

Face inpainting based on edge confrontation combined with hierarchical gated convolution

ZHAI Fengwen*, ZHOU Zhao, SUN Fanglin, JIN Jing

School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

*Corresponding author: ZHAI Fengwen (zhaifw@mail.lzjtu.cn)

Received: September 14, 2022

Revised: October 18, 2023

Accepted: October 23, 2023

Abstract: Aiming at the problems of edge blur and distortion in the current damaged face image inpainting, a two-stage hierarchical gated convolutional network (HGNC) was proposed and then combined with edge adversarial network for face image inpainting. Firstly, the edge adversarial network was adopted to generate edge images. Secondly, the edge images, the masks and the occluded images were combined to train the generative adversarial network (GAN) model of the HGNC to generate the inpainted face images. In the HGNC, traditional convolution was replaced by gated convolution and the dilated convolution was introduced. The main structure of the HGNC is composed of coarse inpainting module and fine inpainting module. In the coarse inpainting module, the encoder and decoder network structure was used for coarse inpainting. In the fine inpainting module, the attention mechanism was introduced to enhance the feature extraction ability so as to further refine the inpainting results. In the experiment, the Celeba-HQ dataset and NVIDIA irregular mask dataset were used as the training datasets, the gated convolution network and attention module were adopted as comparing networks, and PSNR, SSIM and MAE were used as evaluation indicators. The experimental results demonstrated that for the face images with missing areas less than 20%, the proposed network works better than the two other networks on the above three indicators, and for the face images with missing areas greater than 20%, the proposed network is close to the comparison networks on three indicators. In terms of visual effects, the proposed method also surpasses the two contrasting networks in details. The proposed network can evidently improve the inpainting effect, especially image details.

Key words: deep learning; face inpainting; hierarchical gated convolutional network (HGNC); edge confrontation; generative adversarial network (GAN)

0 Introduction

Image inpainting is a process that restores the damaged areas in the image according to the known contents of the image, which makes the repaired images feel real and semantically complete so as to convey correct information^[1,2].

The traditional image inpainting algorithms based on graphics and vision are mainly to find similar parts in the given area to repair the missing area. Such algorithms are based on mathematical and physical theories, and thus readers can understand relevant principles. In addition, they need not massive datasets for training and therefore are high-efficient.

Usually, the traditional image inpainting algorithms are divided into four categories: diffusion based image inpainting^[3,4], texture composite image inpainting^[5,6], data-driven image inpainting^[7,8], and sparse representation image

inpainting^[9,10]. Among them, the diffusion based image inpainting method spreads the known region information to the edge region, which is suitable for inpainting the images with small missing areas. The texture based synthesis method firstly divides the image into a set, and then locates the most similar blocks from the set to fill the missing areas, which is not suitable for the images with complex content. The data-driven image inpainting method searches for similar images in the external database in a data-driven way to fill the missing image areas. This method involves a lot of computation and its application scenarios are limited by the database. The sparse representation image inpainting method utilizes the same sparseness about the known area and the missing area. By use of the collected information of the known area, sparse representation is performed and the image is reconstructed, which is computationally intensive and complex.

Additionally, the traditional image inpainting methods

cannot understand the high-level semantics in the image, thus the inpainting effects of complex images and large-area missing images are relatively poor, which cannot meet the actual needs.

In recent years, the image inpainting methods based on deep learning, such as convolutional neural networks (CNN) ^[11,12] and generative adversarial networks (GAN) ^[13,14], have become a research hotspot in the field of computer vision. Based on the feature information and deep semantic information in the image via massive data learning, the image with complex content can be processed and the realistic inpainting effect for the image with large missing areas can be achieved. In addition, the unknown area of the image can also be expanded and repaired via the texture information and structural information of the known area in the damaged image. Pathak et al. ^[15] adopted a GAN combined with the encoder and decoder network structure to generate the missing areas in the image. Although this method repairs the images with large missing areas, the effect of repairing images is relatively rough. Iizuka et al. ^[16] improved the discriminator of Ref. [15]. They used two discriminators, a global discriminator and a local discriminator to enhance the local and global consistency of the image. In addition, they used dilated convolution to increase the receptive field, making the image clearer after inpainting. Yu et al. ^[17] replaced the traditional convolution and partial convolution with gated convolution and used spectral normalized Markov discriminator to finally complete the repair of irregular missing areas. It can also repair the images with different resolutions and the effect has certain improvement. Nazeri et al. ^[18] proposed a two-stage network structure. It repairs the imaginary edge contour of the damaged image in the first stage, and estimates the pixel value of the missing area according to the imaginary image in the second stage. The texture restoration effect of this algorithm is better than that of previous algorithms. Guo et al. ^[19] proposed a two-stream network to inpaint images by simultaneously modeling structure-constrained texture synthesis and texture-guided structure reconstruction, which mutually promotes the inpainting of image texture information and the inpainting of structural information, but it is less efficient.

There exist notable problems from the current inpainting algorithms: the details of the inpainted image are not rich enough, and the images with complex content are prone to deformation after repair. In response to the above problems, we proposed a two-stage hierarchical gated convolutional network (HGNC) structure with inpainting effect from coarse to fine, which was combined with edge adversarial network was

improve the detail inpainting effect. In this process, the edge adversarial network was responsible for predicting the edges of the missing areas, and the HGNC was responsible for image inpainting according to the predictions.

1 Basic theory

1.1 Gated convolution module

Liu et al. ^[20] proved that conventional convolution does not distinguish between valid and invalid regions of the damaged image during inpainting. Since the missing areas are incorrectly regarded as the valid areas, the repair results will be affected accordingly. Therefore, they replaced the conventional convolution with the partial convolution, which uses certain rules to update the mask to distinguish the missing areas from valid areas, and the result is only affected by the pixels in the valid area rather than that in the missing area, so the repair effect is greatly improved. The rules for partial convolution are given by

$$x' = \begin{cases} W^T (X \odot M) \frac{1}{\text{sum}(M)} + b, & \text{if } \text{sum}(M) > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

$$m' = \begin{cases} 1, & \text{if } \text{sum}(M) > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where W represents the weight of the convolution kernel; X is the pixel value of the image area corresponding to the convolution kernel; M is the binary mask of the mask area corresponding to the convolution kernel; b is the bias; \odot is element-wise multiplication; x' is the result of convolution at one position; and m' is the output of the input mask after convolution operation. It can be seen from Eqs. (1) and (2) that the normal convolution operation is performed only when there are valid pixels in the corresponding position of the convolution kernel, so the output value is only affected by the valid pixel value.

The partial convolution is divided into two parts: the convolution of input damaged image and the convolution of input mask image. As shown in Fig. 1, each convolution operation of partial convolution uses updated mask. As the network deepens, the missing area of the mask gradually decreases, and the valid pixels of the image gradually increase. Partial convolution is able to distinguish between valid and invalid areas of an image, however, its mask update rule is not flexible enough and cannot be learned neither. Gated

convolution^[21,22] is an improvement on partial convolution, and it can solve the above problems of partial convolution.

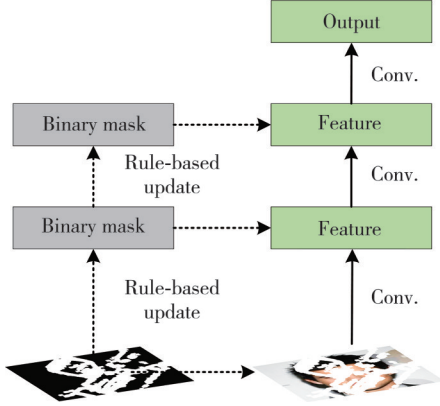


Fig. 1 Process of partial convolution

The most prominent advantage of gated convolution is that it can gradually learn from massive data to update the mask, which replaces the stereotyped rule-based method of updating the mask. Gated convolution can learn a mechanism for each channel and spatial location of the data, and thus enable dynamic feature selection based on image background, mask images, and user sketches. When the number of network layers is relatively large, it can also learn to highlight mask images and edge information in different channels, which can improve the inpainting effect.

It adopts a learnable mask update rule, and the update rule is expressed as

$$Gating_{y,x} = \sum \sum W_g I, \quad (3)$$

$$Feature_{y,x} = \sum \sum W_f I, \quad (4)$$

$$O_{y,x} = \phi(Feature_{y,x}) \odot \sigma(Gating_{y,x}), \quad (5)$$

where σ refers to the sigmoid function; ϕ can be any activation function, such as ReLU or ELU; x and y are the position information; I is the input image; and W_f and W_g are two different convolution filters.

The sigmoid activation function is used to perform dynamic feature selection on the output after the convolution by Eq. (3), and other activation functions are used to achieve image feature extraction on the output after the convolution by Eq. (4). The outputs of these two parts are then dot-multiplied to more efficiently select and extract useful information from the image. Partial convolution based on updating rule uses only 0 or 1 for weighting, while gated convolution uses the number between 0 and 1 for weighting, so the former is more flexible and learnable.

The process of gated convolution is shown in Fig. 2. It uses standard convolution to update the mask for each channel, and then uses the updated mask on the feature

map to continue the next convolution, which makes the mask update process learnable.

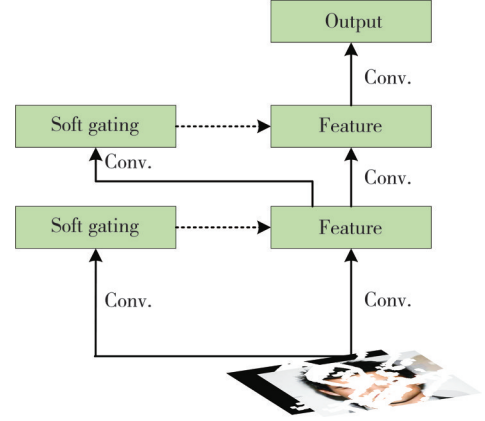


Fig. 2 Process of gated convolution

1.2 GAN

GAN is derived from a zero-sum game and widely used in the field of image inpainting. As for the disadvantages of GAN, unstable training process, gradient disappearance, mode collapse, and over-generation are included.

Aiming at the problem of unstable training process of GAN, we adopted spectral normalization GAN (SN-GAN)^[23,24] and patch GAN^[25,26]. The former realizes the spectral normalization of GAN by constraining the parameters of the convolution kernel in each layer of the network and making the discriminator satisfy the Lipschitz condition. The spectral normalization can effectively control the violent fluctuation of the function, thereby making the network training process more stable. The discriminator of patch GAN adopts a fully convolutional network, which maps the input into an n -order matrix, and takes the average value of the matrix as the final output. A point on the feature map can be traced back to the corresponding position in the original image, so that the influence of a point at a specific position on the output can be observed. In addition, using a matrix for evaluation, more information can be focused on and details can be handled better.

1.3 Attention mechanism module

The attention mechanism^[27,28] can help us pick out the most important parts of the output from the complex input information, and its principle is shown in Fig. 3. The input part is the feature map after convolution, with a size of $c \times w \times h$. The subsequent processing of the feature map is divided into two parts. The upper branch extracts features from different scales of the image to obtain the feature vector, and then weights the feature

vector to obtain a $1 \times 1 \times c$ feature map. The output of the lower branch is the original feature map after

convolution. Finally, the results of the two branches are fused to improve the feature extraction ability.

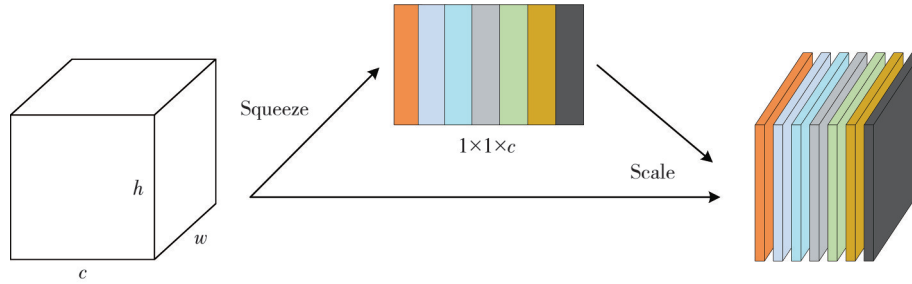


Fig. 3 Attention mechanism module

1.4 Edge adversarial network

The edge adversarial network proposed by Nazeri et al.^[18] includes a generator network and a discriminator network, which is shown in Fig. 4. The input to the generator network consists of mask images, edge images and grayscale damaged images. After a multi-layer

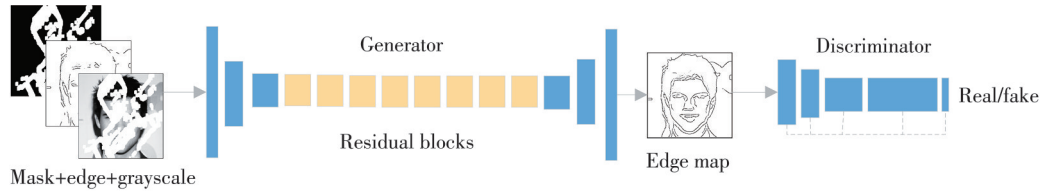


Fig. 4 Edge adversarial network

The discriminator network adopts patch GAN to discriminate the authenticity of the generated edges, which performs multi-scale feature matching between the edges generated by the generator and the real edges, and then continuously updates the parameters of the generator and the discriminator through back-propagation algorithm. Finally, the generator generates ideal edge images. The constructed loss function used to train the generator is shown as

$$L_{adv} = E_{(C_{gt}, I_{gray})} \log [D(C_{gt}, I_{gray})] + E_{I_{gray}} \log [1 - D(C_{pred}, I_{gray})], \quad (6)$$

where I_{gray} stands for the real grayscale image, C_{gt} stands for the real edge image, and C_{pred} stands for the predicted edge image.

When training the discriminator, the adversarial loss combined with the feature matching loss is used, that is,

$$\min_G \max_D L_G = \min_G (\lambda_{adv} \max_D (L_{adv}) + \lambda_{FM} L_{FM}), \quad (7)$$

where L_{FM} is the feature matching loss, and λ_{FM} and λ_{adv} are regularization parameters.

2 Edge adversarial network combined with HGNC

Since the inpainting results based on simple one-stage

convolution operation, the structure and texture features in the image are extracted. The generator network consists of an encoder with twice down-sampling convolutions, eight residual blocks and a decoder with twice up-sampling convolutions. The residual block can prevent network degradation, and the dilated convolution can increase the receptive field.

repair network and traditional convolution method are prone to blurs and incoherence of the edges, a two-stage HGNC structure with inpainting effect from coarse to fine was proposed, which was combined with the edge adversarial network to inpaint the face images with missing areas. The HGNC is more refined than the one-stage network for it introduces gated convolution, dilated convolution and attention mechanisms. The gated convolution can distinguish the valid area and the invalid area of the image, and it also can highlight the mask and the edge information provided by the edge adversarial network. The dilated convolution can extract multi-scale image features and reduce information loss. The attention mechanism was introduced in the second stage of the HGNC, which can enhance the feature extraction ability and makes the network pay attention to the important information. Therefore, the HGNC and the edge adversarial network were designed to achieve a more refined inpainting effect.

2.1 Proposed HGNC

The HGNC proposed mainly consists of three parts: a coarse inpainting part, a fine inpainting part and a discriminant part.

The structure parameters of the HGNC are shown in

Fig. 5. The four-digit numbers connected by underlines in the figure represent the number of input channels, the number of output channels, the size of the convolution kernel and the step size, respectively.

The encoder and decoder structures are used for the coarse inpainting network, and the standard convolution is replaced by gated convolution. At the same time, the network parameters are dynamically learned to effectively distinguish the missing area from the valid area. The coarse inpainting network takes the damaged image, mask image and edge image as inputs, and contains seventeen gated convolutional layers, two of which are down-sampling convolutional layers and two are up-sampling convolutional layers. The gated convolution module takes the sigmoid function and the ELU function as activation functions. After all gated convolution operations done, the tanh activation

function gets the result of the coarse inpainting network to generate a blurred inpainting image.

The fine repair network also adopts the encoder and decoder structure. The blurred image generated in the coarse inpainting stage is used as the input of the fine inpainting stage. The refined inpainting network consists of two parts: a gated convolutional branch and a perceptual layer branch, where the attention mechanism is introduced. The convolutional branch contains ten gated convolutional layers, two of which are down-sampling convolutional layers, and the perception layer branch contains eight gated convolutional layers and an attention module. The gated convolution uses sigmoid function and ReLU function as activation functions. The results of the two branches are concatenated and decoded to get the refined inpainted image results, where the tanh activation function is used.

