the output layer. The output layer is activated using tanh activation in order to ensure that the output is between -1 and 1.

(7) Considering the application, to fix the specific image like having to add the influence of the specific image, the input of the generated model is changed from noise to the broken original image from the beginning of building the model, and the broken image and the original image are used as the training data pair to train the model.

The DCGAN network is applied to image-inpainting by referring to the construction process of CNN networks and GAN networks, where the first input is a high-dimensional feature vector based on the broken image using the decoder part of the generator tensor. The broken incomplete image and the corresponding complete original image are input to the generator and discriminator as a pair of data pairs, and the outputs of the two networks are paired to train both networks simultaneously. Based on the feedback, the generative and discriminative networks are also adjusted simultaneously. In this process, a suitable loss penalty is set so that the generator refines the ability to generate fake images close to the original image, while the discriminator refines the ability to distinguish the original image from the generated image. Ideally, the trained two networks, for the generative network the probability distribution of the input broken image and then the output repaired image is exactly the same as the original image. A simple cropping and stitching is performed to get to the original image. This part of the principle is mostly similar to the restoration principle of GAN networks.

Comparison of the three networks

It is proved that CNN, GAN, and DCGAN can be applied to image restoration, but the advantages and disadvantages of the three networks are different, and the comprehensive Table 1 shows the comparison of the advantages and disadvantages of the three networks. According to the table, DCGAN network is selected for image restoration in the following experiments because it has a better theoretical basis.

CONSTRUCTION OF DCGAN MODEL

Dataset Related

Before deciding to use the CelebA dataset,Liu et al. (2015) building the dataset by using crawlers to crawl icons by themselves can work, but the restoration effect was poor because the collected images tended to be obscure, confusing and few in number. In order to improve the restoration effect of DCGAN network, only one specific class of images is selected for restoration, so in this paper, CelebA image set is used for restoration training. And the images will be pre-processed.

(1) A face-centered crop (64*64*3) is performed on the Celeba image set.

(2) Normalize each image in the range [-1,1] because the pixel values are in [0,255], so divide each pixel by 255 multiply by 2 and subtract 1.

(3) Break up the image, using tensorflow's shuffle method

Method	Advantages	Disadvantages
CNN	Any broken area of the image can be repaired	Only low-resolution images can be generated,and the restored images have poor results
GAN	Can generate images with high resolution restoration	GAN is more representative of the original image than the features learned by CNN It is not easy to converge during training; training is not stable, resulting in the unsatisfactory generation of the generator; the optimization of G generation comes from the feedback of the resolution result of D. If the resolution result of D is too good, it will cause the gradient of G to disappear; if G generates fake images too well, it will cause D to be unable to distinguish between real and fake images, resulting in overfitting when G is stuck
DCGAN	Combine the advantages of CNN and GAN to reduce the instability of GAN network training	The instability during training is reduced, but it is still unstable

Table 1. Comparison of CNN, GAN, DCGAN on image-inpainting

(4) Combine the images into a batch, the value can be freely configured according to the video memory.

Construction of Generators and Discriminators

Images generated by G are used to fill the missing regions, the D diagnose the image whether or not it conform to realistic image when the inpainted image as input images. Here, content loss, gradient loss and a prior loss are reverse update. The DCGAN architecture is shown in Figure 5

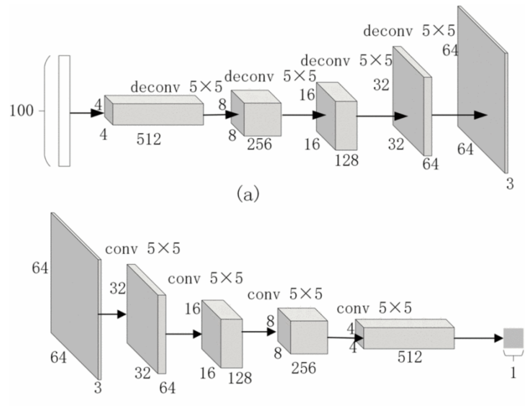


Figure 5. Deep convolutional generative adversarial net-works (DCGAN) architecture. (a) the generator network; (b) the discriminator network.h

