Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/taut20

The image inpainting algorithm used on multi-scale generative adversarial networks and neighbourhood

Jiangchun Mo & Yucai Zhou

To cite this article: Jiangchun Mo & Yucai Zhou (2020) The image inpainting algorithm used on multi-scale generative adversarial networks and neighbourhood, Automatika, 61:4, 704-713, DOI: 10.1080/00051144.2020.1821535

To link to this article: https://doi.org/10.1080/00051144.2020.1821535

Published online: 23 Sep 2020.

AUTOMATIKA
2020, VOL. 61, NO. 4, 704–713
https://doi.org/10.1080/00051144.2020.1821535

REGULAR PAPER

OPEN ACCESS Check for updates

The image inpainting algorithm used on multi-scale generative adversarial networks and neighbourhood

Jiangchun Mo and Yucai Zhou

School of Energy and Power, Changsha University of Science and Technology, Changsha, People's Republic of China

ABSTRACT

Various problems existed in the image inpainting algorithms, which can't meet people's requirements visually. Aiming at the defects of the existing image inpainting algorithms, such as low accuracy, poor visual consistency, and unstable training, an improved image inpainting algorithm used on a multi-scale generative adversarial network (GAN) and neighbourhood model have been proposed in the paper. The proposed algorithm mainly improves the network structure of the discriminator, and introduces a multi-scale discriminator based on the global discriminator and the local discriminator. The multi-scale discriminators were trained on images of different resolutions. Discriminators of different scales have different receptive fields, which can guide the generator to generate more global image views and finer details. Aiming at the problem of gradient disappearance or gradient explosion that often occurs in GAN training, the method of WGAN (Wasserstein GAN) has been used to simulate the sample data distribution using EM distance. The proposed model has been trained and tested on the CelebA, ImageNet, and Place2. The experimental results show that compared with the previous algorithm model, the proposed algorithm improves the accuracy of image inpainting and can generate more realistic repairing images, and it is suitable for many types of images.

ARTICLE HISTORY
Received 7 March 2020
Accepted 1 September 2020

KEYWORDS
Image inpainting; generative adversarial networks; multi-scale; reconstruction loss; adversarial loss

The following variables are matrices:
Repaired Image (Original Image): I
The Noise Mask: z
The Distribution of Real Data x: $P_{data}(x)$
The Distribution of Noise Variable z: $P_z(z)$
Others are common variables.

1. Introduction

With the rapid development of deep learning in the researching field of computer vision, the research works on image editing and image generation have achieved remarkable results. The image inpainting problem, discussed in the paper, is a hot issue between image editing and image generation. It has important applications in the areas of image scaling, protection of cultural relics, facial repair of police detectives, biomedical image applications, and aerospace technology.

Image inpainting is a traditional graphic problem. A certain area is missing at a certain position on an image, and other information is used to restore this missing area, making it impossible for people to identify the repaired part. As shown in Figure 1 (from left to right, the original image, the missing image, and the repaired image), the missing areas in the two images have cups and flowers, respectively. Persons can easily convert the images according to the content of the surrounding image inpainting. Because the human brain has subjective consciousness, different people have different repair effects. Therefore, in the process of image repairing, the principles of structure, similarity, consistent texture, and structure priority must be followed. However, the image repairing task is particularly difficult for computers, because there is no unique solution to this problem. How to use other information to assist the repair and how to determine whether the repair results were sufficiently authentic are the concerns of researchers.

At present, the structure-based image inpainting [1–4], texture-based image inpainting [5–10], and deep learning-based image inpainting [11–16] are the three main directions in the research field of image inpainting. The research in the paper is mainly aimed at image learning algorithms based on deep learning. In recent years, convolutional neural networks (CNN) [17,18] have greatly improved the performance of semantic image classification [19–22], object detection [23–27], and image segmentation tasks [28,29]. Researchers have used CNN models for image inpainting tasks, but the image inpainting methods using only CNNs

CONTACT Jiangchun Mo jiangchunmo@aliyun.com School of Energy and Power, Changsha University of Science and Technology, Changsha 410076, People's Republic of China

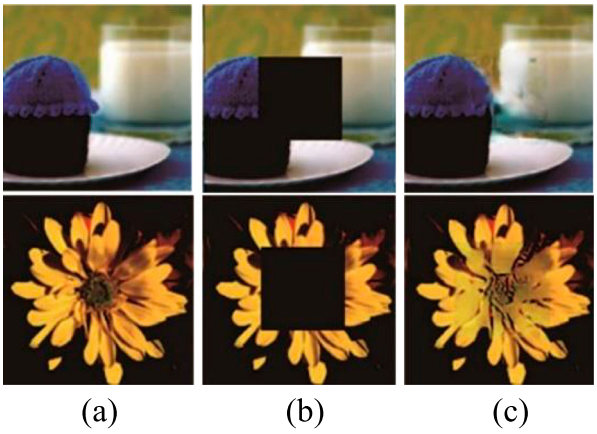

Figure 1. Repair of three different images (a) original image (b) missing image (c) repaired image.

have low accuracy and great room for improvement in performance.

Aiming at the problems of the existing methods, the paper has proposed a multi-scale generative adversarial network and neighbourhood model to obtain high-precision, high-accuracy, and visually consistent inpainting images. Firstly, the generator and discriminator were composed deep generation adversarial inpainting model uses the reconstruction loss and the adversarial loss to synthesize the missing content from random noise. Secondly, a multi-scale discriminator structure has been proposed, and image inpainting is performed by using adversarial training with images of different resolutions. Then, the generator uses dilated convolution to reduce the loss information during down-sampling, and uses current Poisson blending method to perform certain subsequent processing on the repaired image. Finally, the advantages of the proposed method and the effect of image inpainting are illustrated through experiments.

2. The related works

Traditional image inpainting methods, such as the literature [30], have used the diffusion equation to iteratively propagate to the unknown region along the low-level features of the known region along the mask boundary. Although it performs well in inpainting, it is limited to dealing with small and uniform areas. By introducing texture synthesis [31], the inpainting effect is further improved. The literature [32] recovered the image with missing pixels by learning the prior of the image block.