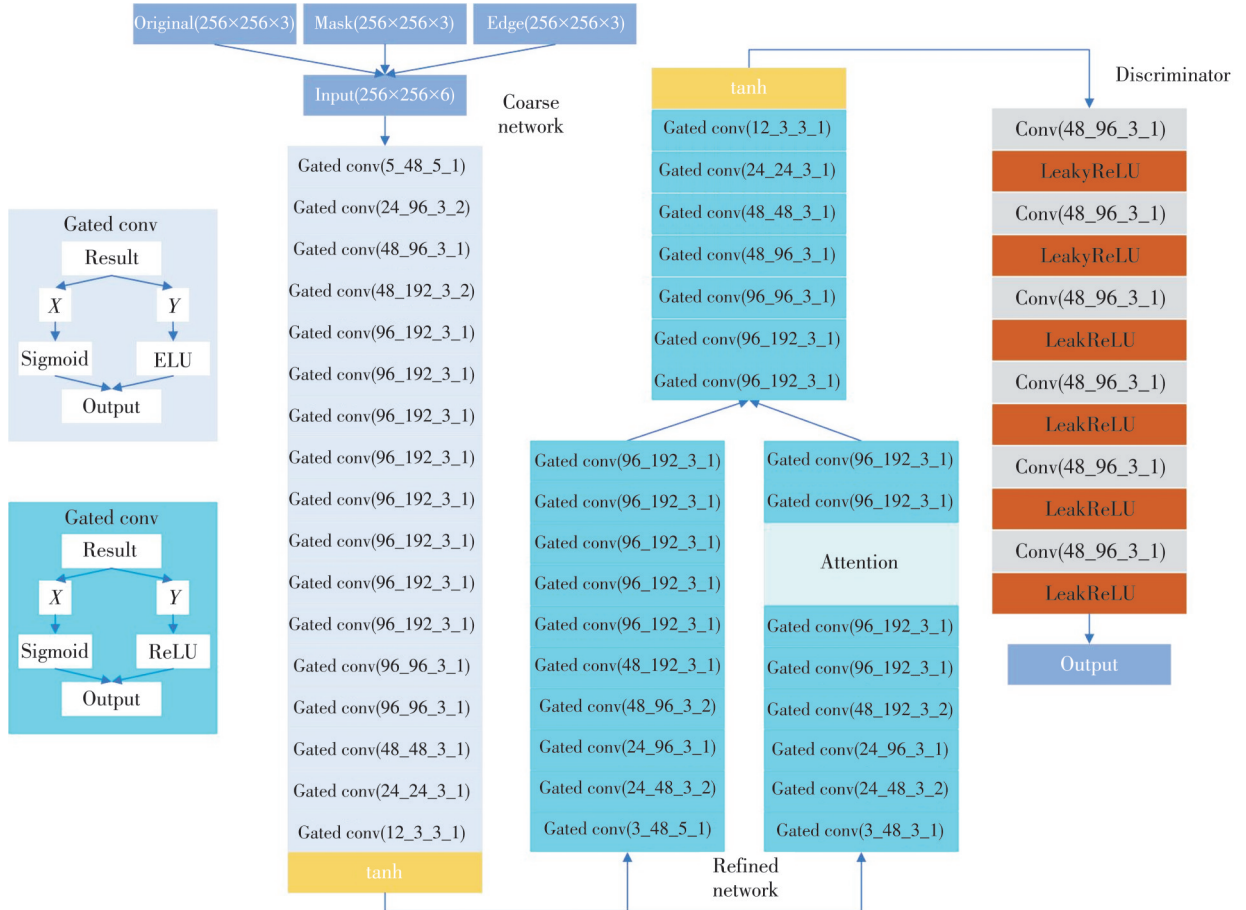


Fig. 5 Structure of proposed HGCN

The input feature image of the attention mechanism module contains known background areas and unknown foreground ones. The known background area is divided into n pixel blocks with a block size of 3×3 pixels, and these pixel blocks are then converted into convolution kernels. Then, the cosine similarity between the foreground unknown area and the background

convolution kernel is calculated, and the attention score is calculated by softmax^[29,30]. The higher the score, the more similar the pixel block and the area to be inpainted. Afterwrd, the missing areas are inpainted by deconvolution using the information of the most similar areas. The image feature map and attention map obtained by attention mechanism module from the space

and channel respectively and then are fused to enhance the feature extraction ability. The refinement network with the perceptual layer can improve the clarity of detailed texture.

The discriminator network is a fully convolution network, which takes the generated image, mask and edge image as inputs. It contains six stridden convolutions with 5×5 convolution kernel and the stride size is 2. Meanwhile, LeakyReLU function is used as the activation function, and SN-Patch GAN is used for discrimination. SN-GAN constrains the discriminator Lipschitz constant by constraining the spectral norm of the weight matrices of all layers of the discriminator network to enhance the training stability. The discriminator of patch GAN has the same dimension as that of the feature map. Each value represents the evaluation result of a small area of the generated image, so it can capture global and local information. Patch GAN encodes the image block information, and the receptive field of each element in the final output feature map covers the whole image, which can pay attention to more detailed information and improves the inpainting effect of image details.

2.2 Loss function for HGCN

The proposed HGCN adopts L1 loss and SN-Patch GAN loss. The L1 reconstruction loss is used both in

coarse inpainting stage and fine inpainting stage, and the loss function is calculated by

$$L_{adv} = |X_{ori} - X_{cor}| + |X_{ori} - X_{ref}|, \quad (8)$$

where X_{ori} stands for the original image, X_{cor} stands for the coarse inpainting result, and X_{ref} stands for the fine inpainting result, among which the loss includes the loss of the missing area and the loss of the given area. Taking the loss of the entire image to be inpainted as the loss of the generation part can improve the learning ability of the network and has a better inpainting effect in terms of image color consistency.

The discriminator network uses SN-Patch GAN loss as well as the generator and discriminator hinge loss, which are calculated by

$$L_G = -E_{z \sim P_z(z)}[D^{sn}(G(z))] \quad (9)$$

$$L_D = E_{x \sim P_{data}(x)}[\text{ReLU}(1 - D^{sn}(x))] + E_{z \sim P(z)}[\text{ReLU}(1 + D^{sn}(G(z)))],$$

where D^{sn} represents the spectral normalization discriminator. Each element in the output feature map has the corresponding loss, which can better focus on different positions in different channels of the image.

2.3 Edge adversarial network combined with HGCN

The structure of edge adversarial network combined with HGCN is shown in Fig.6.

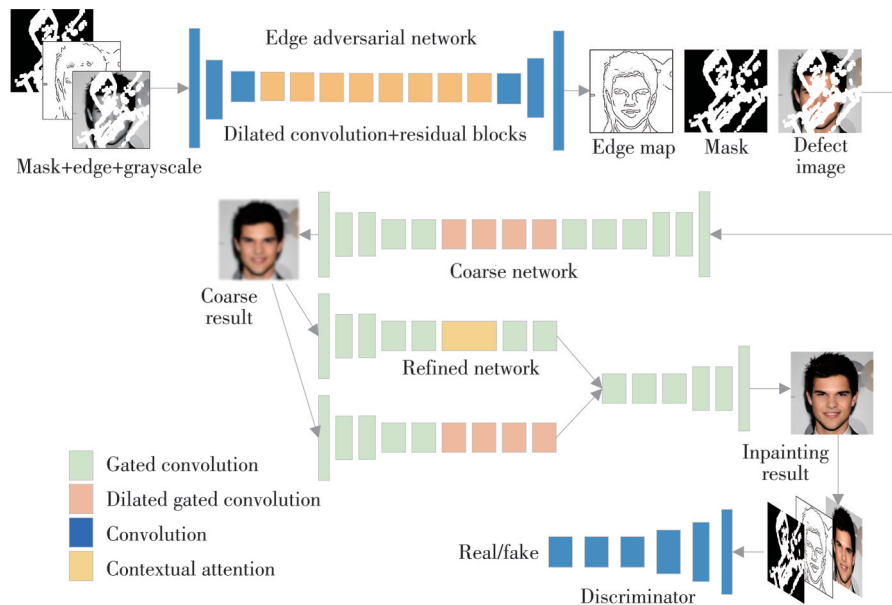


Fig. 6 Structure of edge adversarial network combined with HGCN

Firstly, the masks, image edges and grayscale occluded images are used to train the edge GAN model, so that the generator can generate complete face edge images. The discriminator performs multi-scale feature

matching between the edge generated by the generator and the real edge, and continuously updates the parameters of the generator and the discriminator through back propagation. Along with the confrontation

between the generator and the discriminator, the generating ability of the generator to generate accurate edge map and the discriminating ability of the discriminator become stronger. As a result, the generator can generate the ideal edge image.

Secondly, the edge map is used as the guidance information of the HGCN, and the masks and the occluded images are combined to train the GAN model of the HGCN, so that the network generator can generate the inpainted face images. The discriminator continuously updates the parameters to improve the discriminating ability by comparing the inpainted image with the original image, and at the same time, it can promote the generator to generate a face image that is more consistent with the edge map.

Finally, after the entire network is trained, the edge adversarial network can generate an ideal edge map, and the HGCN can repair a face image that conforms to the edge map. Thus end-to-end face image inpainting is achieved by inputting the occluded image and the corresponding mask through a deep inpainting network.

3 Experiment and analysis

3.1 Experiment platform

Two datasets were used in the experiment, one is the face dataset and the other is the mask dataset. The face dataset used was the Celeba-HQ dataset, which was generated by Liu et al.^[20] based on the CelebA dataset by training high-resolution GAN to provide the images with different resolutions. It contains 30 000 high-definition face images. The mask dataset adopted the testing set in NVIDIA irregular mask dataset, which contains 12 000 irregular mask images.

The experimental platform used the Ubuntu 18.04 operating system, Tensorflow 1.6.0 framework, Python 3.6, the PyCharm 2021 version, and RTX 2080Ti graphics card with 11 GB video memory.

3.2 Algorithm training

The training process of the HGCN is shown in Table 1. Firstly, m images were selected from the training dataset, and then the m images were randomly occluded to obtain the occluded images. Secondly, the generated edge images, masks and occluded images were put into the coarse inpainting network. Thirdly, the coarse inpainting results were put into the fine repair network. Fourthly, m inpainted images were put into the discriminator, SN-Patch GAN loss was used for training, and then the discriminator parameters were

updated. After k rounds of training, the generated edge images and the occluded images were put into the generator, and then the generator parameters were updated with L1 loss. In Table 1, \hat{E}_i represents the i th occluded image, $X_{cor}^{(i)}$ represents the results of the coarsely i th inpainted image, and $X_{pred}^{(i)}$ represents the results of the i th finely inpainted image.

Table 1 HGCN training process

1	While the generator has not converged
2	For k steps do
3	Sample batch images m from training dataset
4	Randomly occlude the map of m images, and generate a dataset of m masked images $\{\hat{E}_1, \hat{E}_2, \dots, \hat{E}_m\}$
5	Enter m final edge maps and m occluded images to the coarse inpainting network and obtain m coarse images $\{X_{cor}^{(1)}, \dots, X_{cor}^{(m)}\}$
6	Enter m coarse images to the refine repair network and obtain m refined images $\{X_{pred}^{(1)}, X_{pred}^{(2)}, \dots, X_{pred}^{(m)}\}$
7	Enter m refined images to the discriminator and update the parameters of the discriminator with the SN-Patch GAN loss
8	End For
9	Sample batch images m from training dataset
10	Randomly occlude the map of m images, and generate a dataset of m masked images $\{\hat{E}_1, \hat{E}_2, \dots, \hat{E}_m\}$
11	Enter batch of m final edge maps and m occluded images to the generator and update the parameters of the generator with reconstruction L1 loss
12	End while

3.3 Qualitative analysis

In this part of the experiment, the visual effects of the original gated convolution method, the attention mechanism and the proposed method were compared and analyzed. The inpainted results of these three algorithms are shown in Fig. 7. The original image is shown in (a), the mask image in (b), the occluded image in (c), the inpainted image using the gated convolution model in (d), the inpainted image using the attention mechanism module in (e), and the inpainted image of the proposed method in (f).

It can be seen from Fig. 7 (d) that the inpainting result of the original gated convolution method is visually reasonable, but there is a problem that the compensation information is insufficient. For example, in Image 1 and Image 5, the color of the eyebrow parts of the inpainted images are light and the texture is sparse. In Fig. 7 (e), the image inpainted by attention mechanism module has some problems such as edge blurring and texture distortion. In Fig. 7 (f), the proposed method uses the predicted edge information to guide the HGCN to inpaint the occluded areas of the images, so that the edges of the inpainted images are clearer and there is usually no deformation problem.

Fig. 8 shows the inpainting results in details from

different algorithms. Fig. 8(a) and (d) are the repaired images using the gated convolution module, Fig. 8 (b) and (e) are the repaired images using the attention mechanism module, and Fig. 8(c) and (f) are the

repaired image using the proposed model. From Fig. 8, it can be seen that the proposed algorithm improves the image restoration effect visually, especially the details.

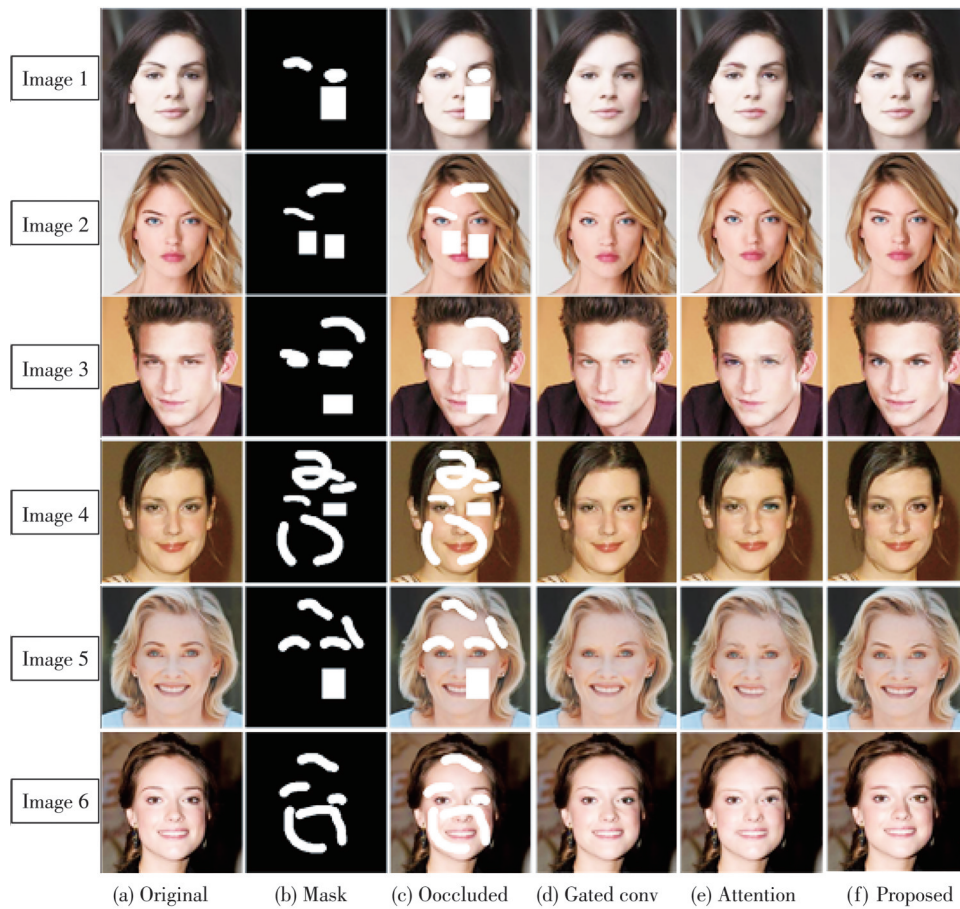


Fig. 7 Comparison of repair effects of different algorithms

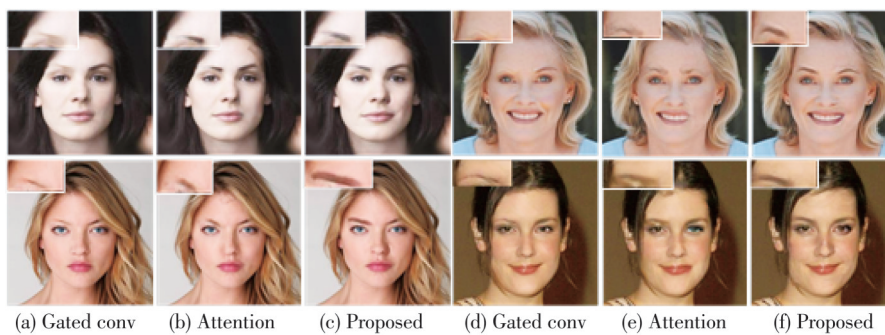


Fig. 8 Comparison of repair details of different algorithms

3.4 Quantitative analysis

In this part of the experiment, the three models were compared and analyzed on peak-signal-to-noise-ratio (PSNR), structural similarity (SSIM) and mean absolute (MAE). There were 200 images randomly selected as the test data, which were divided into two types: the images with small missing areas and the images with large missing areas according to the size of

the mask area, and the corresponding mask sizes were 10%–20% and 20%–30%, respectively.

Among the indicators, PSNR is used to measure the pixel value difference between images. The larger the PSNR value, the smaller the distortion and the better the image inpainting effect. SSIM is used to measure the similarity of two images. The closer the SSIM value is to 1, the better the inpainting effect of the model. MAE is to measure the loss of the mean absolute error. The

smaller the MAE value, the better the inpainting effect of the model. The model comparison results are shown in Table 2, Table 3, and Table 4, respectively. The optimal index values are represented in bold.

Table 2 PSNR comparison

Model	PSNR	
	Mask 10%–20%	Mask 20%–30%
Gated conv	28.578 38	27.052 66
Attention	27.919 96	26.073 68
Proposed	28.980 71	27.491 03

Table 3 SSIM comparison

Model	SSIM	
	Mask 10%–20%	Mask 20%–30%
Gated conv	0.967 522	0.948 372
Attention	0.966 700	0.933 172
Proposed	0.976 086	0.944 972

Table 4 MAE comparison

Model	MAE	
	Mask 10%–20%	Mask 20%–30%
Gated conv	0.029 206	0.031 694
Attention	0.030 27	0.034 226
Proposed	0.023 802	0.032 384

It can be concluded from the above three tables that the proposed method outperforms the other two models when the missing area occupies 10%–20% of the image area. When the missing area occupies 20%–30% of the image area, the proposed method also outperforms attention mechanism model on all three indicators, and outperforms gated convolution model on PSNR, and is close on SSIM and MAE.

4 Conclusions

In this study, a two-stage HGCN structure with inpainting effect from coarse to fine was proposed and combined with edge adversarial network to inpaint the face images with missing areas. The experimental results demonstrate that the proposed model has a better inpainting effect in image details, and the intuitive feeling of the inpainted image is also improved, while the inpainting effect can not significantly improved when the missing area is large. Although experiments have proved the effectiveness of this model, there is still a lot of room to improve, for example, our model ignores the symmetry information of human face, which will help to improve the inpainting result evidently.

Acknowledgement

This work was supported by Natural Science Foundation of Gansu Province (No.21JR11RA062); and University

Innovation Fund of Gansu Province (No. 2022A-047)

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

- [1] GUO J T. Research on face image inpainting and editing based on generative adversarial networks. Beijing: Beijing Jiaotong University, 2021: 1-6.
- [2] SHEN L. Research on image inpainting methods based on semantic perception deep model. Hefei: Hefei University of Technology, 2020: 1-9.
- [3] MU Q, XIA L, LI Z L, et al. Depth image hole inpainting method using curvature diffusion and edge reconstruction. Journal of Xi'an University of Science and Technology, 2021, 40(2): 369-376.
- [4] ZHAI D H, ZUO W J, DUAN W X, et al. Image inpainting algorithm based on double-cross curvature-driven diffusion model. Computer Applications, 2013, 33(12): 3536-3539.
- [5] LIU Y. Research and application on image inpainting based on texture synthesis and image segmentation based on fractal. Doctoral Dissertation. Changchun: Jilin University, 2010: 11-70.
- [6] CAI X X. Exemplar-based texture synthesis and its applications research. Xi'an: Xidian University, 2018.
- [7] SHEN Y F, SUN S L, Xu F S, et al. CT Image Reconstruction via nonlocal low-rank regularization and data-driven tight frame. Symmetry, 2021, 13(10): 1873-1873.
- [8] GAO C, LUO Y, WU H, et al. Data-driven image completion for complex objects. Signal Processing: Image Communication, 2017, 57(2): 21-32.
- [9] CHEN M J. Research on nuclear radiation contaminated image enhancement based on total variation and sparsity representation. Mianyang: Southwest University of Science and Technology, 2020: 61-85.
- [10] DU Y. Learning sparse and deep representations for image restoration. Doctoral Dissertation. Guangzhou: South China University of Technology, 2019.
- [11] ZHAO S X, MEN S Y, YUAN L. MobileNet network optimization based on convolutional block attention module. Journal of Measurement Science and Instrumentation, 2021, 13(2): 225-234.
- [12] HOU X D, LIU H R, LIU H P. High-resolution damaged images restoration based on convolutional auto-encoder generative adversarial network. Journal of Image and Graphics, 2021, 27(5): 1645-1656.
- [13] LIU P W, GAO Y, QIN P Y, et al. Generative adversarial network medical mri image super-resolution reconstruction based on multiscale residuals. Journal of North China University (Natural Science Edition), 2021, 42(5): 449-459.
- [14] WANG J Y. Image generation based on generative adversarial

- networks. Hefei: University of Science and Technology of China, 2021.
- [15] DEEPAK P, PHILIPP K, JEFF D, et al. Context encoders: Feature learning by inpainting//IEEE Conference on Computer Vision and Pattern Recognition, June 26-July, 1, 2016, Las Vegas, USA. Piscataway, N. J.: IEEE, 2016: 2536-2544.
- [16] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion. ACM Transactions on Graphics, 2017, 36(4): 1-14.
- [17] YU J, LIN Z, YANG J, et al. Free-form image inpainting with gated convolution//IEEE/CVF International Conference on Computer Vision, October 27-November 1, 2019, Seoul, South Korea. Piscataway, N. J.: IEEE, 2019: 4471-4480.
- [18] NAZERI K, NG E, JOSEPH T, et al. Edgeconnect: Generative image inpainting with adversarial edge learning. 2019-01-01[2021-01-01]. 2019: 1901-1917. DOI: 10.48550/arXiv.1901.00212.
- [19] GUO X, YANG H, HUANG D. Image inpainting via conditional texture and structure dual generation//IEEE/CVF International Conference on Computer Vision, October 10-17, 2021, Montreal, CA. Piscataway, N. J.: IEEE, 2021: 14134-14143.
- [20] LIU H, JIANG B, SONG Y, et al. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations//European Conference on Computer Vision, August 23-28, 2020, Glasgow, UK. Heidelberg: Springer, 2020: 725-741.
- [21] WAN G L. Research on the restoration algorithm of cultural relics based on gated convolution and coherent semantic attention mechanism. Doctoral Dissertation. Yinchuan: Ningxia University, 2021.
- [22] GAO J, HUO Z. Algorithm for image inpainting in generative adversarial networks based on gated convolution. Journal of Xidian University, 2022, 49(1): 216-224.
- [23] SU M J, WANG B, LIU B Y. Expression image synthesis using GAN by introducing self-attention mechanism and spectral normalization. Intelligent Computers and Applications, 2022, 12(4): 121-125.
- [24] YANG Y, CAO Z, QI Y. Image inpainting algorithm based on multi-angle constraint and spectral normalization. Computer Applications and Software, 2021, 38(8): 233-239.
- [25] WANG H Y, LI H Y, GAO X J. Research on image restoration method based on structure embedding. Computer Engineering and Applications, 2021, 57(22): 241-246.
- [26] LIU A, LIU X, FAN J, et al. Perceptual-sensitive gan for generating adversarial patches//The 34th AAAI Conference on Artificial Intelligence, February 7-12, 2020, New York, USA. Washington: AAAI Press, 2020: 1028-1035.
- [27] ZHANG X F. Research on Image inpainting algorithm based on attention mechanism. Doctoral Dissertation. Nanjing: Nanjing University of Posts and Telecommunications, 2021.
- [28] CHEN Z N, ZHANG H Y, ZENG N Y, et al. Attention mechanism embedded multi-scale restoration method for blurred image. Journal of Image and Graphics, 2022, 27(5): 1682-1696.
- [29] WANG F, CHENG J, LIU W, et al. Additive margin softmax for face verification. IEEE Signal Processing Letters, 2022, 25(7): 926-930.
- [30] WAN L, TONG X, SHENG M W, et al. Review of image classification based on softmax classifier in deep learning. Navigation and Control, 2019, 18(6): 1-9.

边缘对抗结合层次门控卷积的人脸修复研究

翟凤文*, 周 钊, 孙芳林, 金 静

兰州交通大学 电子与信息工程学院, 甘肃 兰州 730070

摘要: 针对目前人脸缺损图像修复存在边缘模糊和失真等问题, 提出了一种两阶段层次门控卷积网络(Hierarchical gated convolutional network, HGCN), 并将其与边缘对抗网络相结合用于人脸图像修复。首先, 采用边缘对抗网络生成边缘图像。其次, 将边缘图、掩模和缺损图像作为输入, 训练HGCN的GAN模型以修复缺损人脸图像。HGCN网络采用门控卷积取代传统卷积, 并引入了扩张卷积, 网络的主体由粗修复模块和精修复模块组成。在粗修复模块中, 编码器和解码器网络结构用于粗修复; 在精细修复模块中, 引入注意力机制来增强特征提取能力, 进一步细化修复结果。实验使用Celeba-HQ数据集和NVIDIA不规则掩码数据集作为训练数据集, 采用门控卷积网络和注意力机制网络作为实验对比模型, PSNR、SSIM和MAE作为实验评估指标。实验结果表明, 对于缺损区域小于20%的人脸图像, 所提出的网络在上述三个指标上优于两种比较网络, 而对于缺失区域大于20%的图像, 所提出的网络与两种比较方法性能指标接近。在视觉效果方面, 所提出的方法在细节上也优于两种对比网络。因而, 所提出的网络可以明显提高图像修复效果, 尤其是对图像细节的修复效果。

关键词: 深度学习; 人脸修复; 层次门控卷积网络; 边缘生成; 生成对抗网络

引用格式: ZHAI Fengwen, ZHOU Zhao, SUN Fanglin, et al. Face inpainting based on edge confrontation combined with hierarchical gated convolution. Journal of Measurement Science and Instrumentation, 2024, 15(1): 33-42.