Structure of the DCGAN

The confrontation between the generator and the judge during the training of DCGAN network is shown in Figure 6

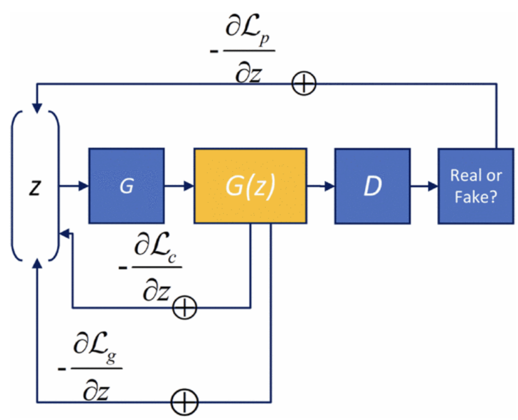


Figure 6. The inpainting process of our approach

The input is a high-dimensional feature vector tensed from the broken image. The first step is to use the broken incomplete image and the corresponding complete original image input to the generative and discriminative networks. The generator takes an corrupt image of size (64, 64, 3) as input as and generates the corresponding best fake image according to the corrupt one. The output of the generator, which is the best fake image, is used as the input of the discriminator along with the original image. The best fake one and original image are classified as a data pair by the discriminator. If the discriminator classifies real images as real and generated images as fake, then the generator's penalty feedback can be large. Otherwise, the discriminator's penalty feedback will be adjusted accordingly according to the degree of deviation.Ren et al. (2018) Train both the generative and discriminative networks based on the feedback.

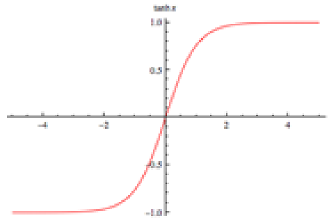
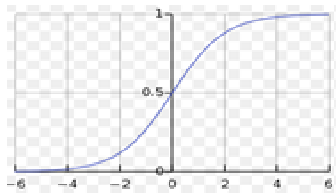
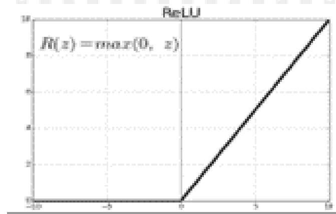
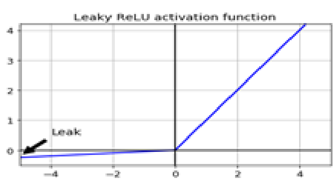
Method	Function Image	Implementation
Tanh		After the tanh function, the output range is -1 to 1. For layers where the output needs to be -1 to 1
Sigmoid		sigmoid function when the output is 0 and 1 when the curve tends to flatten, the output will be mostly 0 or 1. Suitable for two classification problems
Relu		The Relu function converges faster than the sigmoid, but as can be seen from the graph, the output value is positive
LeakyReLU		A modified version of the Relu function with a very small negative area increment added. Can be applied to the problem of input containing negative numbers

Table 2. Activation function of DCGAN network

Activation Function

Table 2 shows the relevant activation functions required for this DCGAN network model construction. Figure 7 shows that the structure of the DCGAN network with different activation functions at each layer.

Loss Function

Discriminator loss

While the discriminator is trained, it classifies both the real data and the fake data from the generator.

It penalizes itself for misclassifying a real instance as fake, or a fake instance (created by the generator) as real, by maximizing the below function.

$$\nabla_{ed} \frac{1}{m} i - 1m[logD(G(x)) + log(1-D(x))]$$

- log(D(x)) refers to the probability that the generator is rightly classifying the real image.
- maximizing log(1-D(G(z))) would help it to correctly label the fake image that comes from the generator.

Generator loss

While the generator is trained, it samples random noise and produces an output from that noise. The output then goes through the discriminator and gets classified as either “Real” or “Fake” based on the ability of the discriminator to tell one from the other.

The generator loss is then calculated from the discriminator’s classification – it gets rewarded if it successfully fools the discriminator, and gets penalized otherwise.

