Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 6948293, 14 pages
https://doi.org/10.1155/2021/6948293

Research Article

A Face Occlusion Removal and Privacy Protection Method for IoT Devices Based on Generative Adversarial Networks

Wenqiu Zhu, Xiaoyi Wang, Yuezhong Wu, and Guang Zou

School of Computer Science, Hunan University of Technology, Zhuzhou 412007, China

Correspondence should be addressed to Yuezhong Wu; wuyuezhong@hut.edu.cn

Received 17 April 2021; Revised 21 May 2021; Accepted 17 June 2021; Published 1 July 2021

Academic Editor: Zhuojun Duan

The device group based on the Internet of Things (IoT) has been used in face recognition in real life, so it is more necessary to discuss the current data security issues and social hot issues. The Internet of Things device combines edge conditions and many recognizers to generative adversarial networks. On the premise of meeting the needs of partial occlusion of users, face recovery is completed through information reorganization. CelebA training set is used to simulate face occlusion, and the model is trained and tested. The results show that the method can recover the complete image of the protection for the facial privacy of specific people. At the same time, the IoT device using this method ensures that the face information is not easy to have tampered with when attacked.

1. Introduction

IoT is "Internet of Things." They are built based on the expansion of the Internet, and with the Internet as the core, the client extends to any object and between objects. These devices can sense each other. By connecting various information sensing devices with the network to form a network, the data can be interconnected and shared anytime and anywhere [1, 2]. The IoT has been applied to people's daily life. In public places, face recognition devices of IoT extract to face data and prompt other devices to get information. With the help of blockchain and cloud computing [3–5], information sharing and management between IoT devices become more convenient. Also, the application of edge computing [6–8] provides support for the rapid popularization of the IoT, which has more advantages than the traditional communication between home and public equipment sensors.

Universal recognition equipment will be applied to image recognition technology. Devices in the IoT can perceive the outside world through image recognition, which can make data collection [9, 10] more efficient, and provide users with more humanized data displays and life suggestions. At the same time, industrial IoT [11–14] also requires recognition equipment with perceptual capabilities. The equipment in each production link handles different functions, and they interact with each other so that the production process and production plans become more flexible and reliable.

With the development of image processing technology, the predecessors proposed pixel-by-pixel filling [15], finding matching blocks [16], using image sets on the Internet to fill similar blocks [17], and suggestions based on matching block technology, such as block completion algorithm and statistical block probability repair method [18, 19]. These methods enable the device to penetrate the limited occlusion of the face when recognizing the face and accurately obtain the user's identity information.

However, in the case of facial defects, the user must remove the occlusion for recognition. This may have potential adverse effects on the user's psychology. In addition, identification devices will also encounter some personal privacy issues related to data collection [20, 21]. For example, in the face of flaws, users still have to perform facial authentication. Since the identification device is part of the IoT, data collection is unavoidable, and sharing with other devices will involve information security [22–24] related issues.

Machine learning [25–28] proposed a solution to improve the recognition of human faces through machine learning, so on improve the recognition ability of face

recognition equipment through fitting and classification. Later, the generative adversarial network developed based on deep learning [29, 30], through the adversarial network to improve the level of generated images and the level of image recognition, so on preventing facial spoof attacks, and at the same time further improve the restoration of the face under concealment conditions for recognizing the face to determine the identity of the user.

Therefore, an end-to-end workflow from edge recovery to face recovery is proposed, and a deep learning network based on edge conditions and multiple discriminators is proposed to judge whether the generated pixels are filled according to the required edge structure. This overcomes these challenges. Fill in the corresponding pixel information according to the integration of different levels of complete image styles and features. A complete module based on self-care mechanism is proposed and applied to Image-inpaint network. So, the face recognition device can restore the entire face from a multifeature level.

The structure of this article is as follows. The second part introduces some technical details of the implementation. The third part introduces the end-to-end deep learning network structure based on edge conditions, including multiple discriminators and self-attention mechanism. After reviewing the literature, the fourth part will introduce in detail the application of multiple local dividers in the Image-inpaint network. The fifth part includes the discussion and conclusion of our survey results.

2. Related Work

Under the influence of generative adversarial network technology, Pathak et al. [31] propose an Encoder-Decoder pipeline model, which uses an unsupervised visual feature learning algorithm driven by context pixel prediction to use surrounding image information. To infer the missing location, Iizuka et al. [32] propose a global discriminator and a local discriminator based on the Context-Encoder to promote the learning of the missing parts and use local convolution to increase the attention of partial blocks by the network to enhance the details of specific parts. The situational attention network proposed by Yu et al. [33] and the self-attention generation adversarial network proposed by Zhang et al. [34] solve the problem that the structure of the surrounding area is distorted, or the texture is not consistent with the fuzzy texture. They can not only synthesize new image structures but also make full use of the surrounding image features as a reference. Among them, the self-attention mechanism proposed to reference [34] can also be weighted according to the importance of features to correlate important information on each other. After that, Yu et al. [35] also combined the contextual attention mechanism and the proposed gat convolution. The discriminator is no longer a combination of local and global discriminators but uses SN-Batch GAN; on the premise of meeting the Lipschitz constraint, the information of the weight matrix of the discriminator is saved to the maximum extent so that the training process is more stable. Kun et al. [36] proposed a face completion algorithm based on a conditional generation adversarial network. The algorithm generates faces that meet the conditional features, but it is still a relatively simple information extraction based on Encoder-Decoder. Moreover, multiple convolutional blocks are not used for sufficiently deep feature extraction. Xie et al. [37] proposed an image restoration method based on a learnable two-way attention map, which can deal with irregular hole repair, and proposed to merge forward and backward attention maps into a learnable two-way attention map. To further improve the visual quality of the image, but because there is no constraint on edge information, the details of specific parts of the portrait are still not ideal. In EdgeConnect [38], Nazeri et al. proposed the concept of " lining first, then color" through image restoration based on edge conditions. The repaired edge information is obtained through the edge generator, and then based on the obtained edge repair image as a condition, the incomplete image data is spliced and input into the image repair network, thereby using the edge information to restore the image. Get images of the image completion network. Its essence is to divide the steps of image restoration of high-frequency information and low-frequency information. Yet, only the discriminator that uses the network has a global identity. Compared with the local recognition network and the facial feature recognition network, the repair details are compared with the image generation, which is inferior to the method of multiple discriminators.

The spectral normalization proposed by Miyato et al. [39] reduces the calculation amount of network normalization and makes the calculation of the discriminator more stable. In addition, the reason for using multiple discriminators [32, 33] as a supervisory network is that multiple discriminators have been proven many times in practical applications to form multilevel constraints on the generated results of more aspects. And let the generator produce better results. Using partial convolution [40], there are only connections to the local area, and the receiving field adopts a connected method, and the interval between the receiving fields adopts a local connection and a convolutional connection. Compared with full convolution, this method will introduce additional parameters in multiples, but it has stronger flexibility and expressive power. Compared with a local connection, it can control the number of parameters and proposed a new convolution to replace the general convolution, which will prevent the training results from generating chessboard artifacts.

For face recognition equipment, the combination of convolutional computer vision and computer image processing can make the face recognition equipment perform better under specific facial features, detail recovery, and state recovery.

To solve the problem that the edge completion method only pays attention to the integrity of the image and ignores the visual connectivity of the complete part and the facial features, based on EdgeConnect, puts forward the following ideas:

(a) The method of a discriminator for face parts (eyes, nose, and mouth) is proposed. After the complementary edge image and damaged image are processed by

the image completion network, the eyes, nose, and mouth of the generated complementary face image are discriminated. So the facial features of the face image are more consistent with the semantic rationality

(b) Propose a gated convolution block based on edge condition to enhance the difference between regions. At the same time, the self-attention mechanism is used to automatically enhance the difference between occluded and nonoccluded areas. The information of the input data is enhanced according to different feature levels, to distinguish the occluded area and the nonoccluded area more accurately

(c) Propose to use a local discriminator. When the completed image is processed by the image completion network, the completed image will have better visual connectivity as a whole

2.1. Deep Learning. Deep learning-based recommendation methods can incorporate multisource heterogeneous data for the recommendation, including explicit or implicit feedback data from users, user portrait and project content data, and user-generated content. Deep learning methods use multisource heterogeneous data as input and use an end-to-end model to automatically train prediction models, which can effectively integrate multisource heterogeneous data into the recommendation system, thereby alleviating the data sparseness and cold start in traditional recommendation system problems and improving the ability of the recommendation system. The application of deep learning to corpus mining is a research hotspot. After 2006, with the publication of Hinton and Salakhutdinov [41], it was wildly sought after by scholars in the artificial intelligence world. This model is based on a neural network model, but it is more complex than a simple neural model, and the problems it deals with are more complex and diverse. Deep learning methods have been successfully used in many applications in the computer field, including speech recognition, speech search, natural language understanding, information retrieval, and robotics. Mokris and Skovajsova [42] applied the neural network model to have a high degree of relevance. It retrieved Slovak-related documents, processed keyword parts of speech, and greatly improved accuracy and recall. Based on the highly nonlinear characteristics of neural network algorithms, using BP network to optimize the weight of each parameter in the entire neural network, constantly revising the weights, Xu et al. [27] constructed a personalized behaviour based on users. Latreche and Guezouli [43] use the correlation characteristics of neighbor nodes in the neural network to combine all documents into a neural network and retrieves the most relevant document according to Query.

The method (Figure 1) consists of two parts, namely, the Edge-inpaint network and the Image-inpaint network. Among them, the Edge-inpaint network is responsible for repairing the edges of the defective face image, and the Image-inpaint network is responsible for completing the face image based on the condition of the completion of the edges.

Before applying the model, the network is trained to generate the model. The data processing in the training process is divided into two steps: First, take the unmasked edge map as the target, and enter the masked, masked Canny Edge maps and masked grayscale images, through training the Edge-inpaint network model, enable the Edge-inpaint network to generate a predicted completion edge map; second, use the prediction completion edge map output by the Edge-inpaint network, combined with the incomplete color face image is input to the Image-inpaint network, and the predicted repaired face image is output. The two networks form an end-to-end solution so that the identification device based on the IoT can ensure that the user can still accurately obtain the correct information of the user when the user is covering his specific part.

3. Networks

3.1. Edge Completion Generative Adversarial Network. The Edge-inpaint network (Figure 2) is composed of an edge completion generation network (edge generator, hereinafter referred to as G1) and an edge completion discriminator network (edge discriminator, hereinafter referred to as D1). Among them, G1 network generates edge completion image, inputs masked gray image, masked edge image, and mask, takes real edge image as label, generates a completion edge image after G1 operation, and then calculates adversarial loss and feature matching loss [44] through D1, and this loss value is used to backpropagate the network associated with it. The purpose of G1 is to minimize the gap between the generated edge image and the real edge image to improve the quality of the image generated by the generator, while D1 is to expand this gap as much as possible, thus making the progress of G1 more difficult; of course, the result will be better.

G1 is composed of 3 convolutional layers, 8 residual blocks, and 3 deconvolutional layers cascaded. Among them, the convolution kernel of the convolution layer is composed of $7*7$, $4*4$, and $4*4$, and the deconvolution layer is composed of convolution kernel sizes of $4*4$, $4*4$, and $7*7$. Convolution both the layer and the deconvolution layer have spectral normalization processing and setting the ReLU activation function after the convolution operation. The first convolution of the convolution layer and the last convolution of the deconvolution layer are done separately reflective filling treatment.

D1 takes the real edge face image as the target (label) and fights against G1. When G1's ability becomes stronger and stronger, D1 can also improve the complementary edge image generated by G1 with reasonable parameter settings. The Canny edge detector is used to extract the edge features of the image as the expected positive samples learned by the discriminator, and the negative samples generated by the network G1 that do not meet the experimental expectations and Canny edge features are merged to improve D1 with its distinguishing ability and supervise G1 to generate an image with edge information that is more in line with the original image.

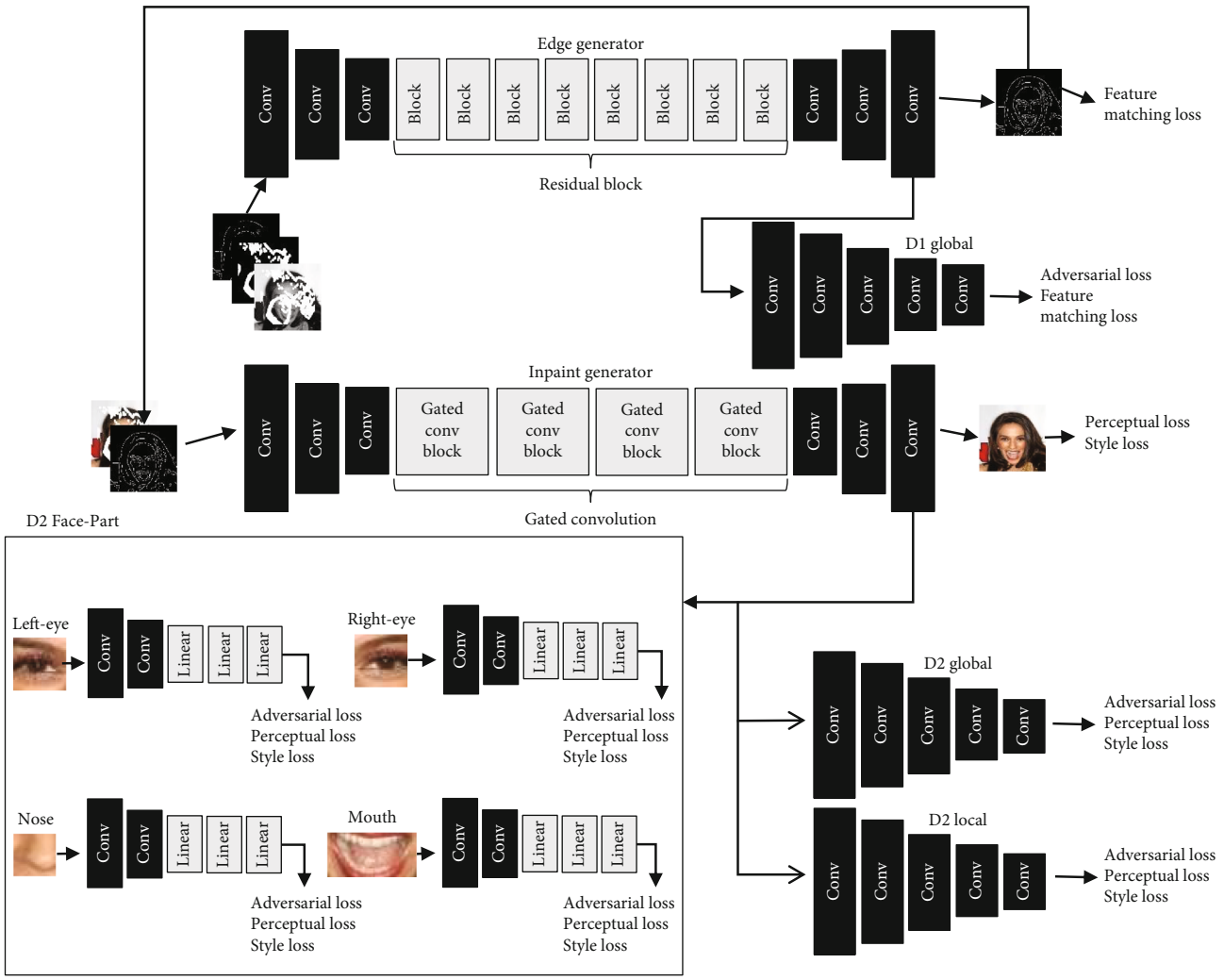


Figure 1: Structure of network, which has edge-GAN, inpaint-GAN, and multiple discriminators.

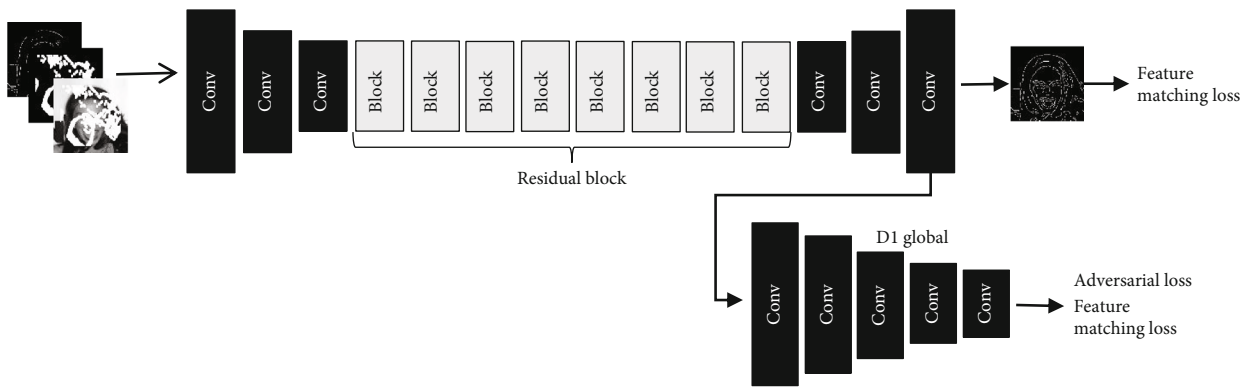


Figure 2: Structure of edge-GAN, which has edge-generator and edge-global discriminator.

The Edge-inpaint network allows the recognition device to reconstruct the structural information of the face from the occluded part of the face, thereby providing a basis for the subsequent feature and texture completion. Through spectral normalize processing of convolution, residual block [45] and deconvolution, and fusion of different information sources, deep texture features are extracted.

3.2. Image Completion Based on Self-Attention Mechanism

3.2.1. Composition of Image-Inpaint Network. The image completion network (Figure 3) is composed of an image completion generation network (Image-inpaint generator, referred to as G2 hereinafter) and an image completion discriminator network (Image-inpaint discriminator, hereinafter

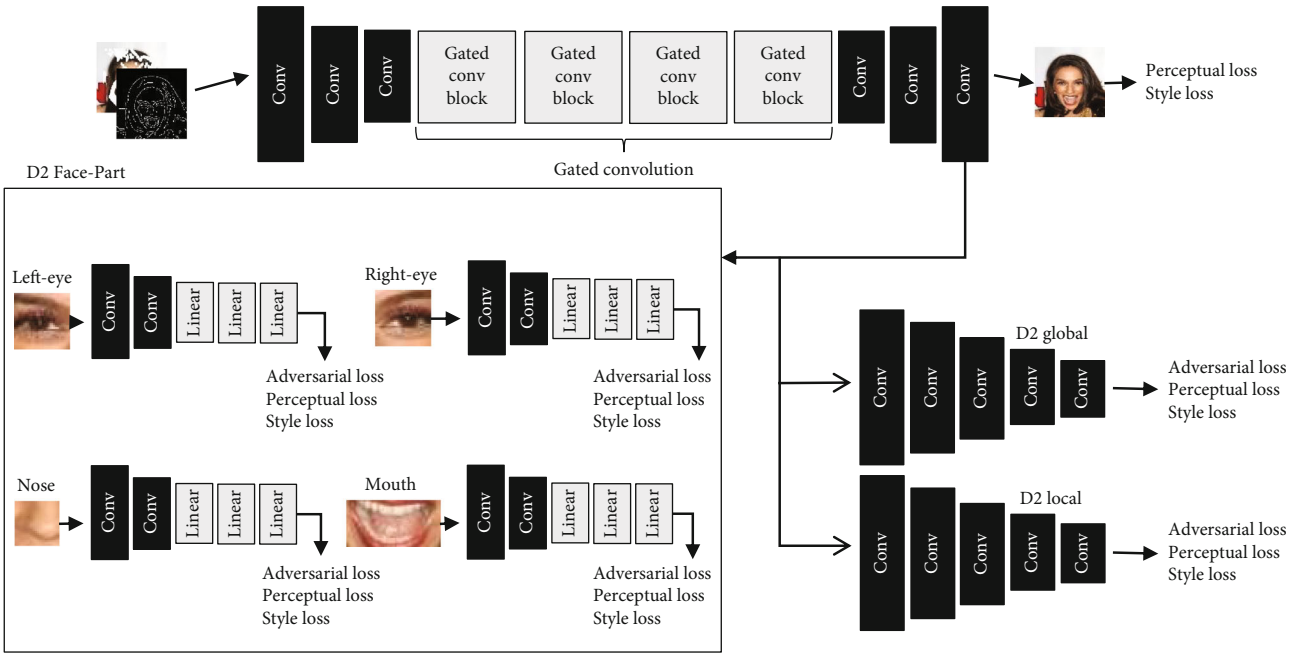


Figure 3: Structure of inpaint-GAN, which has inpaint-generator and multiple discriminators in inpaint-GAN, and multiple discriminator includes global discriminator (D2 global), local discriminator (D2 local), and face-part discriminator (D2 Face-Part).

referred to as D2). The role of the G2 network is to generate a complementary image, input the complementary edge map and the incomplete face image, and use the real image as a label and cascade the input data; after the G2 operation, generate the completed image and then pass the D2. Each discriminating network calculates adversarial loss, perceptual loss, and style loss [44]. The loss of the image completion network is the sum of the counter losses of the D2 Face-Part network, and this loss value is used to backpropagate the network associated with it. The purpose of G2 is to allow G2 to dynamically learn the parameters through training to effectively distinguish between the effective area and the mask area and reduce the adverse effect of the mask on the image completion so that the color and structure of the image completion are more reasonable and minimized. Generate the gap between the completed image and the real image to improve the quality of the image generated by the generator. And D2 is to widen this gap as much as possible, so that G2 and D2 identify the network combination to fight and promote the progress of G2.

G2 is composed of 3 convolutional layers, 4 proposed gated convolutions, and 3 deconvolutional layers in cascade. Among them, the convolution kernel of the convolution layer is composed of 7∗7, 4∗4, and 4∗4, and the deconvolution layer is composed of the convolution kernel size of 4∗4, 4∗4, and 7∗7. The convolution layer and the deconvolution layers all have spectral normalization processing and setting of the LeakyReLU activation function after the convolution operation. The first convolution of the convolution layer and the last convolution of the deconvolution layer are, respectively, filled with reflection.

D2 takes the real face image as the target and fights against G2. When G2's ability is getting stronger and stronger, D2 can improve its ability to discriminate the complementary image generated by G2 and use G2 to generate it with reasonable parameter settings. The edge feature of the mask completes the image. As the expected positive samples learned by the discriminator, the negative samples generated by the network G2 that do not meet the experimental expectations and the mask edge features are merged to improve the discrimination ability of D2 and supervise G2 to generate images that are more in line with the original image.

3.2.2. Design and Implementation of Self-Attention Mechanism. Ordinary convolution has limitations in completing the image under any mask. It extracts local features in a sliding manner, and the pixels under the sliding window are all valid by default. But for image completion, when the window contains the boundary of the mask, its invalid pixels and effective pixels will be processed by the convolution window, which will cause the information to be blurred, and the sides of the complement part do not match the actual results.

The gated convolution module is used to automatically learn the soft occlusion mechanism from the data through self-attention, dynamically identify the effective pixel position in the image, and process the transition between the masked area and the unmasked area. The proposed gated convolution not only can retain features at long distances that the residual block has but also has the ability of the gated convolution itself to enhance the distinction between features on both sides of the edge. The following formula (1) is the operation of gated convolution [33]

$$\begin{cases} \text{Gating}_{y,x} = \sum\sum W_g \cdot I, \\ \text{Feature}_{y,x} = \sum\sum W_f \cdot I, \\ O_{y,x} = \phi\left(\text{Feature}_{y,x}\right) \odot \phi\left(\text{Gating}_{y,x}\right). \end{cases} \tag{1}$$

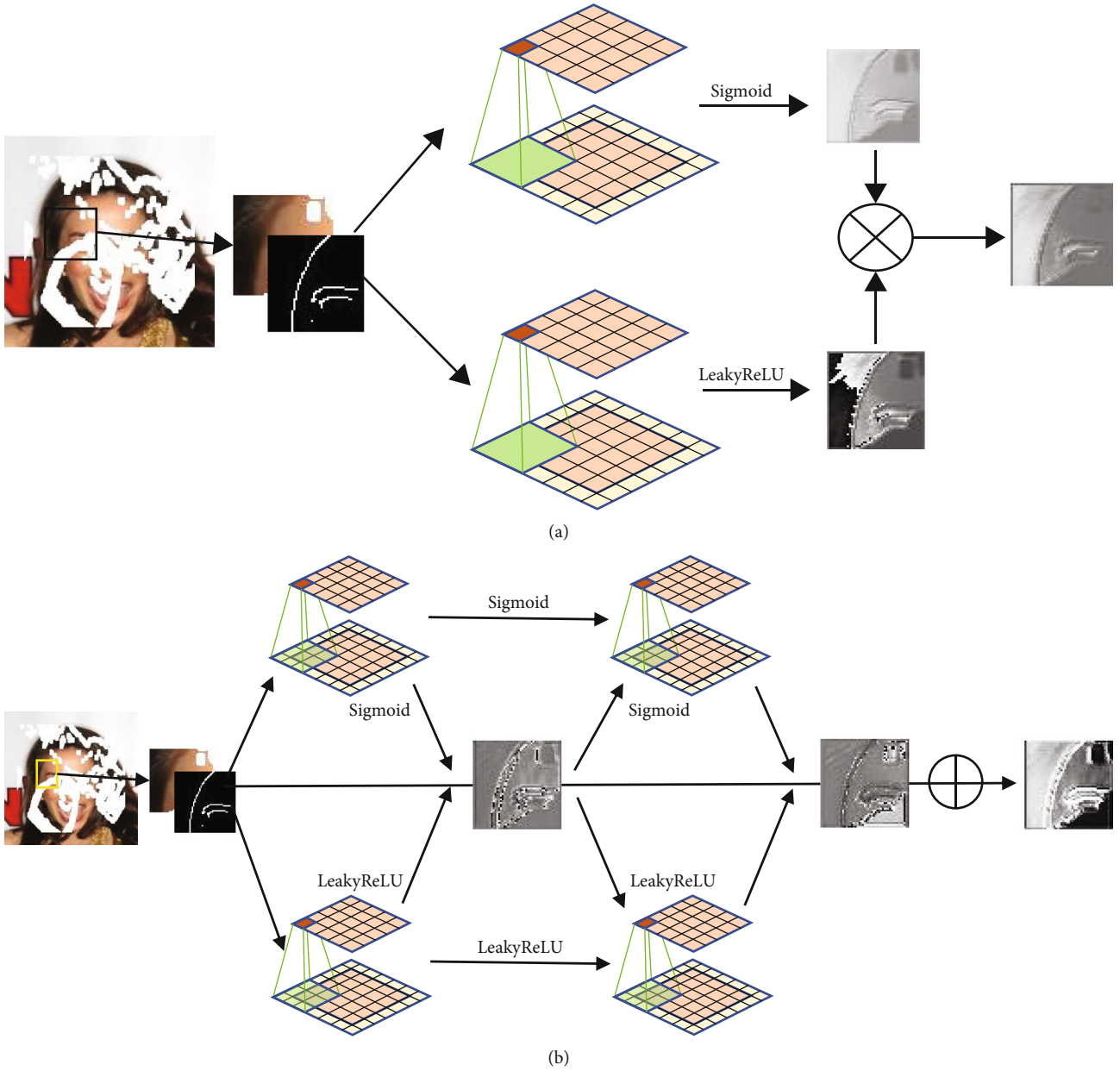


Figure 4: Results produced under different gated convolutions: (a) normal gated convolution; (b) proposed gated convolution.

Among them, y, x is the center point of the current sliding area; W_g is the convolution filter that selects mask and non-mask space; W_f is the convolution filter to distinguish between mask and non mask, and I/O is input convolution filter and output convolution filter, respectively. Gating convolution and sigmoid (Figure 4(a)) [46] activation function realize dynamic feature selection, that is, selecting mask coverage space and normal space; feature convolution and LeakyReLU activation function realize feature extraction, that is, select part of the transition map of the masked and non-masked areas and then use the dot product of gating and feature to more effectively select useful information. The proposed gated convolution has stronger pixel selectivity so that the convolution can accurately describe local features even with a larger range of pixel deletions.

At the same time, the general gated convolution still has a small amount of boundary blur on both sides of the mask boundary, which will cause the problem of invalid pixels as valid pixels when the window of the gated convolution is sliding, which will cause the concealed information will be regarded as part of the face itself rather than blocked, making the device unable to effectively restore the original information of the face. On this basis, different features generated by different subgated convolution are added to the input to get clearer feature differences on both sides of the mask boundary (Figure 4(b)), to distinguish the differences on both

sides of the edge more accurately. The boundary pixels between the region and the normal region affect the visual connectivity of the whole image. If the defective part of the mask edge is close to the normal part of the pixels, gated convolution can easily confuse the two, and different types of features are strengthened by feature addition to achieving more effectiveness on both sides of the mask edge. The feature distinction of, so that the image completion network is better targeted at the completion part, and the part of the human face itself and the covered part are logically distinguished effectively.

4. Discriminator Network for Partial Completion

4.1. Local Discriminator Network Combining Structure and Texture. Aiming at the problem that the Image-inpaint network only has a global discriminator network and the partial restoration is not ideal, a local discriminator network is proposed.

The local discriminator network proposed in this paper is a part of D2 Local in Image-inpaint network. Its network structure is the same as the D2 Global, it consists of five convolution blocks, and each convolution block contains a layer of Convolution, Spectral Normalize, and LeakyReLU (the last convolution block has no leakyrelu activation layer).

The completion part and the real part of the corresponding position are input into the global discrimination network to generate two 15×15 matrices, and then, these matrices are input into the adversarial network. The process is calculated by

$$L_{\text{adv},l} = E_{\left(I_{gt,l}, C_{\text{comp},l}\right)}\left[\ln D_2\left(I_{gt,l}, C_{\text{comp},l}\right)\right] + E_{\text{comp},l} \ln \cdot \left[1 - D_2\left(I_{\text{pred},l}, C_{\text{comp},l}\right)\right]. \tag{2}$$

Among them, $I_{\text{pred},l}$ is the completion part; $I_{gt,l}$ is the real part of the corresponding position; $C_{\text{comp},l}$ is the local complement edge map; this formula is the loss function of the local adversarial network. It is responsible for identifying the authenticity of the complete part so that the complete result will not deviate from the authenticity.

The general local discriminator network is based on authenticity, but because it only judges the authenticity of the complete part itself, the style of the generated part is weak, and the generated part affects the visual connectivity of the whole image. Moreover, in the reconstruction of high-level feature levels, factors such as color, texture, and exact shape of the face are not taken into consideration. Therefore, it is proposed to use style loss and perceptual loss for additional constraints. The recognition device can also restore the occluded part under the condition of conforming to the subject characteristics of the face. Style loss is

$$L_{\text{style},l} = L_{\text{style}}^{\phi,j}(\hat{y}, y) = \left\| G_j^{\phi}(\hat{y}) - G_j^{\phi}(y) \right\|_F^2, \tag{3}$$

where G_j^{ϕ} is the activation map of the j-th layer of the pretrained network and j is a set of integers from 1 to 5, which corresponds to the activation maps of the relu1_1, relu2_1, relu3_1, relu4_1, and relu5_1 layers of the pretrained VGG-19 [47] network. These activation maps are also used to calculate the style loss to measure the difference between the covariances of the activation maps. The Euclidean distance of each image feature is used to measure the degree of dissimilarity between perception and the real part.

Although the style loss corrects the texture and pixel completion error to a certain extent, it does not well preserve the shape and structure of the image completion part. To solve the loss of style, only the texture and color information are retained, but the shape and structure information is not effectively retained. The perceptual loss is proposed to restrict the structure and shape of the generated result. Perception loss as

$$L_{\text{perc},l} = E\left[\sum_i \frac{1}{N_i} \left\| \Phi_i\left(I_{gt}\right) - \Phi_i\left(I_{\text{pred}}\right) \right\|_1\right], \tag{4}$$

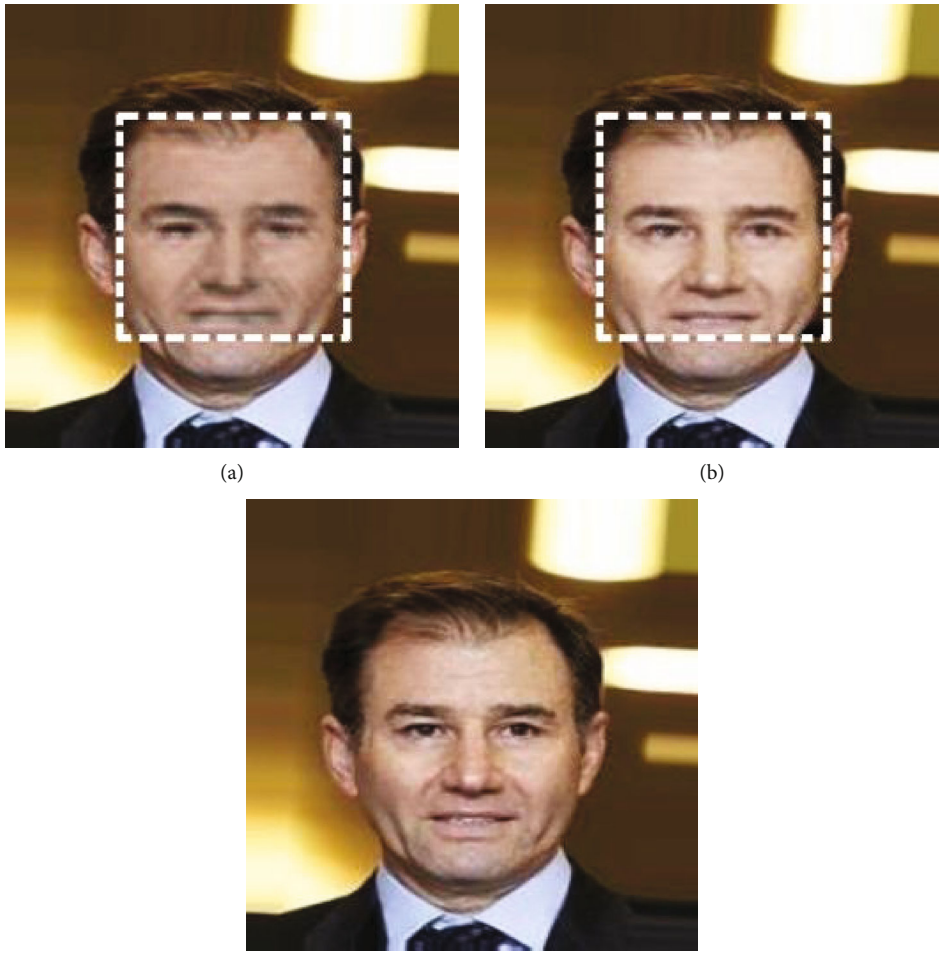
where Φ_i is the activation map of the i-th layer of the pretraining network and i is the set of integers from 1 to 4, which corresponds to the activation maps of the relu2_2, relu3_4, relu4_4, and relu5_2 layers of the pretrained VGG-19 [47] network. I_{gt} and I_{pred} are the real image and the generated image, respectively. The structure and shape features are extracted from each activation graph, and the Euclidean distance between activation is calculated to promote the reconstruction of high-level information.

Add formulas (2), (3), and (4) to obtain the total loss formula (5) of the local discriminator network and the Image-inpaint network

$$L_{\text{local}} = L_{\text{adv},l} + L_{\text{style},l} + L_{\text{perc},l}. \tag{5}$$

The main function of the local discriminator network is to measure some blocks generated by the image and obtain the combined losses to ensure the semantic rationality of the complement.

Figure 5(a) is the result obtained when only formula (2) is used as the loss function; Figure 5(b) is the result obtained when formula (5) is used as the loss function, and Figure 5(c) is the original image. The processed portrait is the information after the occluded part is restored. The restored part should not only pay attention to its authenticity to the whole portrait but also consider its authenticity. Therefore, the function of formulas (3) and (4) is to pay attention to the generic structure and style of the restored part on the premise that the global discriminator network also pays attention to the generation structure and style of the generation part. Let the image completion network be subject to more restrictions in training, just as human beings learn to pay attention to more information when considering a problem.



(a) (b) (c)

Figure 5: Effects under different losses: (a) adversarial loss; (b) combined losses; (c) original.

4.2. Discriminator Network Based on Face Local Position Constraint. Since the general local discriminator network is essentially the same as the global discriminator network, it only focuses on integrity, and it is difficult for the local discriminator network to describe the texture of the face image in detail when the face is completed. Therefore, a special local discriminator network is proposed to deal with the generation details of human face parts, so that the completed part and the whole image meet the visual connectivity.

The face-part discriminator network (D2 Face-Part) (Figure 6) is a special partial discriminator network. The face-part discriminator network is composed of four subnetworks. These four subnetworks are the left-eye discriminator network and the right-eye discriminator network, eye discriminator network, nose discriminator network, and mouth discriminator network. It targets the key parts of the human face, namely, the left eye, right eye, mouth, and nose. The four subnetworks of D2 Face-Part will extract the left eye, right eye, nose, and mouth and send them to the face recognition network. The network will be sent to the corresponding four networks, and finally, four scores will be generated. Calculate the respective adversarial losses, where the adversarial loss is 1 as true and 0 as false. After these scores are calculated for the adversarial loss, the four adversarial loss values are added and averaged. The four values of the adversarial loss are added and averaged to obtain

$$L_{\mathrm{adv},fp} = \frac{1}{4}\sum_{i=1}^{4}\left(E_{gt,i}\left[\ln D_2\left(I_{gt,i}\right)\right] + E_{I_{\mathrm{pred},i}}\left[1 - \ln D_2\left(I_{\mathrm{pred},i}\right)\right]\right). \quad (6)$$

Formula (6) is the loss function of D2 Face-Part and G2 adversarial, where $I_{gt,i}$ is the i-th real Face-Part and $I_{\mathrm{pred},i}$ is the i-th Face-Part of the completed image. The purpose of this function is to hope that G2 (image completion generation network) will pay more attention to the reliability of the generation of the facial "features" during training and to pay more attention to its effects on the microlevel of the face.

Same as the improved local network, to overcome the drawbacks of only paying attention to the true and false of the image itself, which is brought about by the counter loss, and not paying attention to its texture and structure. It is proposed to use the style loss function and the perceptual loss function for the discriminator of each part of the face so that the generated part makes the whole image have better visual connectivity. Formula (7) and formula (8) are the style loss function and the perceptual loss function of the human face, respectively.

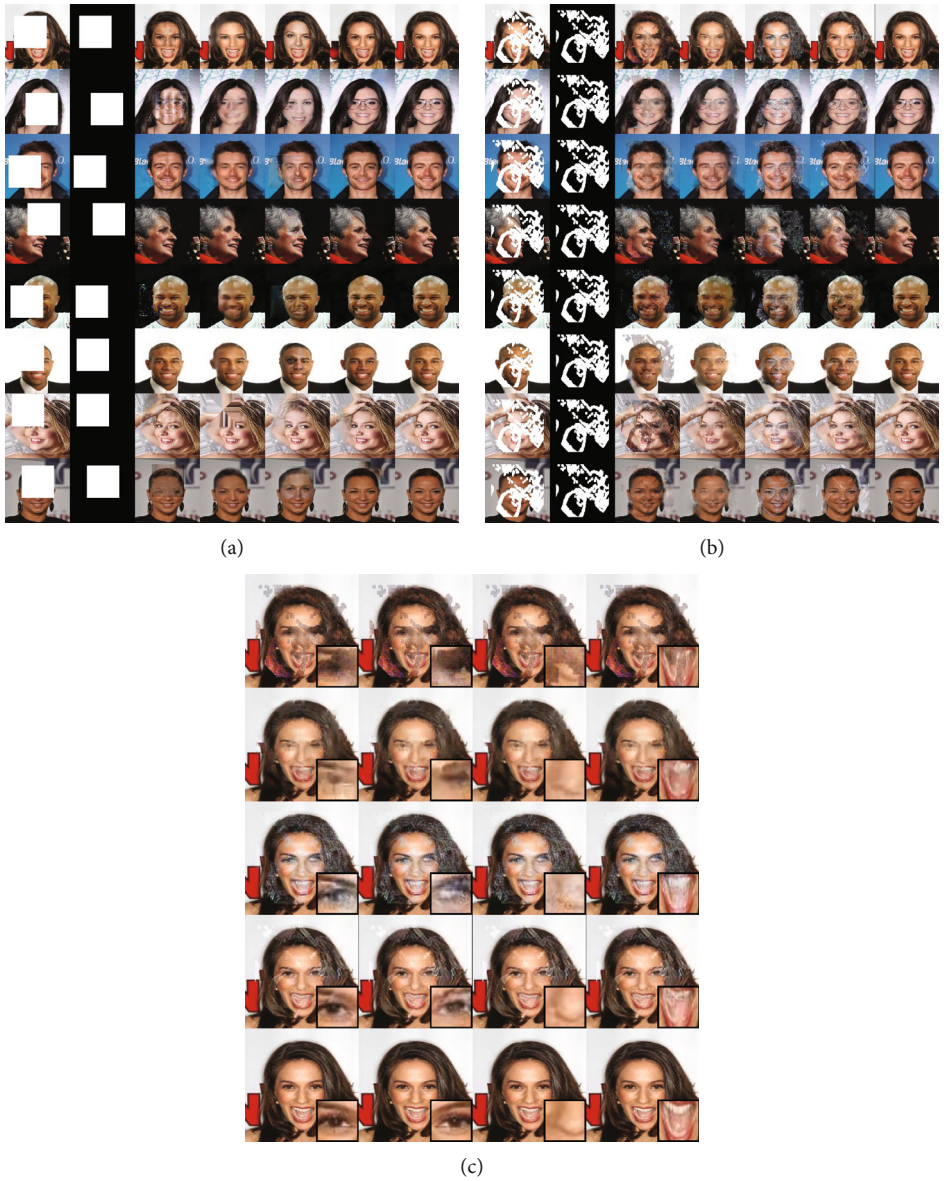


FIGURE 6: This is the result of each method under different mask conditions: (a) restoration results of different techniques under block mask; (b) restoration results of different techniques under random mask; (c) restoration Face-Part of different techniques under random mask.

$$L_{\text{style},fp} = L_{\text{style}}^{\phi,j}(\hat{y}, y) = \frac{1}{4}\sum_{i=1}^{4}\left\|G_j^{\phi}(\hat{y}_i) - G_j^{\phi}(y_i)\right\|_F^2, \quad (7)$$

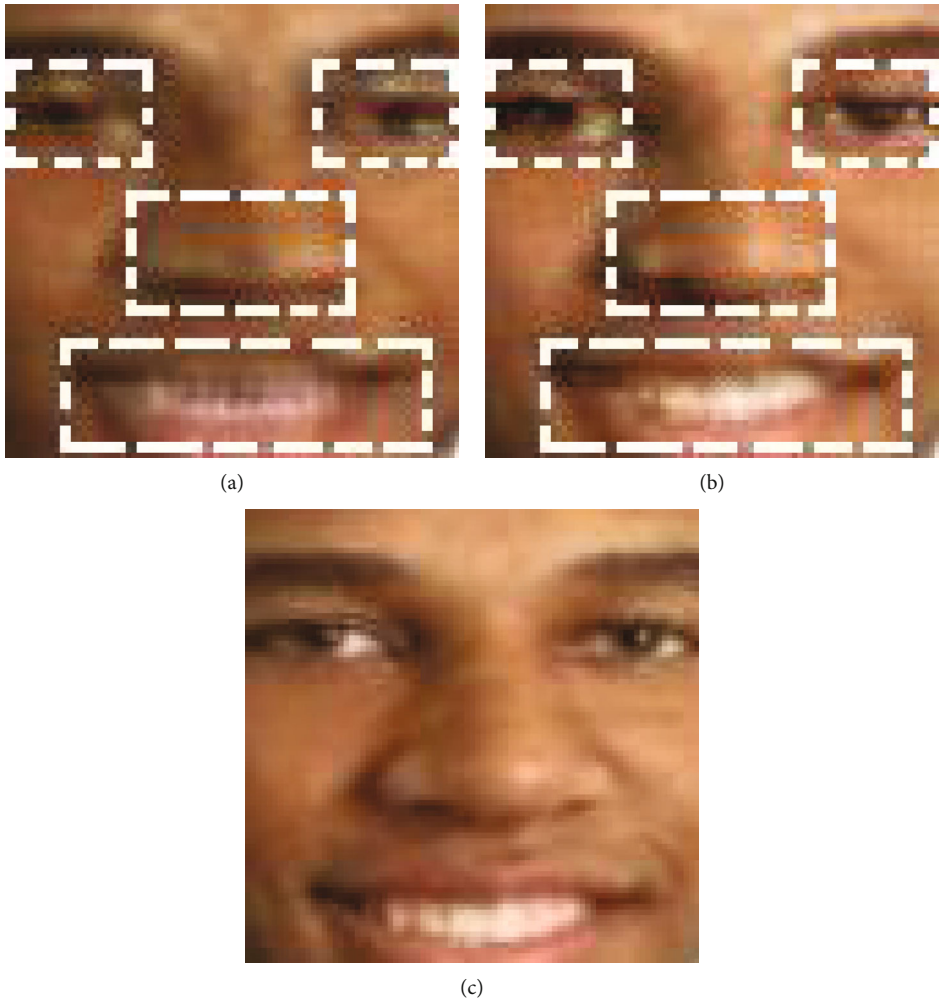
$$L_{\text{perc},fp} = E\left[\frac{1}{4}\sum_{j=1}^{4}\sum_{i}\frac{1}{N_i}\left\|\Phi_i(I_{gt,i}) - \Phi_i(I_{\text{pred},i})\right\|_1\right]. \quad (8)$$

The $\hat{y}_i$ and y_i in formula (7) are the restored part and the corresponding real part, respectively, the i in formula (8) is consistent with that in formula (3), and the j is an integer set from 1 to 4, representing the four parts of the face, respectively.

Formula (7) is the style loss of different parts of the face. Its function is to calculate the style loss through the activation map to measure the difference between the covariance of the activation map. The Euclidean distance of each image feature is used to measure the degree of perception different from the real part. Taking the features extracted in the calculation as the style, the Euclidean distance difference is calculated to constrain the texture of the face to be compensated and ensure that the texture and pixels are consistent with the theme of the whole image.

Formula (8) is the perception loss of different parts of the face. Its function is to limit the overall "features" of the completed face image, keep the structure of the restored part in line with the requirements of the original image, and improve the visual connectivity of the "features" after completion.

In order to measure the difference between the covariance of the activation map, the Euclidean distance of each image feature is used to measure the similarity between the perceptual part and the real part. The features extracted in the calculation are used as the style features of the face image, and the Euclidean distance is used to calculate the degree of



(a) (b) (c)

Figure 7: Effects under different losses: (a) adversarial loss; (b) combined losses; (c) original.

difference in style between the real Face-Part and the complemented Face-Part, to compare the completed face. The location is subject to texture constraints, to ensure that the texture and pixels conform to the theme of the whole image. Constrain the overall "features" of the completed face image to improve the visual connectivity of the "features" after the completion. The face-part discriminator network also uses the normalized convolution filtering in the network to reconstruct the style of the unmasked part of the face, so that the facial features are clearer and texture. The subject of the face image is closer. The effect is shown in Figure 7.

Figure 7 shows the effect of face part generation. Each part of D2 Face-Part network is equipped with adversarial loss, style loss, and perception loss, which restrict the authenticity, texture, and structure of the image, respectively, so that the details of the eyes, nose, mouth, and other parts of the portrait are more in line with common sense that people should have.

Add formulas (6), (7), and (8) to get the total loss of G2 and D2 Face-Part losses (formula (9))

$$L_{fp} = L_{\text{adv},fp} + L_{\text{style},fp} + L_{\text{perc},fp}. \tag{9}$$

In general, the discriminator network for image completion is composed of three discriminator networks, which are based on the integrity of the face image completion, the rationality of the face image completion part, and the generation of facial image "facial senses." It is logical. Use D2 Face-Part to analyze the structure of the face, so that the structure of the glasses, nose, and mouth is closer to the real shape. Through the constraint of external loss, the overall, partial, texture, structure, and pixel supervision of G2 can be achieved.

5. Experiment and Evaluation

5.1. Experimental Setting. Use 202599 data sets in CelebA [48] for training, testing, and evaluation. Before training, each face image is rescaled to 256 pixels × 256 pixels × 3 pixels.

The experimental environment is Windows 10 as the platform, with Pytorch 1.7 implemented on Python 3.8, the processor is i5-9400F 2.9 GHz, the memory is 32 GB, and the graphics card is RTX2070 SUPER 8 GB.

The learning rate is 0.0001 by default, and the Adam [49] gradient descent method is used to backpropagate the

Table 1: PSNR/SSIM of Figures 6(a) and 6(b).

Figure 6 no.	Context-Encoder	Globally-Locally	EdgeConnect	Ours	Ground-Truth
a	21.03	26.88	23.96	27.89	Inf
	0.9467	0.9218	0.8653	0.9508	1
b	14.89	18.08	21.00	22.76	Inf
	0.7824	0.8871	0.7731	0.8297	1

update gradient, and the Beta 1 and Beta 2 are, respectively, 0.0 and 0.9.

5.2. Results and Analysis. To compare the experimental results more intuitively, the classical algorithms with better performance in recent years are adopted, which are Context-Encoder in [31], Globally-Locally in [32], and EdgeConnect in [38]. To intuitively express the superiority of the proposed complementary algorithm, the peak signal-to-noise ratio (PSNR) is used to measure the distance between the complementary image and the original image. The larger the value of PSNR, the better the performance of the complemented image. At the same time, to reflect the authenticity of the proposed algorithm in the structure of the complementary image, the structure similarity (SSIM) is used to measure the difference between the structure of the complementary image and the original image. SSIM uses 0 as the lowest score. The higher the score, the structure of the complementary image. Logically, the more it conforms to the standard of the original image structure, the highest score is 1.

Figures 6(a) and 6(b) are the results of block occlusion completion and irregular occlusion completion, respectively. There are 8 rows and 7 columns. The rows represent different test images, and the columns represent the performance of the same type or the same method in different images. Among them, the first column represents the original image occluded by a specific mask, the second column is the mask image of a specific type of occlusion original image, and the third to sixth columns are article [31], article [32], article [38], and the completion results of the proposed method under different masks. The seventh column is the original image.

Figure 6(c) has 5 rows and 4 columns, and each row represents the results of different methods. They are the Context-Encoder method of article [31], the Globally-Locally method of article [32], the EdgeConnect method of article [38], the method and original image proposed in this article. Each column represents the left eye, right eye, nose, and mouth.

5.2.1. Analysis of Occlusion Restoration and Completion Results. It can be seen from the results in Table 1 that the effect of the proposed method is better than that of other control groups.

From the perspective of qualitative analysis, in Figures 6(a) and 6(b), there are still some noises after the completion of the image in the third column, the recovery level of details is generally poor, and the hue and brightness of the generated part are different from that of the complete image. The fourth column of facial features is too flat, and the feature recovery rate of each part of the face is low, which cannot reflect the facial features of the face well. The fifth column is better in the case of random occlusion, but it is not good in the case of block occlusion, and the complementary color does not match the basic tone of the whole image. The sixth column was based on the control group. In addition to basically avoiding noise, the hue and facial contour of the complementary color part are more consistent with the original image, and the detail texture of facial features is also better.

From the perspective of quantitative analysis, the reconstruction loss formula of the third column fits the surrounding texture according to its results. It is essentially a linear operation using L2 distance, and its fitting ability is not as good as that of adversarial loss. The fourth column uses multiple discriminators to calculate, which not only makes the generated part more specific but also takes into account the overall information. However, because it is unconditional input and there is no edge information as the condition, the visual connectivity of the image restoration results is poor, but its score shows that this idea is feasible. The fifth column method uses multiple loss functions and takes edge conditions as input, which greatly improves the visual connectivity of the generated results, but the results are slightly lower than the fourth column method. Based on the fifth column, the sixth column method further enhances the generated results.

5.2.2. Analysis of the Results of Face-Part Completion. LE, RE, N, and M in Table 2 represent the English abbreviations for the left eye, right eye, nose, and mouth, respectively. Combining the results of Figure 6(c), the proposed method has a better ability to restore Face-Part than the control group.

From the perspective of qualitative analysis, the details of the facial parts after the completion of the first line in the figure are blurred; there are obvious noises, and the structure is unclear. The complete structure of the second line is not obvious, the texture of the complete effect is flat, and the facial parts are not clear. The hue expression in the third line deviates too much from the entire image, and the supplementary details of Face-Part are not ideal. The repair effect is slightly worse when the fourth line is defective, but overall, the repair effect of the facial part is better than the control group, and the color tone and brightness are consistent with visual connectivity.

From the perspective of quantitative analysis, the Context-Encoder method in the first line only focuses on the wholeness of the local generation to fit the visual connectivity of the entire image, resulting in the inadequate generation of Face-Part (LE, RE, N, and M) details. The method

Table 2: PSNR/SSIM of Figure 6(c).

Face-Part	Context-Encoder	Globally-Locally	EdgeConnect	Ours	Ground-Truth
LE	18.99	18.66	16.16	19.37	Inf
	0.6798	0.6675	0.6640	0.7051	1
RE	17.74	18.94	15.28	19.19	Inf
	0.6252	0.6797	0.3602	0.7411	1
N	19.50	21.96	17.14	22.28	Inf
	0.6493	0.7736	0.5731	0.7628	1
M	20.99	23.36	21.17	23.41	Inf
	0.6565	0.8133	0.7407	0.7618	1
Average	19.30	20.73	17.44	21.06	Inf
	0.6527	0.7335	0.5898	0.7374	1

proposed in the second line also has the problem of the method in the first line, but it is more closely related to the local generation and the overall generation, and the generation effect is better. Although the method in the third row is not good in terms of data performance, it is a portrait restoration based on edge conditions, and its performance in the reconstruction of texture and structure is more in line with the look and feel of real portraits. The method in the fourth line uses a gated convolution block with a self-attention mechanism to identify both sides of the mask boundary more accurately. At the same time, it uses a loss function and multiple discriminators that focus on different factors of the portrait, giving a human-like Face-Part (LE, RE, N, and M) that are more real.

6. Conclusion

We have determined that GAN can be trained on external standard datasets. To generate face occlusion recovery in the adversarial network, a complete structure based on edge conditions, a convolution block based on self-attention mechanism, and a discriminator based on multiple discriminators are introduced in the GAN. The hidden part is repaired by an edge generator, and the hidden part is distinguished from the normal part by self-attention convolution block. Based on the constraints of local and facial feature parts, multiple discriminators are used to completing the recovery results of style texture and different levels. To verify the validity of this method, three methods, Context-Encoder, Globally-Locally, and EdgeConnect, are used to compare. The results show that the comprehensive level of the proposed method is higher than that of the control group.

However, the method still has some deficiencies in the details of face integrity, and the effect for small-sized parts of the face still needs to be improved. In the complex texture part, the restoring effect is also limited, and the restoring effect is relatively simple. So overcoming these problems is also our future work. We have determined that image inpainting based on edge conditions and deep learning using a GAN can effectively solve this problem. Next, we will further improve the image quality based on the latest research results and the advantages of the proposed method.

Data Availability

The URL of the public dataset used to support the results of this study is http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported in part by the National Key R&D Program of China under grant numbers 2017YFC0821602, 2019QY1604, and 2019YFE0122600; in part by the National Natural Science Foundation of China under grant number U1836217; and in part by the Open Platform Innovation Foundation of Hunan Provincial Education Department under grant number 20K046.

References

[1] M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, X. Du, I. Ali, and M. Guizani, "A survey of machine and deep learning methods for Internet of Things (IoT) security," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1646–1685, 2020.

[2] Z. Xiao, F. Li, H. Jiang et al., "A joint information and energy cooperation framework for CR-enabled macro–femto heterogeneous networks," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2828–2839, 2020.

[3] Y. Xu, C. Zhang, G. Wang, Z. Qin, and Q. Zeng, "A blockchain-enabled deduplicatable data auditing mechanism for network storage services," *IEEE Transactions on Emerging Topics in Computing*, p. 1, 2020.

[4] C. Zhang, Y. Xu, Y. Hu, J. Wu, J. Ren, and Y. Zhang, "A blockchain-based multi-cloud storage data auditing scheme to locate faults," *IEEE Transactions on Cloud Computing*, p. 1, 2021.

[5] Y. Xu, J. Ren, Y. Zhang, C. Zhang, B. Shen, and Y. Zhang, "Blockchain empowered arbitrable data auditing scheme for network storage as a service," *IEEE Transactions on Services Computing*, vol. 13, pp. 289–300, 2020.

[6] Z. Jiale, Z. Yanchao, C. Bing, H. Feng, and Z. Kun, "Survey on data security and privacy-preserving for the research of edge computing," *Journal on Communications*, vol. 39, pp. 1–21, 2018.

[7] Z. Xiao, X. Dai, H. Jiang, and D. Wang, "Vehicular task offloading via heat-aware MEC cooperation: a game-theoretic method with correlated equilibrium," *IEEE Internet of Things Journal*, vol. 7, pp. 2038–2052, 2019.

[8] Y. Wu, Q. Liu, R. Chen, C. Li, and Z. Peng, "A group recommendation system of network document resource based on knowledge graph and LSTM in edge computing," *Security and Communication Networks*, vol. 2020, Article ID 8843803, 11 pages, 2020.

[9] Z. Cai and Z. He, "Trading private range counting over big IoT data," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pp. 144–153, Dallas, TX, USA, July 2019.

[10] X. Zhou, W. Liang, K. I. K. Wang, R. Huang, and Q. Jin, "Academic influence aware and multidimensional network analysis for research collaboration navigation based on scholarly big data," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 1, pp. 246–257, 2021.

[11] X. Zheng and Z. Cai, "Privacy-preserved data sharing towards multiple parties in industrial IoTs," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 968–979, 2020.

[12] X. Zhou, X. Xu, W. Liang et al., "Intelligent small object detection based on digital twinning for smart manufacturing in industrial CPS," *IEEE Transactions on Industrial Informatics*, p. 1, 2021.

[13] X. Yan, Y. Xu, X. Xing, B. Cui, Z. Guo, and T. Guo, "Trustworthy network anomaly detection based on an adaptive learning rate and momentum in IIoT," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 6182–6192, 2020.

[14] X. Zhou, Y. Hu, W. Liang, J. Ma, and Q. Jin, "Variational LSTM enhanced anomaly detection for industrial big data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3469–3477, 2021.

[15] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of Graphics Tools*, vol. 9, no. 1, pp. 23–34, 2004.

[16] F. Tang, Y. Ying, J. Wang, and Q. Peng, "A novel texture synthesis based algorithm for object removal in photographs," in *Advances in Computer Science - ASIAN 2004. Higher-Level Decision Making. ASIAN 2004. Lecture Notes in Computer Science, vol 3321*, M. J. Maher, Ed., pp. 248–258, Springer, Berlin, Heidelberg, 2004.

[17] J. Hays and A. A. Efros, "Scene completion using millions of photographs," *Communications of the ACM*, vol. 51, no. 10, pp. 87–94, 2008.

[18] C. Barnes, D. B. Goldman, E. Shechtman, and A. Finkelstein, "The PatchMatch randomized matching algorithm for image manipulation," *Communications of the ACM*, vol. 54, no. 11, pp. 103–110, 2011.

[19] J. Sun, "Computing nearest-neighbor fields via Propagation-Assisted KD-Trees," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 111–118, Providence, RI, USA, June 2012.

[20] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, pp. 577–590, 2018.

[21] Z. Cai and X. Zheng, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 766–775, 2020.

[22] Y. Xu, C. Zhang, Q. Zeng, G. Wang, J. Ren, and Y. Zhang, "Blockchain-enabled accountability mechanism against information leakage in vertical industry services," *IEEE Transactions on Network Science and Engineering*, p. 1, 2020.

[23] Y. Xu, Q. Zeng, G. Wang, C. Zhang, J. Ren, and Y. Zhang, "An efficient privacy-enhanced attribute-based access control mechanism," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 5, pp. 1–10, 2020.

[24] L. Qi, C. Hu, X. Zhang et al., "Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4159–4167, 2021.

[25] X. Yan, Y. Xu, B. Cui, S. Zhang, T. Guo, and C. Li, "Learning URL embedding for malicious website detection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6673–6681, 2020.

[26] X. Zhou, W. Liang, S. Shimizu, J. Ma, and Q. Jin, "Siamese neural network based few-shot learning for anomaly detection in industrial cyber-physical systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5790–5798, 2021.

[27] Y. Xu, X. Yan, Y. Wu, Y. Hu, W. Liang, and J. Zhang, "Hierarchical bidirectional RNN for safety-enhanced B5G heterogeneous networks," *IEEE Transactions on Network Science and Engineering*, p. 1, 2021.

[28] X. Zhou, Y. Li, and W. Liang, "CNN-RNN based intelligent recommendation for online medical pre-diagnosis support," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 3, pp. 912–921, 2021.

[29] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.

[30] Z. Cai, Z. Xiong, H. Xu, P. Wang, W. Li, and Y. Pan, *Generative Adversarial Networks: A Survey Towards Private and Secure Applications*, ACM Computing Surveys, 2021.

[31] P. Krahenbuhl, J. Donahue, T. Darrell, and A. Efros, "Context-encoders: feature learning by inpainting," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2536–2544, Las Vegas, NV, USA, June 2016.

[32] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–14, 2017.

[33] J. Yu, Z. Lin, J. Yang, X. Shen, and X. Lu, "Generative image inpainting with contextual attention," in *In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5505–5514, Salt Lake City, UT, USA, 2018.

[34] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," 2018, https://arxiv.org/pdf/1805.08318.pdf.

[35] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," in *In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4470–4479, Seoul, Korea(South), 2019;.

[36] C. Kun, W. Fei, L. Lizhi, Y. Zhaokun, and W. Qian, "Face completion algorithm based on condition generation adversarial network," *Transducer and Microsystem Technologies(China)*, vol. 38, pp. 129–132, 2019.

[37] C. Xie, S. Liu, C. Li et al., "Image inpainting with learnable bidirectional attention maps," in *In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8857–8866, Seoul, Korea(South), 2019.

[38] K. Nazeri, E. Ng, F. Joseph, F. Qureshi, and M. Ebrahimi, "EdgeConnect: generative image inpainting with adversarial edge learning," 2019, https://arxiv.org/pdf/1901.00212v3.pdf.

[39] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, *Spectral Normalization for Generative Adversarial Networks*, International Conference on Learning Representations, Vancouver Convention Center, Vancouver Canada, 2018.

[40] G. Liu, F. Reda, K. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *In Proceedings of the European Conference on Computer Vision*, pp. 89–105, 2018.

[41] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[42] I. Mokris and L. Skovajsova, "Proposal of cascade neural network model for text document space dimension reduction by latent semantic indexing," in *In Proceedings of the 2008 6th International Symposium on Applied Machine Intelligence and Informatics*, pp. 79–84, 2008.

[43] A. Latreche and L. Guezouli, "Similarity measure for semi-structured information retrieval based on the path and neighborhood," in *In Proceedings of the 2012 International Conference on Information Technology and e-Services*, pp. 1–5, 2012.

[44] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *In Proceedings of the Computer Vision - ECCV 2016*, pp. 694–711, Amsterdam, The Netherlands, 2016.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, 2016.

[46] W. Fupin, L. Wenlou, L. Ying, L. Jin, and G. Yanchao, "Face inpainting algorithm combining edge information with gated convolution," *Journal of Frontiers of Computer Science and Technology(China)*, vol. 15, pp. 150–162, 2021.

[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, vol. 4, pp. 1409–1421, 2014.

[48] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 3730–3738, Sadversarial ago, Chile, 2015.

[49] D. Kingma and J. Ba, "Adam: a method for stochastic optimization," *Computer Research Repository*, vol. 12, pp. 1–6, 2014.