Early, the deep learning-based image inpainting methods, such as the literature [33], have learned a convolutional network, which greatly improves the performance of image inpainting through an efficient image block matching algorithm [34]. When a similar image block is found, it performs well, but when the dataset does not contain enough data to fill the unknown area, it is likely to fail. The literature [35] has used image inpainting as a task to recover sparse signals from the input. By solving a sparse linear system, the image can be repaired based on some corrupted input image. However, the algorithm requires images to be highly structured. The literature [36] has proposed Variational Auto-Encoders (VAEs). By applying a priori on the latent units, the images can be generated by latent unit sampling or interpolation. However, due to training objects based on pixel-level Gaussian likelihood, the images generated by VAEs were usually blurred.

With the further development of deep learning [58–60] [37–39], the Generative Adversarial Networks (GAN) have been proposed by the literature [40]. The literature [40] is a milestone in the development of deep learning. With the advent of GAN, the problem of blurring of images generated using traditional VAEs was solved, and shocking results were achieved. In theory, a large number of clear images can be generated. The literature [41] has improved VAEs by adding an adversarial trained discriminator, which came from a generative adversarial network and proved that it can generate more realistic images.

One of the main problems in image inpainting using GANs is the instability during model training, such as the inability of the network to converge, prone to gradient disappearance, and gradient descent, which led to a lot of research on this problem [42]. The latest research shows that cross-entropy (JS divergence) in traditional GANs isn't suitable for measuring the distance between the generated data distribution and the real data distribution. If you train GANs by optimizing the JS divergence, you will not find the correct optimization object. The Wasserstein GAN (WGAN) [43] has been proposed by the literature [43] to improve the GAN from the perspective of the loss function. The improved WGAN after the loss function can obtain good performance results even on the fully connected layer, and solves the problem of unstable training. The literature [44] has improved on the basis of Wasserstein GAN [44], optimized the continuity constraints, solved the problem of training gradient disappearance and gradient explosion, and accelerated the convergence speed. The LSGAN (Least Squares GAN) [45] model has used the least squares loss function instead of the GAN loss function, which also alleviates the problems of unstable GAN training, poor image quality, and insufficient diversity.

As a person has higher requirements for the resolution of GAN-generated images, another problem that comes with it is that the network will down-sample the images during the pooling process to extract low-dimensional features, resulting in the loss of much key information in the images [46–48]. The discriminator is easier to distinguish real and fake images, so that the gradient can't indicate the correct optimization direction. How to effectively use the features extracted from

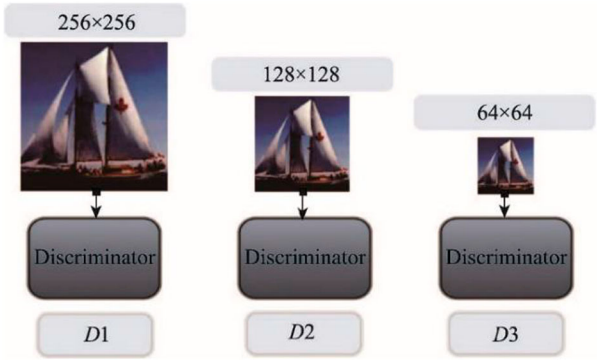

Figure 2. Multi-scale discriminator model.

each layer of the neural network, minimize the loss caused by the down-sampling process, and fully extract the low-dimensional features of the image, it is a hot spot in the current research. The literature [49] has proposed a method of expanding convolution, which can expand the receptive field and keep the size of the feature map unchanged during the convolution process, which effectively reduces the banding caused by down-sampling in the traditional convolution process. The incoming information is lost and used for image processing. The "pix2pixHD" model proposed by the literature [50] utilizes conditional generative adversarial networks (Conditional GANs) [51] to synthesize high-resolution realistic images, using a latest multi-scale generator-discriminator structure to stabilize. At the same time, the quality of the image has improved and the resolution of the image has improved. Figure 2 shows a schematic diagram of a multi-scale discriminator model. It has the same network structure, but works at different image scales. These discriminators are called D_1, D_2, and D_3. Specifically, the down-sampling of real and synthetic high-resolution images, respectively D_1, D_2, and D_3 are then trained to distinguish between real and synthetic images on three different scales.

By using the Mean Squared Error (MSE) loss combined with the GAN loss, an image inpainting network can be trained, avoiding the common ambiguity when only using the MSE loss. Merely using this method can make network training unstable. The paper has used the loss in WGAN to replace the loss of traditional GAN, uses the EM distance to measure the difference between data distributions, and does not train generative models and adjust the learning process to prioritize stability to avoid this problem. In addition, the architecture and training process were optimized for image inpainting problems. In particular, instead of using a single discriminator, multiple discriminators were used, and a multi-scale discriminator similar to the "pix2pixHD" model [50] is used to improve visual quality.

3. The proposed algorithm

Section 3 introduces the multi-scale generative adversarial networks model and neighbourhood, including a generative network for image inpainting, four additional discriminator networks-assisted training, namely two multi-scale discriminator networks, a global discriminator network, and a local discriminator network in order to train the entire network to perform image repair tasks outstandingly. During training, the discriminator is trained to determine if the image has been successfully repaired, while the generator is trained to fool all discriminators. Only through

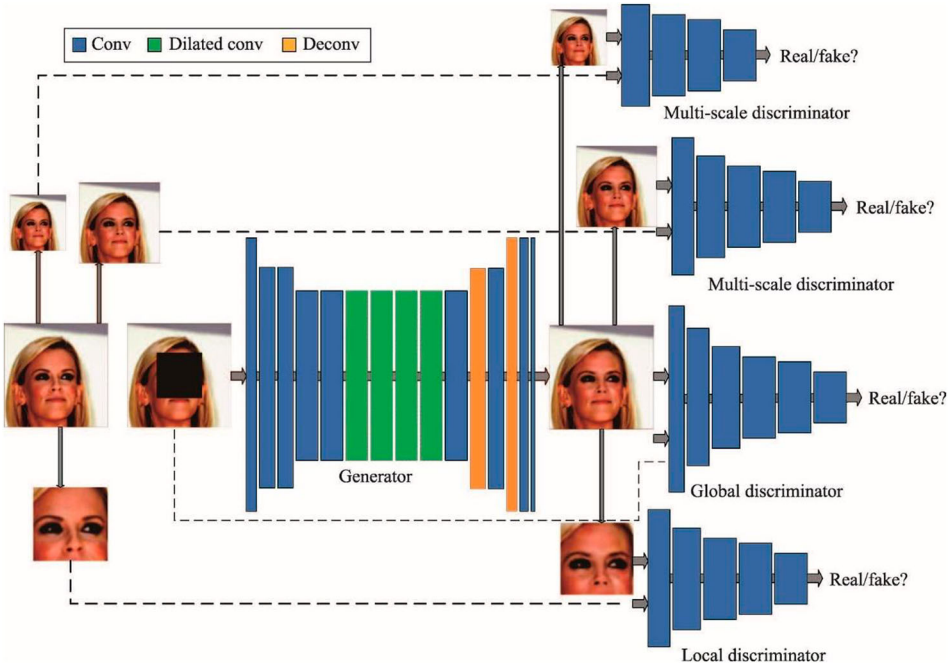

Figure 3. Network architecture.

Table 1. Architecture of generator *G*.

Layer	Kernel	Dilation	Stride	Outputs
conv	5 × 5	1	1 × 1	64
conv	3 × 3	1	2 × 2	128
conv	3 × 3	1	1 × 1	128
conv	3 × 3	1	2 × 2	256
conv	3 × 3	1	1 × 1	256
d-conv	3 × 3	2	1 × 1	256
d-conv	3 × 3	4	1 × 1	256
d-conv	3 × 3	8	1 × 1	256
d-conv	3 × 3	16	1 × 1	256
conv	3 × 3	1	1 × 1	256
deconv	4 × 4	1	1/2 × 1/2	128
conv	3 × 3	1	1 × 1	128
deconv	4 × 4	1	1/2 × 1/2	64
conv	3 × 3	1	1 × 1	32
deconv	3 × 3	1	1 × 1	3

all the networks trained together can the generator really repair various images. The network architecture is shown in Figure 3.

3.1. Generator

A convolutional self-encoder is used as the generator model *G*, which is a standard encoder–decoder structure. The encoder structure takes an image with missing regions as input content, and generates a latent feature representation of the image through a convolution operation. The decoder structure uses this latent feature representation to restore the original resolution through a transposed convolution operation, resulting in the image content of the missing area. Unlike the original GAN model, which starts directly from the noise vector, the hidden representation obtained from the encoder captures more changes and relationships between unknown regions and known regions, and then inputs the decoder to generate content. The intermediate layer uses dilated convolution, which allows a larger input area to be used to calculate each output pixel. There are no additional parameters and calculations. Compared with the standard convolution layer, the dilated convolution network model can use larger pixels in the input image to calculate each output pixel under the influence of the area. If you don't use dilated convolution, it will only use smaller pixel areas and cannot use more context information for image synthesis.

The generator uses a standard auto-encoder network [12], and on this basis, an expansion convolution layer is added, that is, the generator network proposed by the literature [44] to remove the middle two convolution layers. The network architecture is shown in Table 1. From left to right are the network layer type (*conv* is the convolution layer, $d - conv$ is the expansion convolution layer, *deconv*is the deconvolution layer), the size of the convolution kernel, the number of zero-filling of the convolution kernel, the step size, and the layer of output channels.

3.2. Discriminator

After training, the generator can fill the corresponding pixels of the missing area with a small reconstruction loss. Just using the generator does not ensure that the filled area is visually true and consistent. The pixels in the missing area of the generated image are very blurred, and only the general shape of the missing area can be captured. In order to obtain a more realistic effect, global discriminators, local discriminators, and multi-scale discriminators were added as binary classifiers to distinguish real image and fake image. The purpose is to distinguish whether the image is real or restored. The discriminator helps the network improve the quality of the repairing results, and the trained discriminator will not be fooled by unrealistic images. These discriminators were based on convolutional neural networks to compress images into corresponding small feature vectors. The prediction corresponds to the probability value that the image is true.

The first is the local discriminator D_l, which determines whether the synthesized content of the missing area is authentic information that helps the network to generate missing content. It encourages the generated objects to be semantically valid. Due to the locality of the local discriminator, its limitations are also obvious. Local discriminator can neither normalize the global structure of a face, nor guarantee the consistency of the inner and outer edges of the missing region. Therefore, the inconsistency of the pixel values of the repair image along the boundary of the repair area is obvious.

Due to the limitations of the local discriminator, another network structure called global discriminator D_g is introduced to determine the accuracy of the image as a whole. The basic idea is that the content of the generated image repair area must not only be true, but also be consistent with the context. A network with a global discriminator greatly alleviates the problem of inconsistency and further improves the effect of generating repaired images to make them more realistic.

Finally, a multi-scale discriminator network structure has been proposed. The basic idea is to downsample real and synthesized images with downsampling coefficients of two and four, and train two discriminators, D_{m1} and D_{m2}, to distinguish between real and repaired images on two different scales. Through two discriminator networks whose inputs are images of different resolutions, the process of repairing the image by the generator is strictly controlled. The two multi-scale discriminators and global discriminators had similar architectures but different receptive fields of different sizes. Compared with using the global discriminator, training with a joint multi-scale discriminator can guide the generator to generate a more globally consistent repairing image and finer details, and the repairing effect of the entire image is more reasonable visually. By adding two multi-scale discriminators

Table 2. Architecture of local discriminator D_l.

Layer	Kernel	Stride	Outputs
conv	5 × 5	2 × 2	64
conv	5 × 5	2 × 2	128
conv	5 × 5	2 × 2	256
conv	5 × 5	2 × 2	512
conv	5 × 5	2 × 2	512
conv	5 × 5	2 × 2	512

Table 3. Architecture of global discriminator D_g.

Layer	Kernel	Stride	Outputs
conv	5 × 5	2 × 2	64
conv	5 × 5	2 × 2	128
conv	5 × 5	2 × 2	256
conv	5 × 5	2 × 2	512
conv	5 × 5	2 × 2	512

Table 4. Architecture of multi-scale discriminator D_{m1}.

Layer	Kernel	Stride	Outputs
conv	5 × 5	2 × 2	64
conv	5 × 5	2 × 2	128
conv	5 × 5	2 × 2	256
conv	5 × 5	2 × 2	512
conv	5 × 5	2 × 2	512

Table 5. Architecture of multi-scale discriminator D_{m2}.

Layer	Kernel	Stride	Outputs
conv	5 × 5	2 × 2	64
conv	5 × 5	2 × 2	128
conv	5 × 5	2 × 2	256
conv	5 × 5	2 × 2	512

to the network, a better repaired image can be obtained.

The global discriminator and local discriminator are removed from the last two fully connected layers, and other structures aren't changed. The global discriminator, local discriminator, and multi-scale discriminator network architectures are shown in Tables 2–5, respectively. From left to right are the network layer type, the size of the convolution kernel, the step size, and the number of output channels of the layer. Tables 2–5 are D_l, D_g, D_{m1} and D_{m2}, respectively.

3.3. Loss function

There are usually many reasonable ways to fill in missing image areas that are consistent with the context. This behaviour can be modelled, for example, by a loss function. So, first introduce the reconstruction loss L_r to the generator, which is responsible for capturing the structural information of the missing area and keeping it consistent with the context, that is, L_2 distance between the pixels of the repair image and the original image, and z is the noise mask.

$$L_r(x, z) = ||z \odot (x - F((1 - z) \odot x))||_2^2 \quad (1)$$

But using L_r loss, it is observed that the content of the generated repaired image tends to be blurred and smooth. Because the cause of the L_2 distance loss is due to severe penalties for outliers, the network is encouraged to smoothly cross various assumptions to avoid large penalties. By using a discriminator, an adversarial loss is introduced, which reflects how the generator fools the discriminator to the maximum and how the discriminator distinguishes between real and fake. Adversarial loss is a loss based on the GAN model. In order to learn the generative model of data distribution, GAN learns an adversarial discriminator model D to provide a loss gradient for the generator model. The adversarial discriminator D predicts both the samples generated by the generator G and the real samples, and tries to distinguish them, while the generator G obfuscates the discriminator D by generating as "true" samples as possible:

$$L = \min_G \max_D E_{x \sim P_{data}(x)}[\lg(D(x))] + E_{z \sim P_z(z)}[\lg(1 - D(G(z)))] \quad (2)$$

Among them, $P_{data}(x)$ and $P_z(z)$ represent the distribution of real data x and noise variable z, respectively. Optimize the network by minimizing the generator loss and maximizing the discriminator loss.

Due to the instability in the training process of the traditional GAN model, the loss function and method of WGAN are used to train the GAN. The specific method is to remove the sigmoid of the last layer of the discriminator D. The loss function of G and D doesn't take the log of the loss function. The proposed method is used instead of the traditional GAN objective function:

$$L = \min_G \max_{D \in 1-Lipschitz} E_{x \sim P_{data}(x)}[D(x)] - E_{z \sim P_z(z)}[D(G(z))] \quad (3)$$

The discriminator D meets the $1 - Lipschitz$ limitation factor, which essentially requires that the degree of fluctuation of the network can't be too large. The specific method is to update the parameters of D every time and then cut off the absolute value so that it does not exceed a fixed constant, namely clipping.

The four discriminant networks $\{D_l, D_g, D_{m1}, D_{m2}\}$ have the same definition of the loss function. The only difference is that the local discriminator provides training loss gradients only for the missing regions, and the global discriminator and multi-scale discriminators back-propagate the loss gradients on the entire image with different resolutions. The input of the local discriminator D_l is a repaired part of the output image of the generator G and a part corresponding to the real image. The input of the global discriminator D_g is the output image and real image of the generator G. The input of the multi-scale discriminator D_{m1} is the output image and real image of the generator G output image and the real image down-sampled twice, respectively.

The input of the multi-scale discriminator D_{m2} is the output image and real image of the generator G output image and the real image, respectively, down-sampled four times. The discriminators are defined for formulas (4–7).

$$L_{D_l} = \min_G \max_{D_l \in 1-Lipschitz} E_{x\sim P_{data}(x)}[D_l(x)] - E_{z\sim P_z(z)}[D_l(G(z))] \quad (4)$$

$$L_{D_g} = \min_G \max_{D_g \in 1-Lipschitz} E_{x\sim P_{data}(x)}[D_g(x)] - E_{z\sim P_z(z)}[D_g(G(z))] \quad (5)$$

$$L_{D_{m1}} = \min_G \max_{D_{m1} \in 1-Lipschitz} E_{x\sim P_{data}(x)}[D_{m1}(x)] - E_{z\sim P_z(z)}[D_{m1}(G(z))] \quad (6)$$

$$L_{D_{m2}} = \min_G \max_{D_{m2} \in 1-Lipschitz} E_{x\sim P_{data}(x)}[D_{m2}(x)] - E_{z\sim P_z(z)}[D_{m2}(G(z))] \quad (7)$$

In summary, the total loss function of the entire network optimization is defined as formula (8).

$$L_{All} = L_r + \lambda_1 L_l + \lambda_2 L_g + \lambda_3 L_{m1} + \lambda_4 L_{m2} \quad (8)$$

λ_1, λ_2, λ_3 and λ_4 are the corresponding weights of different losses, which are used to balance the effects of different losses on the entire loss function. The specific values of λ_1, λ_2, λ_3 and λ_4 need to be set and adjusted manually during the experiment.

4. Training details

The proposed work is based on the implementation of deep convolutional adversarial neural networks. In order to effectively train the network, the training process is divided into three stages: (1) The generator network G is trained, and the network is trained using the reconstruction loss. The generator can get Vague repair content, this stage does not include confrontation training and confrontation loss. (2) Use the generator network trained in the first stage to train all discriminator networks $\{D_l, D_g, D_{m1}, D_{m2}\}$, and use the adversarial loss to update all discriminators. (3) The final stage trains the generator and all discriminators in joint confrontation. Each stage can prepare for the next stage of improvement, which greatly improves the effectiveness and efficiency of network training. The training process is completed by the Back Propagation method.

When training for adversarial loss, a method similar to the literature [52] is adopted to avoid the situation where the recognizer is too strong at the beginning of the training process. The default hyperparameters (such as learning rate), suggested in the literature [53], were used. Set λ_1, λ_2, λ_3 and λ_4 to 0.001. The training is done by adjusting the image size, and the image is cropped to a 256×256 image as the input image. For the missing area, the input of the central square area

Algorithm 1: Multi-Scale GAN Algorithm
1. while t $< T_3$ do
2. Collect small-batch images from training data x;
3. Generate a mask with random holes for each image x in the mini-batch;
4. if t $< T_1$ then
5. Update generator network G with weighted L_2 loss;
6. else
7. Generate a mask with random holes for each image x in the mini-batch;
8. Down-sampling of image x and $G(z)$ with ×2 and ×4, respectively;
9. Update all discriminators $\{D_l, D_g, D_{m1}, D_{m2}\}$ with EM distance loss;
10. if t $> T_1+T_2$ then
11. Joint adversarial loss gradient update generator networks G and $\{D_l, D_g, D_{m1}, D_{m2}\}$;
12. end if
13. end if
14. end while

in the image is set to 0. That is the missing part of the image covers about 1/4 of the image. The input of the global discrimination is a complete image of 256 × 256 size, the input of the local discriminator is an image of a repair area of 128 × 128 size, and the inputs of the two multi-scale discriminators are complete images of 128 × 128 and 64 × 64 sizes. The network model in the paper can reasonably fill the missing areas, but sometimes the generated areas will be inconsistent with the surrounding areas. To avoid this, simple post-processing is performed by blending the repaired area with the colours of surrounding pixels. In particular, the paper has used Poisson blending [54] for subsequent processing of images.

5. The experimental results' analysis

The paper has used 100,000 images obtained from the CelebA dataset to train a multi-scale generative adversarial network model. A total of 80,000 images were used for training and 20,000 images were used for the testing procedure. The dataset includes a variety of face images. The inpainting of face images is more difficult than the scene images. The inpainting of facial images requires more details. For example, the position of the facial features and the symmetry of the face make repairing relatively difficult, so higher requirements are imposed on the design of the neural network, and the batch size is set to 32. The generator network goes through 20,000 iterations, then trains the discriminator through 10,000 iterations, and finally trains the entire network 70,000 times. The device conditions are CPU, Intel i7-8700; GPU, RTX2080Ti; Main Memory, DDR4 16GB. The code runs under the Pytorch deep learning framework, and the entire network training takes about five days to complete.

We can also try to add more multi-scale discriminators. It is found in the experiment that two discriminators are enough to improve the network repairing effect. Adding too many discriminators will complicate the entire network, increase network parameters and operation time.

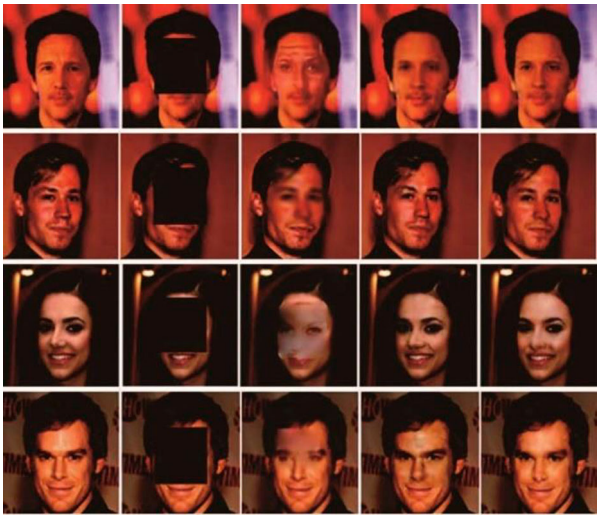

Figure 4. Comparison of repair results of different models (a) original image (b) unrepaired (c) GIICA (d) GLCIC (e) proposed method.

The experimental results were compared with the inpainting results of the GIICA (Generative Image Inpainting with Contextual Attention) [55] using only one discriminator acting on the repairing area, and the GLCIC (Globally and Locally Consistent Image Completion) [56] using the generator and two discriminators. In order to compare the fairness, the above model was retrained and the same number of iterations was performed, and the results are shown in Figure 4.

Figure 4 shows the face repairing results on the CelebA testing dataset [57]. In each testing image, the network will automatically cover the area in the middle of the image, because generally the middle part will contain important parts of the face (for example, eyes, mouth, eyebrows, hair, and nose). The column (a) in Figure 4 corresponds to four original non-missing images. The column (b) in Figure 4 is a masked missing image. Due to the lack of a globally consistent understanding of this structure, it can be seen that the repair results obtained by the proposed method not only have obvious global inconsistencies, but also repair the missing area. It is also very blurry and can't meet the requirements of image repairing tasks. The column (d) in Figure 4 is the repairing effect diagram of the "globally and locally consistent image completion" method with the global discriminator and local discriminator. The introduction of adversarial loss can enable the network to repair the image more reasonably. The area has an impact, so that the repair of missing area can be successfully completed. The global discriminator will affect the entire image in response to the global inconsistency of the repaired image, forcing the network to generate a globally consistent image, eliminating obvious edge differences. It is good. The column (e) in Figure 4 is the repair result of the algorithm in this paper, which uses the loss function of WGAN to make the training of the entire adversarial network more stable. Multi-scale discriminators are added, and training is performed jointly with global discriminators and local discriminators. It can be seen that compared with the result of column (d) in Figure 4, column (e) in Figure 4 has a certain improvement in the details of repair, the image is more integrated, and the repair effect is better.

Table 6. Quantitative experimental results on PSNR

Algorithm	PSNR/dB
CE	18.61
GLCIC	19.45
Proposed	19.61

In addition to the visual effects, the paper also uses the PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) for quantitative evaluation of the CelebA testing dataset. These two indicators are the repair results and original faces obtained through different methods. The first indicator is PSNR, which is an objective standard for evaluating images. It directly measures the difference in pixel values. The unit of PSNR is dB. The larger the value, the smaller the distortion. Assuming that the two images input are x and y, the calculation formula is given as formula (9) and formula (10).

$$MSE(x,y) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (X(i,j) - Y(i,j)) \quad (9)$$

$$PSNR(x,y) = 10 \lg \left(\frac{(2^n - 1)^2}{MSE} \right) \quad (10)$$

Among them, MSE represents the Mean Square Error of the repaired image x and the real image y, H and W are the height and width of the image, respectively, n is the number of bits per pixel, generally eight is taken, that is, the number of pixel grey levels is 256. The results are shown in Table 6.

The second indicator is the Structural Similarity Index (SSIM), which is a measure criterion of the similarity between two images. It is a number between zero and one. The larger the value, the smaller the difference between the repaired image and the real image. That is, the better the image quality. When the two images are identical, and the value is one. Assuming that the two image inputs are x and y, the calculation formula is formula (11).

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (11)$$

Among them, μ_x and μ_y represent the average values of x and y. σ_x and σ_y represent the standard deviations of x and y. σ_{xy} represents the covariance of x and y, and c_1 and c_2 are constants. The calculation results are shown in Table 7.

Table 7. Quantitative experimental results on SSIM.

Algorithm	SSIM
CE	0.773
GLCIC	0.799
Proposed	0.805

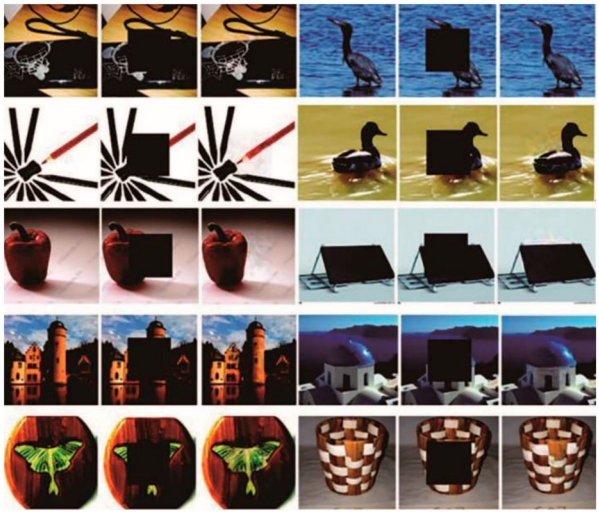

Figure 5. Repair results of the proposed algorithm on ImageNet dataset.

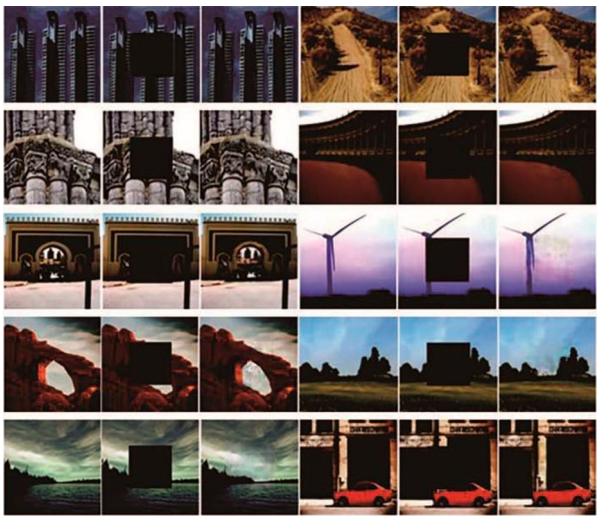

Figure 6. Repair results of the proposed algorithm on Place2 dataset.

In addition, in order to prove that the proposed algorithm in the paper can be applied to many types of image repairing, 50,000 images acquired in ImageNet dataset and 50,000 images acquired in Places2 dataset were used to train the model in the paper. The network model training method is the same as the training method used in the CelebA dataset. The experimental results are shown in Figures 5 and 6, indicating that the proposed model also performs well on the ImageNet dataset and Places2 dataset. Table 8 shows experimental results on PSNR and SSIM in ImageNet and Place2 datasets.

Table 8. Experimental results on PSNR and SSIM in ImageNet and Place2 datasets.

Dataset	SSIM	PSNR
ImageNet	0.7483	18.55
Place2	0.7869	19.21

6. Conclusions

In recent years, the deep learning had achieved great results in the field of computer vision. The research of image inpainting technology based on deep learning has achieved initial results and has a wide application prospect. The paper firstly introduces the background and significance of image inpainting technology, briefly reviews the current research status at home and abroad, and analyses the shortcomings of those existing algorithms. Then, it introduces the principle of generative adversarial network, analyses the problems existing in generative adversarial network, and applies the improved generative adversarial network model to the research of image repairing problems. Multi-scale generative adversarial networks model consists of adversarial discriminators. The reconstruction loss and multiple confrontation losses are used to synthesize the missing content from random noise. Combined with the idea of WGAN, EM distance is used to simulate the data distribution, which improves the stability of the network and improves the effect of image repairing. Finally, it is verified on the CelebA dataset. Using qualitative and quantitative evaluation methods, it is proved that the image repairing algorithm, based on multi-scale generative adversarial networks proposed in the paper, has better repairing effect than the current image repairing methods. The corresponding training and testing were also performed on the ImageNet and Place2, respectively, which proves that the proposed algorithm can be applied to the repair of many types of images and has good results.

In addition, during the image repairing experiment, it was found that in most cases, the network output image repairing effect is good, but in some cases, the image repairing output by the network will show some strange pixels, that is, artefacts, making the whole image looks very unnatural. The reason for this may be that the network has extracted the features of some invalid pixels during the convolution process. The situation isn't good for image repairing tasks. The purpose of the image repairing task is to complement the missing area as much as possible with the existing information in the image. The appearance of artefacts makes the repairing effect worse. The next work of the paper will improve the network model for this problem, and find a method that can eliminate artefacts, such as partial convolution, to achieve better image repair results.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 51408069).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the National Natural Science Foundation of China [No. 51408069].

ORCID

Jiangchun Mo http://orcid.org/0000-0002-2315-0664

References

[1] Chen YT, Wang J, Xia RL, et al. The visual object tracking algorithm research based on adaptive combination kernel. J Ambient Intell Humaniz Comput. 2019;10(12):4855–4867.
[2] Zhang JM, Wang W, Lu CQ, et al. Lightweight deep network for traffic sign classification. Ann Telecommun. 2020;75:369–379. DOI:10.1007/s12243-019-00731-9.
[3] Chen YT, Wang J, Liu SJ, et al. Multiscale fast correlation filtering tracking algorithm based on a feature fusion model. Concurr Comput Pract Exp. 2019. DOI:10.1002/cpe.5533.
[4] Lu WP, Zhang YT, Wang SJ, et al. Concept representation by learning explicit and implicit concept couplings. IEEE Intell Syst. 2020DOI:10.1109/MIS.2020.3021188.
[5] Zhang JM, Xie ZP, Sun J, et al. A cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection. IEEE Access. 2020;8:29742–29754.
[6] Chen YT, Wang J, Chen X, et al. Single-image super-resolution algorithm based on structural self-similarity and deformation block features. IEEE Access. 2019;7:58791–58801.
[7] Chen YT, Tao JJ, Liu LW, et al. Research of improving semantic image segmentation based on a feature fusion model. J Ambient Intell Humaniz Comput. 2020. DOI:10.1007/s12652-020-02066-z.
[8] Sun L, Wu F, Zhan T, et al. Weighted nonlocal low-rank tensor decomposition method for sparse unmixing of hyperspectral images. IEEE J Sel Top Appl Earth Obs. 2020;13:1174–1188. DOI:10.1109/JSTARS.2020.2980576.
[9] Sun L, Ma C, Chen Y, et al. Low rank component induced spatial-spectral kernel method for hyperspectral image classification. IEEE Trans Circ Syst Viedo T. 2019DOI:10.1109/TCSVT.2019.2946723.
[10] Chen YT, Tao JJ, Zhang Q, et al. Saliency detection via improved hierarchical principle component analysis method. Wirel Commun Mob Comput. 2020;vol.2020, Article ID 8822777.
[11] Yu F, Liu L, Xiao L, et al. A robust and fixed-time zeroing neural dynamics for computing time-variant nonlinear equation using a novel nonlinear activation function. Neurocomputing. 2019;350:108–116.
[12] Chen YT, Liu LW, Tao JJ, et al. The improved image inpainting algorithm via encoder and similarity constraint. Vis Comput. 2020. DOI:10.1007/s00371-020-01932-3.
[13] Chen YT, Liu LW, Tao JJ, et al. The image annotation algorithm using convolutional features from intermediate layer of deep learning. Multimed Tools Appl. 2020. DOI:10.1007/s11042-020-09887-2.
[14] Yu F, Qian S, Chen X, et al. A new 4D four-wing memristive hyperchaotic system: dynamical analysis, electronic circuit design, shape synchronization and secure communication. Int J Bifurcation Chaos. 2020. DOI:10.1142/S0218127420501412.
[15] Yu F, Zhang Z, Liu L, et al. Secure communication scheme based on a new 5D multistable four-wing memristive hyperchaotic system with disturbance inputs. Complexity. 2020;vol.2020, Article ID 5859273.
[16] Yu F, Liu L, Qian S, et al. Chaos-based application of a novel multistable 5D memristive hyperchaotic system with coexisting multiple attractors. Complexity. 2020;vol.2020, Article ID 8034196.
[17] Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern. 1980;36(4):193–202.
[18] Lecun Y, Boser B, Denker JS, et al. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1989;1(4):541–551.
[19] Luo YJ, Qin JH, Xiang XY, et al. Coverless real-time image information hiding based on image block matching and dense convolutional network. J Real-Time Image Process. 2020;17(1):125–135.
[20] He SM, Li ZZ, Tang YN, et al. Parameters compressing in deep learning. CMC-Comput Mater Contin. 2020;62(1):321–336. DOI:10.32604/cmc.2020.06130.
[21] Wang J, Qin JH, Xiang XY, et al. CAPTCHA recognition based on deep convolutional neural network. Math Biosci Eng. 2019;16(5):5851–5861.
[22] Zhang JY, Zhong SQ, Wang T, et al. Blockchain-based systems and applications: a survey. J Internet Technol. 2020;21(1):1–14. DOI:10.3966/160792642020012101 0100.
[23] Yu F, Liu L, He B, et al. Analysis and FPGA realization of a novel 5D hyperchaotic four-wing memristive system, active control synchronization, and secure communication application. Complexity. 2019;vol.2019, Article ID 4047957.
[24] Gao P, Zhang QQ, Wang F, et al. Learning reinforced attentional representation for end-to-end visual tracking. Inform Sciences. 2020;517:52–67. DOI:10.1016/j.ins.2019.12.084.
[25] Tang Q, Chang L, Yang K, et al. Task number maximization offloading strategy seamlessly adapted to UAV scenario. Comput Commun. 2019. DOI:10.1016/j.comcom.2019.12.018.
[26] Gao P, Yuan RY, Wang F, et al. Siamese attentional keypoint network for high performance visual tracking. Knowl Based Syst. 2019. DOI:10.1016/j.knosys.2019.105448.
[27] Lu WP, Zhang X, Lu HM, et al. Deep hierarchical encoding model for sentence semantic matching. J Vis Commun Image Represent. 2020;71:102794.
[28] Chen YT, Xu WH, Zuo JW, et al. The fire recognition algorithm using dynamic feature fusion and IV-SVM classifier. Cluster Comput. 2019;22:7665–7675.
[29] Chen YT, Xiong J, Xu WH, et al. A novel online incremental and decremental learning algorithm based on variable support vector machine. Cluster Comput. 2019;22:7435–7445.
[30] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic

segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, 24–27 Jun 2014, Columbus. Piscataway: IEEE. p. 580–587.
[31] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. ACM Commun. 2017;60(6):84–90.
[32] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell. 2017;39(4):640–651.
[33] Bertalmio M, Sapiro G, Caselles V, et al. Image Inpainting. Proceedings of ACM SIGGRAPH, 23–28 July 2000, New Orleans. p. 417–424.
[34] Bertalmio M, Vese L, Sapiro G, et al. Simultaneous structure and texture image inpainting. IEEE Trans Image Process. 2003;12(8):882–889.
[35] Zoran D, Weiss Y. From learning models of natural image patches to whole image restoration. Proceedings of the IEEE International Conference on Computer Vision, Barcelona, 6–13 Nov 2011. Piscataway: IEEE. p. 479–486.
[36] Ren JS, Xu L, Yan Q, et al. Shepard convolutional neural networks. Proceedings of the Advances in Neural Information Processing Systems, Montreal, 7–12 Dec 2015. Denver (CO): NIPS. p. 901–909.
[37] Zhang JM, Lu CQ, Wang J, et al. Training convolutional neural networks with multi-size images and triplet loss for remote sensing scene classification. Sensors. 2020;20:1188.
[38] Xiang LY, Yang SH, Liu YH, et al. Novel linguistic steganography based on character-level text generation. Mathematics . 2020;8(9):1558–1558. DOI:10.3390/math8091558.
[39] Liao ZF, Peng JS, Chen YT, et al. A fast Q-learning based data storage optimization for low latency in data center networks. IEEE Access. 2020;8:90630–90639.
[40] Barnes C, Shechtman E, Finkelstein A, et al. Patchmatch: a randomized correspondence algorithm for structural image editing. ACM Trans Graph. 2009;28 (3):24.
[41] Wright J, Yang AY, Ganesh A, et al. Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell. 2009;31(2):210–227.
[42] Larsen A, Sonderby S, Winther O. Autoencoding beyond pixels using a learned similarity metric. Proceedings of the International Conference on Machine Learning, New York, 19–24 Jun 2016. New York (NY): ACM. p. 1558–1566.
[43] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN [EB/OL]. arXiv:1701.07875v3 (2017-12-06). https://arxiv.org/abs/1701.07875.
[44] Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image competition. ACM Trans Graph. 2017;36(4):107.
[45] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training GANs. Proceedings of the Advances in Neural Information Processing Systems, Barcelona, 5–10 Dec 2016. Denver (CO): NIPS. p. 6125–6134.
[46] Cao D, Zheng B, Ji BF, et al. A robust distance-based relay selection for message dissemination in vehicular network. Wirel Netw. 2020;26:1755–1771. DOI:10.1007/s11276-018-1863-4.
[47] Chen YT, Wang J, Chen X, et al. Image super-resolution algorithm based on dual-channel convolutional neural networks. Appl Sci. 2019;9(11):2316.
[48] Yang L, Wang J, Tang Z, et al. Using conditional random fields to optimize a self-adaptive bell-LaPadula model in control systems. IEEE Trans Syst Man Cybern Syst. 2019. DOI:10.1109/TSMC.2019.2937551.
[49] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. Proceedings of the International Conference on Mach ine Learning, Sydney, 6–11 Aug 2017. New York (NY): ACM. p. 214–233.
[50] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein GANs. Proceedings of the Advances in Neural Information Processing Systems, Long Beach, 22–29 Oct 2017. Denver (CO): NIPS. p. 5769–5779.
[51] Mao XD, Li Q, Xie HR, et al. Least squares generative adversarial networks. Proceedings of the IEEE International Conference on Computer Vision, Venice, 6–13 Nov 2017. Piscataway: IEEE. p. 2794–2802.
[52] Yu F, Koltun F. Multi-scale context aggregation by dilated convolution. Proceedings of the International Conference on Learning Representations, Utah, 2–4 May 2016. San Juan: ICLR. p. 514–526.
[53] Wang TC, Liu MY, Zhu JY, et al. High-resolution image synthesis and semantic manipulation with conditional GANs. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 18–22 Jun 2018. Piscataway: IEEE. p. 8798–8807.
[54] Perez P, Gangnet M, Blake A. Poisson image editing. ACM Trans Graph. 2003;22(3):313–318.
[55] Yu JH, Lin Z, Yang JM, et al. Generative image inpainting with contextual attention. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 18–22 Jun 2018. Piscataway: IEEE. DOI:10.1109/CVPR.2018.00577.
[56] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. arXiv:1511.06434 (2016-01-17) [2019-04-05]. https://arxiv.org/abs/1511.06434.
[57] Wang XL, Gupta A. Generative image modeling using style and structure networks [EB/OL]. arXiv:1603.05631 (2016-07-26) [2019-04-05]. https://arxiv.org/abs/1603.05631.
[58] Kingma DP, Welling M. Auto-encoding variational Bayes [EB/OL], arXiv:1312.6114 (2013-12-20) [2019-04-05]. https://arxiv.org/abs/1312.6114v1.
[59] Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. Proceedings of the International Conference on Neural Information Processing Systems, Montreal, 8–13 Dec 2014. Denver (CO): NIPS. p. 2672–2680.
[60] Mirza M, Osindero S. Conditional generative adversarial nets [EB/OL]. arXiv:1411.1784 (2014-11-06) [2019-04-05]. https://arxiv.org/abs/1411.1784.