The following equation is minimized to training the generator:

$$\nabla_{ed} \frac{1}{m} \sum_{i-1}^{m} log(1-D(G(x)))$$

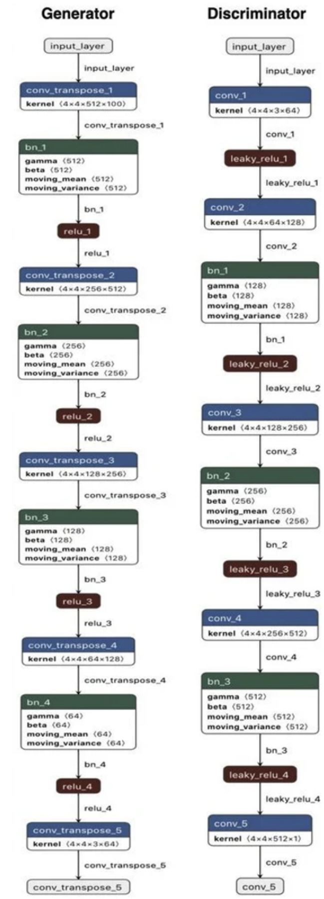


Figure 7. Generator and Discriminator network architectures

Training Process

1. Processing image dataset.

a) Unify the image format by face-centered, cut to reduce irrelevant features (clothing, etc.), reduce the computation and mention the extraction effect.

b) Normalize the images so that the value of each pixel point of the image dataset is in [-1,1] to reduce the computation.

c) Perform random disruption and divide into multiple patches to train in patches, the size of the batch does not affect the result but affects the training speed.

2. Build the generator.

a) Input is a broken image of (64, 64, 3) and output is a repaired image of (64, 64, 3).

b) Determine the number of layers of transposed convolution layers and the size of the convolution kernel for each layer.

c) Determine the uniform normalization and activation function selection after transposed convolution.

3. Construct the determiners.

a) The input is a broken image of (64, 64, 3) and the output is a numerical true image probability.

b) Determine the number of layers of convolution and the size of convolution kernel for each layer.

c) Determine the uniform normalization and activation function selection to be used after convolution. Construct the optimizer.

d) Determine the learning rate and associated values of the Adam optimizer

e) Construct the cost function as in Table 3.1 using the cross-entropy cost function.

4. Train the model using the dataset.

5. Call the trained model to perform a repair.

6. Optimize the model based on the repair.

Figure 8 shows that whole process of how DCGAN model does image-inpainting.

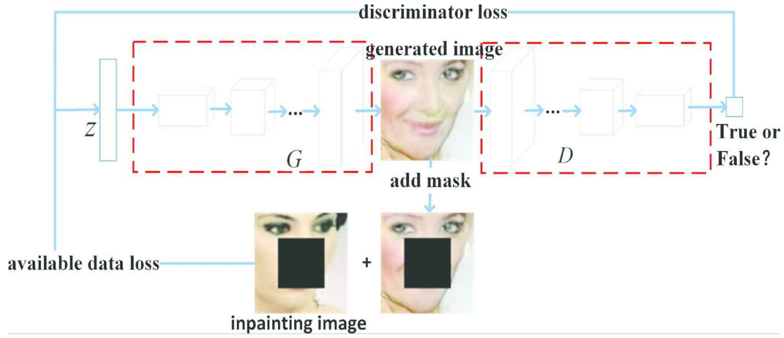


Figure 8. Whole process of DCGAN

RESULTS AND ANALYSIS

For the training of the image inpainting model, 10,000 pictures in the CelebA database were selected as the training set. The results are shown below. Use a more high-performance configuration by using a cloud server with your own code deployed to run on a remote server, such as Google ML platform or AWS. This training is completed by a local Nvidia 3070 GPU.

As shown in the Figure 9, the generation loss and discriminator loss analysis graph. From the images, it can be seen that the losses of G and D networks are relatively evenly distributed within a certain range of transformations, which proves that G and D networks are effectively fighting against each other and the generation results are continuously optimized. The trend of both sides is smooth. It indicates that the generating ability and judgment accuracy of the networks are improving.

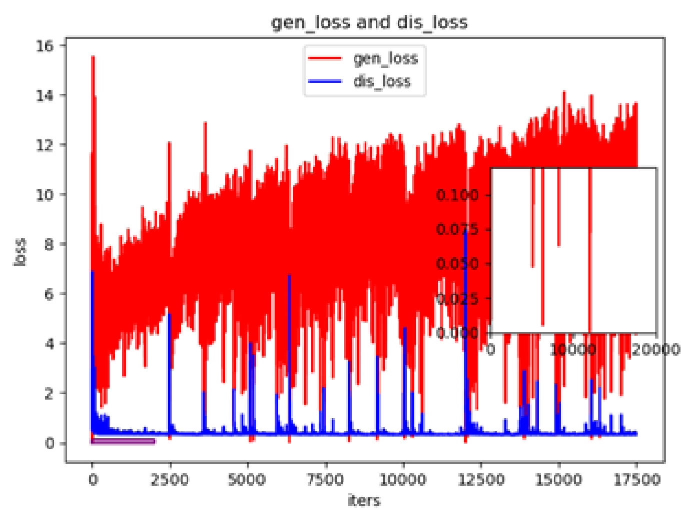


Figure 9. Analysis of Genloss and Disloss

Figure 10 shows that the image inpainting result, In each example, the left and middle images are the input corrupted image and the output repaired image, respectively. The picture on the right is the actual real picture for comparison.

After a period of model training, the repair effect of the network has been particularly impressive. For most of the clear frontal pictures, the fake pictures directly generated have reached the point where people can't realize the difference without looking carefully. However, there are exceptions, as shown in the last group of pictures of white-haired women in Figure 9. If look carefully at the generated pictures, the traces of splicing can still be noticed.

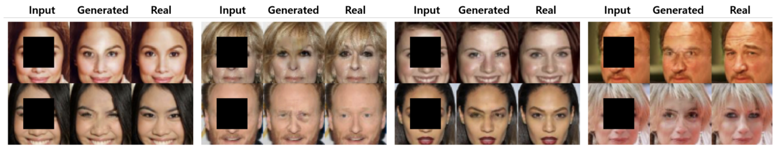


Figure 10. Semantic image inpainting result

Most of the problems are due to the fact that the trend of the training set is not obvious enough and the number is not large enough. There is still room for improvement in the construction of the data set in this experiment. If permitted, increasing the depth of the model can also improve the repairing ffect.

CONCLUSION AND DISCUSSION

The report introduces the DCGAN network developed from the CNN network and the GAN network, combined with the principles of image restoration from scratch in the main body of this article, and implements these principles step by step. Finally, training, experiments, and optimization adjustments were carried out according to the built DCGAN network. After a structural rework, the DCAGN network using a new construction idea achieved a relatively ideal restoration effect.

The final image restoration result of the project is not bad, but the model structure in the experiment can also be improved. But the optimization of the network itself might not play a role in qualitative change. The cognition of the whole network is based on the training set and limited to the training set. It is experience that guides restoration as it proceeds. However, in reality, what really guides us to perform image restoration is information. People know what a person looks like, and what the distribution of muscles and bones in the face looks like. These are the basis of artificial restoration, and these basis come from the interpretation results of other neural network models. Therefore, the true aim of the improvement should focus on the combination of different neural network models.

REFERENCES

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE signal processing magazine*, 35(1):53–65.

Iizuka, S., Simo-Serra, E., and Ishikawa, H. (2017). Globally and locally consistent image completion. volume 36, pages 1–14. ACM New York, NY, USA.

Li, Y., Liu, S., Yang, J., and Yang, M.-H. (2017). Generative face completion. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3911–3919.

Liu, Z., Luo, P., Wang, X., and Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738.

O'Shea, K. and Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., and Efros, A. A. (2016). Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544.

Ren, K., Meng, L., Fan, C., and Wang, P. (2018). Least squares dcgan based semantic image inpainting. In *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, pages 890–894. IEEE.

Xia, L.-m., Wang, H., and Guo, W.-t. (2019). Gait recognition based on wasserstein generating adversarial image inpainting network. *Journal of Central South University*, 26(10):2759–2770.

Yeh, R. A., Chen, C., Yian Lim, T., Schwing, A. G., Hasegawa-Johnson, M., and Do, M. N. (2017). Semantic image inpainting with deep generative models. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5485–5493.

Zhang, R., Isola, P., and Efros, A. A. (2016). Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer.