

Article

A Robust Face Recognition Algorithm Based on an Improved Generative Confrontation Network

Huilin Ge , Yuewei Dai *, Zhiyu Zhu and Biao Wang

School of Electronic Information, Jiangsu University of Science and Technology, Zhenjiang 212003, China; ghl1989@just.edu.cn (H.G.); zzydzz@163.com (Z.Z.); wangbiao@just.edu.cn (B.W.)

* Correspondence: dywjust@163.com

Abstract: Objective: In practical applications, an image of a face is often partially occluded, which decreases the recognition rate and the robustness. Therefore, in response to this situation, an effective face recognition model based on an improved generative adversarial network (GAN) is proposed. Methods: First, we use a generator composed of an autoencoder and the adversarial learning of two discriminators (local discriminator and global discriminator) to fill and repair an occluded face image. On this basis, the Resnet-50 network is used to perform image restoration on the face. In our recognition framework, we introduce a classification loss function that can quantify the distance between classes. The image generated by the generator can only capture the rough shape of the missing facial components or generate the wrong pixels. To obtain a clearer and more realistic image, this paper uses two discriminators (local discriminator and global discriminator, as mentioned above). The images generated by the proposed method are coherent and minimally influence facial expression recognition. Through experiments, facial images with different occlusion conditions are compared before and after the facial expressions are filled, and the recognition rates of different algorithms are compared. Results: The images generated by the method in this paper are truly coherent and have little impact on facial expression recognition. When the occlusion area is less than 50%, the overall recognition rate of the model is above 80%, which is close to the recognition rate pertaining to the non-occluded images. Conclusions: The experimental results show that the method in this paper has a better restoration effect and higher recognition rate for face images of different occlusion types and regions. Furthermore, it can be used for face recognition in a daily occlusion environment, and achieve a better recognition effect.

Keywords: generative confrontation network; face recognition; convolutional neural network; loss function; double discriminator



Citation: Ge, H.; Dai, Y.; Zhu, Z.; Wang, B. A Robust Face Recognition Algorithm Based on an Improved Generative Confrontation Network. *Appl. Sci.* **2021**, *11*, 11588. <https://doi.org/10.3390/app112411588>

Academic Editors: Kelvin K. L. Wong and Fabio La Foresta

Received: 27 September 2021

Accepted: 29 November 2021

Published: 7 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the development of artificial intelligence technology, biometric recognition technology has received unprecedented attention. As a kind of biological characteristics, face recognition technology has great potential application in security, access control systems, finance and other fields due to its advantages such as non-interference and uniqueness [1]. Face recognition offers easy security allowance, quick operation and being able to identify diversified features. In practice, many scholars have studied face recognition and developed new and improved techniques used in security [2]. However, due to the under-developed network, limited resources of face images and poor quality of images, many researchers mostly studied it from the perspective of algorithms, but the recognition accuracy is low, much worse than the human eye's recognition effect. With the gradual improvement of machine learning technology, many powerful algorithms have been developed, such as the genetic algorithm, Bayesian classifier and support vector machines. The application of these algorithms [3] in face recognition technology improves the accuracy of face recognition to a certain extent, but its feature extraction is complex and single, especially when the faces are occluded, which is greatly affected by human factors. Thus,

these methods are not widely used. With the development of computer technology, a large number of face image resources can be obtained, and the rapid development of deep learning provides new vitality for face recognition; thus, recognition accuracy has greatly improved. With the gradual expansion of the application field of face recognition, especially since the COVID-19 pandemic, people have been wearing masks and other protections more frequently. Face recognition in a sheltered environment is facing certain technical challenges. At present, traditional methods are usually used for partial occlusion face recognition, and the ideas are generally divided into two types: the discarding method and the filling method.

The discarding method is designed to simplify or discard the occluded part of the information through some algorithms, and then perform face recognition based on the irrelevant part. Some researchers [4] proposed a method based on sparse tables to realize occlusion facial expression recognition. However, because some important parts of the human face contain a large amount of face information, such as mouth, eyes and nose, it is obviously unreasonable to discard these parts directly when some parts are blocked. The filling method is designed to repair the occluded part first, and then identify it. Xue et al. [5] used robust principal component analysis and saliency detection methods to reconstruct the occluded pixels, and then the AdaBoost classifier [6] updated by the weights performs facial expression recognition on the irrelevant face images. However, the image restoration result of this method is not ideal, and therefore the recognition rate is not ideal either. In image recognition research, face identification technology has been introduced in recent years, but the image processing that is related to face recognition has developed rapidly, and has achieved remarkable results in recognition accuracy etc., and related technical achievements have attracted worldwide attention. Due to the lack of prior knowledge of face images, large illumination changes, complex backgrounds, and variable face angles, the demand for face images is large, expression changes are large, and face occlusion leads to low accuracy of face recognition. Deep learning applications often use convolutional neural networks to achieve image processing and recognition with high efficiency and accuracy [7–10]. Facial images are highly structured images. Combining prior facial knowledge is a very popular method in face super-resolution technology based on deep learning.

Due to its strong learning ability, deep learning methods have become a better solution to the problem of face recognition in occluded environments, but still face many problems that need to be solved. Reducing the impact of occlusion on the performance of face recognition algorithms is one of the key and difficult problems in this field. This paper proposes a deep generative confrontation network model based on a dual discrimination network to recognize occluded faces. A face image contains rich information, but only part of the information plays a key role in the recognition process. Before recognition, the face image needs to be preprocessed to extract useful feature information, thereby improving the speed and accuracy of face recognition. The model uses the generator and two discriminators to conduct adversarial learning in the early stage to repair the occluded image [11]. Due to the addition of the local discriminator and the global discriminator, the model can be better in detail and globally when restoring the image. The occluded image is restored to make the repaired face image more realistic. On this basis, a convolutional network based on ResNet-50 is constructed to re-identify the repaired face image [12]. Furthermore, the introduction of the Regular Face loss function and the SoftMax loss function, to deal with the feature fusion problem that may exist in the face after the repair, can improve the classification effect in one step.

2. Related Theories

2.1. Face Recognition

Face recognition is a technology that uses images containing human faces collected by a camera and detects them through related technologies. It mainly recognizes the facial

feature information of the human body to distinguish features, and finally realizes the classification and recognition of human individuals [13].

In summary, the process of face recognition can be roughly divided into four steps, as shown in Figure 1.

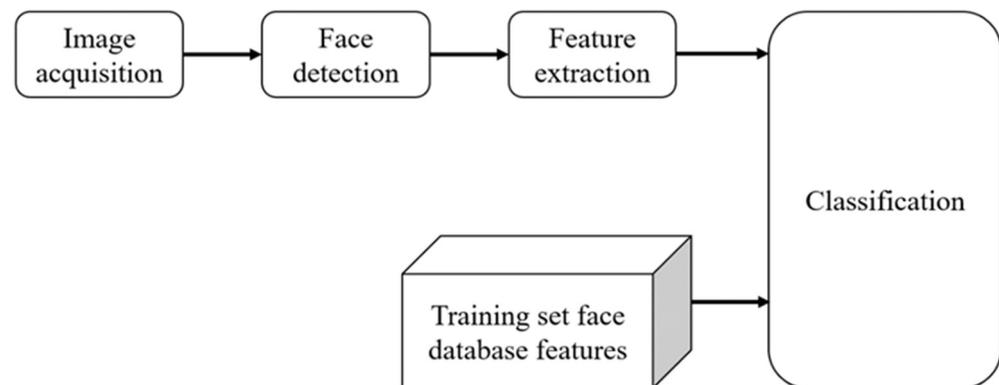


Figure 1. Face recognition flow chart.

At present, reducing the impact of the features of the irrelevant area and repairing the inherent features of the occluded area are two common ideas. When the face is in a clearly visible and irrelevant environment, with the support of deep learning technology and a large number of data sets, its feature extraction is easier. Deep learning frameworks [14] can be used for face recognition and improve efficiency in identifying face photos that are of poor quality, including solving the problem of image artifacts or occlusion; if it is partially occluded, not only the features of the occluded area but also the extraction of the entire facial features will be affected. The research begins by highlighting the face area in the image and weakening the background area of the non-face in the image, and trying to expand the data set used for model training and testing to improve the recognition effect. During the study, they restore the occluded part of the face, and then classify the repaired face through an appropriate recognition network [15]. This method restores the original missing feature information, so that the face to be recognized has rich features close to the original image. The methods used include automatic encoder and generative adversarial network (GAN). The autoencoder is used as the generator, and the global and local discriminators are used to make the semantics of the generated image richer. The method based on deep learning is an end-to-end method that combines feature extraction and feature classification into one model and through the introduction of Regular Face loss function and SoftMax loss function to deal with the feature fusion problem that may exist in the face after repair.

2.2. Effective Use of Irrelevant Facial Features

When face information occlusion occurs, the feature extraction of the occluded face can be assisted by using the features of other irrelevant parts of the face, that is, to supplement, restore and predict the image content of the missing area according to the neighboring information of the occluded area. Then, feature extraction is performed, according to the following method.

(1) The deep convolutional neural network algorithm extracts the attribute features of the face. The convolutional neural network is an important part of deep learning face recognition. Its function is to extract richer and deeper features from the face. It is mainly composed of a convolutional layer, a pooling layer and a fully connected layer [16]. The core of the convolutional layer is the convolutional kernel, and the input features are passed through a series of convolutional operations through the convolutional kernel to obtain deeper features. After the pooling layer is connected to the convolutional layer, the main function is to reduce the size of the feature map obtained through the convolutional layer, thereby reducing the amount of parameter calculation.

Generally, there are two types of commonly used pooling layer: global tie pooling and maximum pooling. The fully connected layer is usually used at the end of the network to convert two-dimensional feature maps into one-dimensional features for identification and classification. Designing a proper convolutional network to effectively extract facial features will also have a greater impact on the final recognition accuracy [17]. The self-integrating neural network AlexNet was released, and researchers designed many network structures to extract image features at a deeper level, such as VGGNet, GoogleNet, and ResNet [18]. In face recognition algorithms, deep learning applications [14] are commonly implemented and tested in situations whereby the humans are photographed in front of various colored and complex backgrounds.

Wang et al. used the anchor strategy and data enhancement strategy to build a face recognition network (face attention network, FAN) that integrates an attention mechanism, as shown in Figure 2 [19]. During model training, different attention mechanisms are set for the feature maps at different positions of the feature pyramid based on the face size, that is, the attention function is added to the RetinaNet anchor through multi-scale feature extraction, multi-scale anchor and semantic segmentation.

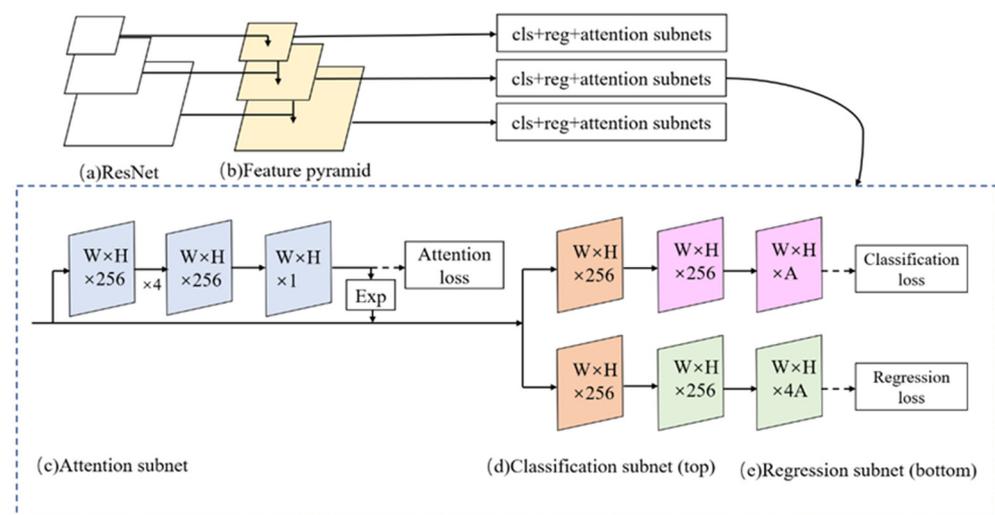


Figure 2. FAN network model.

The scale attention mechanism implicitly learns the face in the occluded area, and improves the detection effect of the occluded face. The condition of training is that the features of the face area and the occlusion area in the data set are mixed together. This will cause the attention mechanism to simultaneously enhance the facial features and the occlusion features contained in the face area, and the method of dividing different attention maps based on size does not guarantee that the face is divided into appropriate feature maps, thereby affecting the recognition effect [20].

2.3. Generative Adversarial Network

As generative adversarial networks (GANs) have achieved good results in machine learning tasks, a GAN-based generative model is derived from this to solve the problem of occluded face image restoration [21]. The classic GAN model is shown in Figure 3.

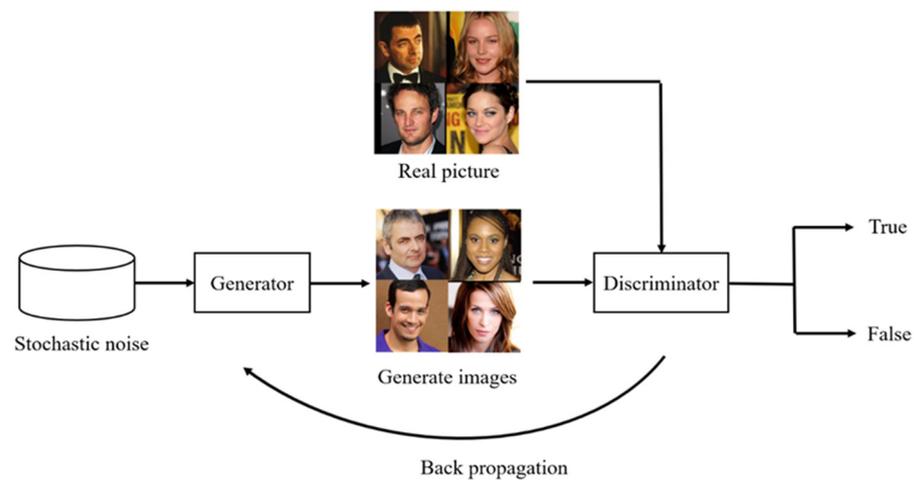


Figure 3. Classical GAN model.

It consists of a generator and a discriminator. The generator is used to learn the distribution of real images. The discriminator compares the image obtained by the generator with the original image to determine the authenticity of the generated image. The objective function of GAN training is

$$\min_G \max_d V(D, G) = E_{x \sim p_{data}(x)} [\lg D(x)] + E_{z \sim p_z(z)} [\lg(1 - D(G(z)))] \quad (1)$$

In Equation (1), x represents the real picture, z represents the noise input to the G network and $G(z)$ represents the picture generated by the G network. $D(x)$ represents the probability that the D network judges whether the real picture is real, and $D(G(z))$ is the probability that the D network judges whether the picture generated by G is real [22].

Then, we input the generated image $G(z)$ and the real image x into the discriminator D , and train the discriminating ability of the discriminator D so that the probability $D(G(z))$ of the output of the discriminating network D tends to 0 and $D(x)$ tends to 1. This error is fed back to the generator, and the generator adjusts the parameters according to this error, so as to adjust itself to generate an image closer to the real image, to deceive the discriminator so that $D(G(z))$ is close to 1. With continuous training, the generator and the discriminator have reached a balance in the confrontation iteration, that is, $D(G(z))$ tends to 0.5. At this time, in the case of a fixed generator, the objective function formula in Equation (2) can obtain the optimal solution of the discriminator:

$$D_G(x) = P_{data}(x) / (P_{data}(x) + P_g(x)) \quad (2)$$

In this case, the discriminator cannot determine whether the generated image is real or generated by training, because the generative adversarial network can learn image features in similar image data sets through a large number of feature learning and training, so it generates more realistic images [23].

3. Methods

3.1. Face Recognition Model Based on Dual Discriminant Confrontation Network

This paper proposes an occlusion facial expression recognition model based on a generative confrontation network. The model is divided into two modules, namely, the occlusion face image restoration part and the face recognition part, as shown in Figure 4.

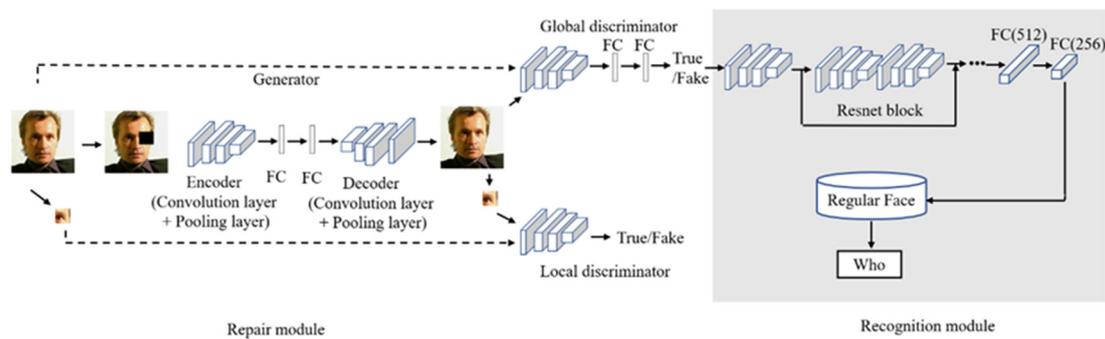


Figure 4. Occlusion face recognition model based on dual discrimination network.

The difference between this model and traditional GAN is that the input to the repair network is an occluded image, rather than a set of random noise [24]. The model is divided into two modules, namely, the occlusion face image restoration module and the face recognition module. In this repair model, dual discriminators, namely, local discriminator and global discriminator, are used most often. The introduction of a local discriminator can better repair the details of the occluded part, and the global discriminator is used to identify whether the entire image after repairing the damaged area is true and consistent. Through these two discriminators and generators against training, a better image restoration effect can be produced. The recognition part uses part of its convolutional layer and pooling layer on the basis of the global discriminator, and uses it as a feature extractor.

The generator G is designed as an automatic encoder to generate new content for the input missing image. First, the model input is mapped to a hidden layer through the encoder, which contains two known areas and missing areas in the original occlusion image. The decoder uses this hidden information to generate filling content. Unlike the original GAN, the input of the generator G in this paper is no longer random noise, but an occluded face image. The generator's network structure will be based on the VGG19 network "conv1" to "pool3" architecture. On this basis, two convolutional layers and a pooling layer will be superimposed, and a fully connected layer will be added as an encoder. The input dimension of the encoder is 128×128 , each convolutional layer uses a 3×3 convolutional kernel, and each convolutional layer is followed by a Leaky ReLU activation layer. The maximum pooling is performed in the pooling layer, and the window size is 2×2 [25]. The decoder has a symmetrical structure. The face image is gradually restored through the convolutional layer Conv and the upsampling layer Upsampling. Between the encoder and the decoder, two fully connected layers with 1024 neurons are used as the middle floor.

Only the image generated by the generator can capture the rough shape of the missing facial components or generate the wrong pixels. In order to obtain a clearer and more realistic image, this paper will use two discriminators D : a local discriminator and a global discriminator. If only a local discriminator is used, it has certain limitations. First, it cannot standardize the global structure of the face, and cannot guarantee the consistency of the occluded area and the non-occluded area and the continuity of the global picture. Secondly, when the newly generated pixels are constrained by the surrounding environment, due to the "inverse pooling" structure of the decoder. In the process of backpropagation, it is difficult for the local discriminator to directly affect the area outside the occluded area, and the inconsistency of pixel values along the boundary area is very obvious [26]. Therefore, two discriminators are used to perfect the details of the generated images, making the generated images more realistic. The basic structure of the ordinary discriminator is the convolutional layer, the fully connected layer, the densely connected layer and the fully connected layer, and then the result of the discrimination, that is, the probability that the input sample is a true sample, is output as a real number, and its input is generally a real image and a generated image.

However, the image range received by this type of discriminator is too large, which may cause low resolution and unclear images when the image is generated. In this paper, the global discriminator adopts PatchGAN, as shown in Figure 5. The discriminator is composed of 5 convolutional layers, and the output is an $n \times n$ matrix, and the mean value of the output matrix is output as True or False. Each output in the output matrix represents a receptive field of the original image, corresponding to a part of the original image. Compared with the output of a general GAN network discriminator, the discriminator takes into account the influence of different parts of the image, and the discriminative output is performed on a small part of the image, so that the model can pay more attention to the details of the image during training, and the generated image is clearer. In this paper, the input image is processed into different blocks with a size of $8 \times 8 \times 1$, and the input size of the discriminator is $256 \times 256 \times 3$.

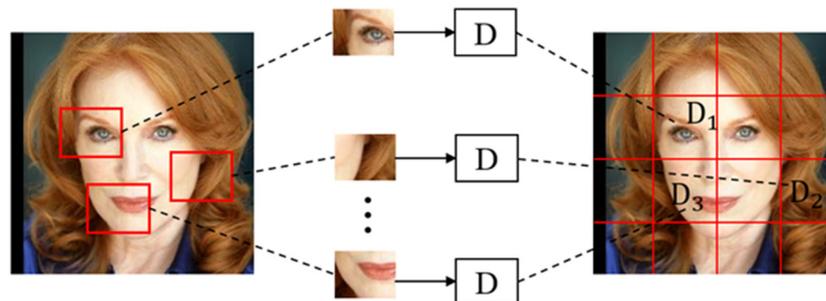


Figure 5. The principle of PatchGAN.

The input dimension of the local identification network is 64×64 , and the entire network structure is a fully convolutional network with a convolutional kernel size of 3×3 and 4×4 . There are a total of 11 convolutional layers, and each of the first 10 convolutional layers is followed by a LeakyReLU activation layer, and the last convolutional layer is followed by a TanH activation layer, which can better train the network. The alternating structure of multiple convolutional layers and non-linear activation layers makes its feature-extraction ability strong, and the model depth and performance are also more suitable. Therefore, this paper builds a local discriminator based on this model.

3.2. Loss Function

First, the reconstruction loss L_r is introduced into the generator, that is, the L_r distance between the output image of the generator and the original image. When there is only L_r , the generated content is often fuzzy and smooth. This is because the L_r loss excessively penalizes the outliers, and the network can avoid excessive penalties. By using two discriminators, this paper adopts an adversarial loss function, which reflects how the generator deceives the discriminator to the greatest extent, and how the discriminator distinguishes the authenticity from the fabrication. It is defined as

$$L_{a_1} = \min_G \max_d E_{x \sim p_{data}(x)} [\lg D(x)] + E_{z \sim p_z(z)} [\lg(1 - D(G(z)))] \quad (3)$$

In Equation (3), $p_{data}(x)$ and $p_z(z)$ respectively represent the distribution of noise variable z and real data. The two discriminant networks $\{a_1, a_2\}$ have the same definition as the loss function. The single difference is that the local discriminator only provides the loss gradient for the missing area, while the global discriminator backpropagates the loss gradient in the entire image. When the loss gradient is provided, the global discriminator backpropagates the loss gradient in the entire image.

Because facial features are very rich, the shallow convolutional neural network structure cannot extract the deep features well, but blindly stacking convolutional layers and pooling layers to deepen the network depth will cause gradient disappearance and network degradation, affecting the network optimization performance.

In this paper, the identification network adopts the residual network ResNet-50 structure, which eliminates the network degradation problem in the deep convolutional neural network through jump connections, reduces the difficulty of network training, and does not increase the additional parameters and calculations. However, blindly stacking convolutional layers and pooling layers to deepen the network depth will cause gradients to disappear and network degradation, and affect network optimization performance [27].

The loss function is used to measure the difference between the predicted value and the actual value of our network output. It is a non-negative function. Generally speaking, the smaller the loss function, the better the robustness of the model. The purpose of the loss function is to increase the inter-class distance of different categories and reduce the intra-class distance of the same category. Common loss functions such as Center Loss, Softmax and the variants can better reduce the inter-class distance of features of the same category. However, because many facial features are similar, the distance between different categories may also be small; thus, the above loss function may cause low recognition accuracy in face recognition. In this paper, the recognition module uses the Regular Face loss function, which uses the angular distance of the center points of different categories to quantify the distance between different categories, as shown in Equation (4):

$$sep_i = \max_{j \neq i} \cos(\varnothing_{(i,j)}) = \max_{j \neq i} \frac{\omega_i \cdot \omega_j}{|\omega_i| \cdot |\omega_j|} \quad (4)$$

The loss function is

$$L_r(\omega) = \frac{1}{c} \max_{j \neq i} \frac{\omega_i \cdot \omega_j}{|\omega_i| \cdot |\omega_j|} \quad (5)$$

In Equation (5), ω represents the weight vector of the category, ω_i represents the i th column of ω , ω_j represents the cluster center closest to i , and $\varnothing_{(i,j)}$ represents the angle between ω_i and ω_j . It is hoped that the larger the value of $\varnothing_{(i,j)}$, the better; that is, the smaller the value of sep_i , the better. In the identification module of this paper, the loss function is used in conjunction with Softmax, which can effectively reduce the intra-class distance of the same class and increase the inter-class distance of different classes. Softmax is formulated as follow:

$$L_1 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{\omega_{y_1}^T X_i + b_{y_1}}}{\sum_{j=1}^n e^{\omega_{y_1}^T X_i + b_{y_1}}} \quad (6)$$

Therefore, the loss function of the recognition module is

$$L_{total} = L_1 + \lambda L_r(\omega) \quad (7)$$

In Equation (7), λ is used to balance the two loss functions, and the value is 1 in this paper [20].

4. Experiments

4.1. Simulation Experiment

In order to verify the effect of the restoration method in this paper, we carried out multiple rounds of instance verification on image restoration. The experimental platform is a programming environment combining Windows 10, Python 3.6 and TensorFlow, Intel 4.20 GHz CPU clock frequency and 16.0 GB memory.

4.2. Data Preprocessing

In this paper, 20 human subjects with different identities in the CASIA-WebFace data set totaling about 12,000 images were used as the data set. Since the restoration method in this paper was mainly used for object occlusion, when selecting images, we mainly chose images with sufficient light and frontal faces to avoid the influence of light occlusion and self-occlusion on the experiment. After calibration, the size of the image was uniformly adjusted to 128×128 .

Occlusion in reality is caused by various factors, so there is currently no universal, mature and standard occlusion facial expression data set [28]. Therefore, in the training process of this model, for the image filling part, the training data were occluded by the simulation system. The image size of the occlusion part was 64×64 , and the occlusion position was random.

In the test process, images with different occlusion areas were selected to test the face recognition rate and repair effect under different occlusion areas. Random occlusion 10%, 20%, 30%, 40% and 50% were used to simulate temporary occlusion with an unfixed position. The occluded image is shown in Figure 6.

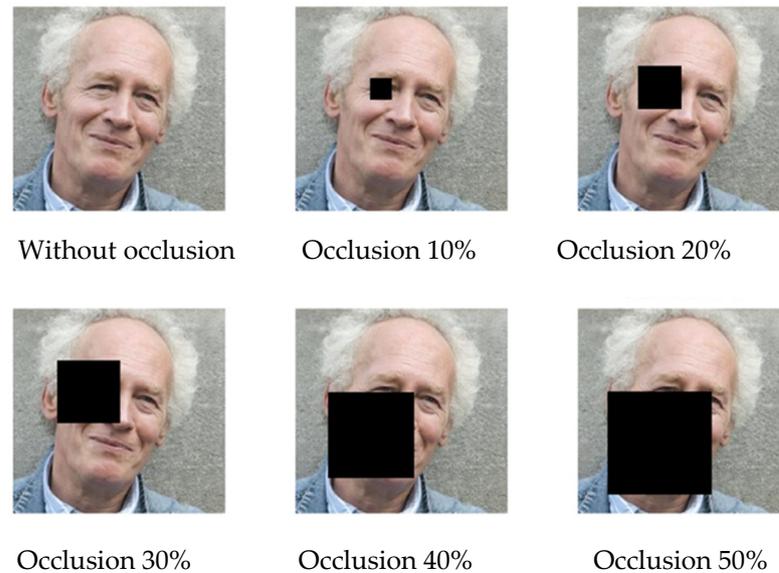


Figure 6. Image preprocessing and occlusion simulation.

4.3. Model Training

The training for this paper was divided into two parts: the training of the repair module and the recognition module.

The model first trained the face repair network on the CelebA data set. The training process of the face repair network was divided into three stages: (1) Use reconstruction loss to train the generative network to obtain fuzzy content; (2) add the local discriminator loss to fine-tune the generative model; (3) use the global discriminator loss to adjust the parameters of the generated model. This can prevent the discriminator from being too strong at the beginning of training. When the GAN model is close to the Nash equilibrium and the two-classification accuracy is about 0.5, the discriminator can hardly judge the authenticity of the input sample [29]. The hyperparameter settings of the repair network are shown in Table 1.

Table 1. Hyper-parameter settings for the training of inpainting network.

Optimization	G Learning Rate	D Learning Rate	Number of Batches	Number of Iterations
Adam	0.0002	0.0002	250	300

At the beginning of the recognition phase, all of the 12,000 (approx.) images were divided: 4/5 were used as the training set and 1/5 was used as the test set. To better compare the performance of the recognition network, we first trained the recognition model to achieve a higher recognition accuracy on the normal face data set, and used the trained network at this time as a benchmark to recognize the repaired face [30]. This benchmark can be adjusted with the size of the face data set, and the recognition accuracy after the final restoration will also change with the benchmark, but the ultimate goal of

this paper is to explore the restoration effect and improve the recognition accuracy, so the size of the benchmark does not affect a comparative experiment. To better compare the recognition network performance, this benchmark can be adjusted with the size of the face data set, and the recognition accuracy after the final restoration will also change with the benchmark, but the ultimate goal of this paper is to explore the restoration effect and improve the recognition. The network hyperparameter settings are shown in Table 2.

Table 2. Hyper-parameter settings for the training of identification network.

Optimization	G Learning Rate	D Learning Rate	Number of Batches	Number of Iterations	Balance Parameter λ
Adam	0.0005	1×10^{-8}	128	100	1

5. Results and Discussion

5.1. Image Restoration Results

(1) Image restoration results of different occlusion types

The filling effect of the model in this paper is shown in Figure 7A. The occlusion area is random occlusion, where the first line is the original image before occlusion, the second line is the face image after occlusion and the third line is the result of face filling and image restoration.

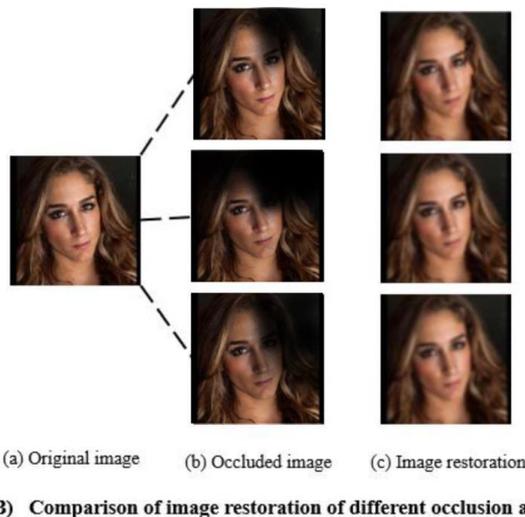
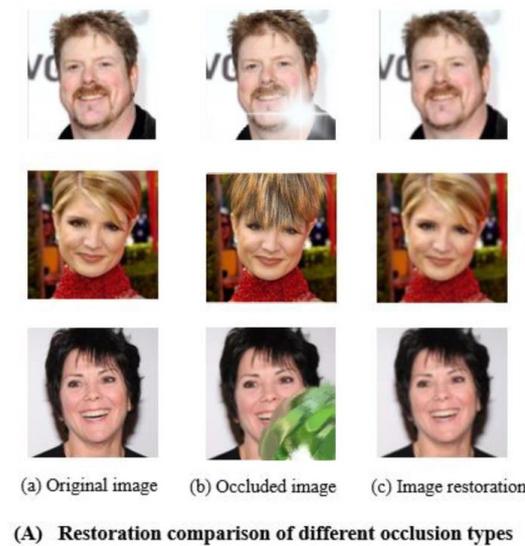
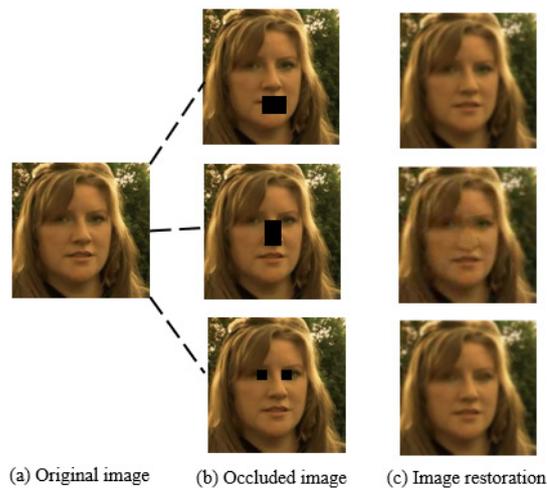


Figure 7. Cont.



(C) Comparison of image restoration of different occlusion site

Figure 7. Different occlusion conditions and repair results.

It can be seen from Figure 7A that after face filling and image restoration, the entire face image looks true and coherent, and subtle changes may occur in the unoccluded area.

(2) Image restoration results of different occlusion areas

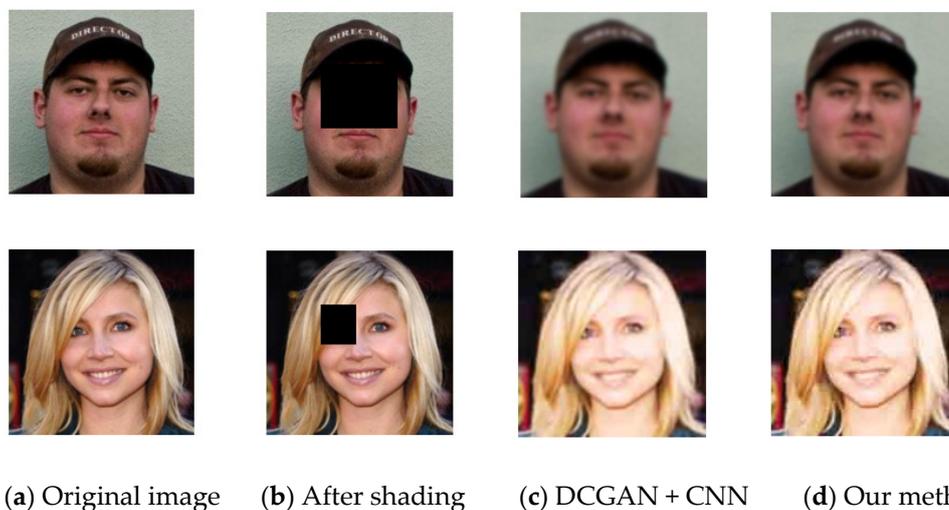
The analysis of different occlusion areas is based on the same occlusion shape of the same person, as shown in Figure 7B. In this paper, linear occlusions with different areas and rectangular occlusions with different areas are selected for experiments. It can be seen from the image restoration results that the image restoration effect gradually deteriorates as the occlusion area increases.

(3) The image restoration effect of different occluded parts

The facial features have a great influence on the accuracy of face recognition, so this paper experimented with different occluded areas of face repair effects, as shown in Figure 7C. It can be seen from the figure that when the mouth and nose are occluded, the image restoration results are relatively good, but when the eyes are occluded, the repaired details are lost more seriously.

(4) Comparison with other algorithm repair effects

Since the PCA + SVM method and the SRC and CNN methods directly recognize the occluded image, only the repair effect of the DCGAN + CNN method and the method in this paper will be compared, as shown in Figure 8.



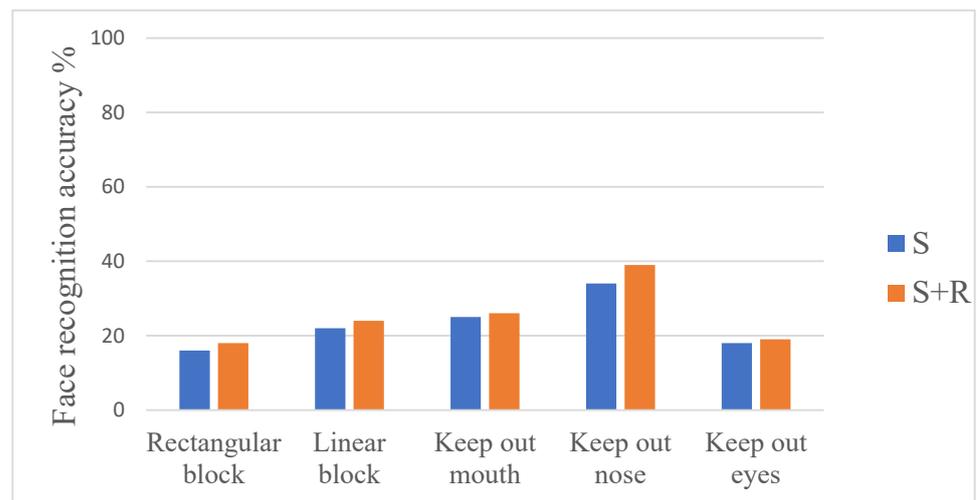
(a) Original image (b) After shading (c) DCGAN + CNN (d) Our method

Figure 8. Comparison of image restoration results between the model in this paper and the DCGAN + CNN method.

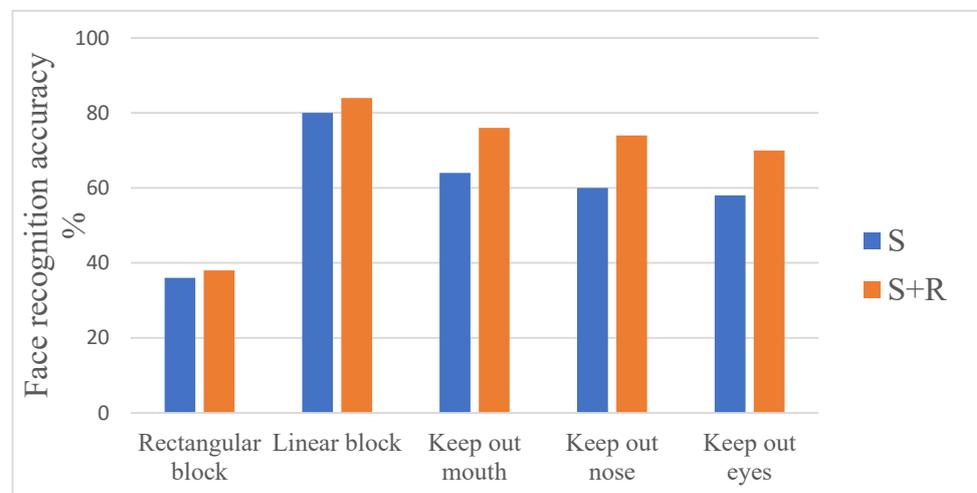
Figure 8 compares the image restoration results of the two methods under different occlusion areas. The repair effect of the method in this paper is more real and smooth, and the repair effect of the DCGAN + CNN method makes the face look unreasonable, strange and not real enough, which makes the recognition effect poor, and the recognition rate drops. Through this comparison, the effectiveness and advantages of this method can also be proved from the side.

5.2. Network Performance Analysis

This paper proposes a joint optimization method of Regular Face loss and Softmax loss, which integrates the expression classification task based on Softmax loss and the metric learning task based on Regular Face loss into the proposed network. Softmax loss can make the model learn expression features, and the Regular Face loss can ensure that the expression features are learned by the model. Through multi-task learning, the advantages of the two tasks are brought into play, so as to improve the discrimination capacity and robustness of the face recognition. In order to verify the performance of the Regular Face loss function, this paper compares the recognition accuracy before and after repair when the Regular Face + SoftMax (abbreviated as R + S) loss function is used and only the SoftMax (abbreviated as S) loss function is used [31]. The experimental results are shown in Figure 9.

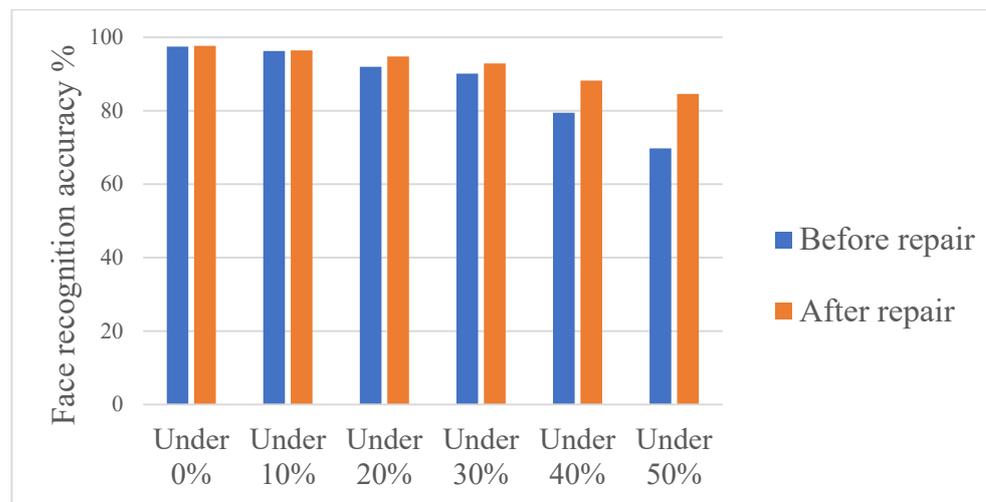


(A) Comparison of recognition accuracy before repair



(B) Comparison of recognition accuracy after repair

Figure 9. Cont.



(C) Comparison experiment of recognition rate before and after occlusion

Figure 9. Comparison of recognition accuracy after different levels of occlusion restoration.

It can be seen from Figure 9A that the face before repair has lost too many features. No matter which loss function is used, the recognition accuracy is very low, but the loss function of Regular Face + SoftMax is still partly improved compared to SoftMax. It can be seen from Figure 9B that the recognition accuracy after the repair has been significantly improved, and the improvement effect of using the Regular Face + SoftMax loss function is also more obvious, especially the occlusion of facial features. This is because Regular Face has the characteristic of quantifying the distance between classes. The effect of these difficult-to-classify features has been improved more obviously.

For face images with different occlusion areas, the filled face images are input into the face recognition network, and the faces are discriminated. The experimental results are shown in Figure 9C. It can be seen that the face filling model improves the accuracy of occluded expression recognition as a whole. When the occlusion area is less than 50%, the overall recognition rate of the model is still above 80%, which is close to 20% higher than the direct use of convolutional neural networks. When the occlusion area is small, the recognition accuracy of the model has a small improvement. Considering the processing time of the model, system resources and other factors, when the occlusion area is less than 10% of the face area, no filling is required, and the convolutional neural network is directly used for simple classification and recognition [32].

In this paper, the PCA + SVM method, sparse representation-based classification (SRC) and CNN, DCGAN + CNN methods are selected for comparison. In the DCGAN + CNN method, DCGAN is used to fill the occluded face image, and the CNN is a fine-tuned VGG model. All methods used for facial expression classification were tested on the CK + data set, and the results are shown in Table 3. It can be seen from the table that regardless of whether the face image is occluded, the facial expression recognition rate of the method in this paper is high. Although the method based on DCGAN + CNN also fills in the face image, the coherence of the image after the repair is poor, which affects the accuracy of expression recognition [33], especially the occlusion area that is less than 40%. The accuracy of facial expression recognition of this method is even lower than that of CNN alone. The images generated by the method in this paper are truly coherent and have little impact on facial expression recognition. Therefore, when the occlusion area of the face is 50%, a recognition accuracy rate of more than 80% can still be achieved.

Table 3. Comparison experiment with other algorithms (recognition accuracy rate) %.

Method	Unobstructed	Random Occlusion 10%	Random Occlusion 20%	Random Occlusion 30%	Random Occlusion 40%	Random Occlusion 50%
PCA + SVM	91.24	89.41	85.94	84.13	79.23	68.88
SRC	92.56	90.34	87.95	95.23	79.82	55.82
CNN	97.65	94.92	92.75	88.96	78.11	67.57
DCGAN + CNN	97.65	92.76	89.95	82.45	78.82	70.24
Paper model	97.65	96.45	94.78	92.89	92.23	84.56

6. Discussion

In face recognition, there are still many challenges and difficulties for researchers to solve. For example, regarding facial expression recognition in real scenes, there are research gaps for face recognition based on different expressions. The expression of facial emotion may vary with region, culture and environment. There are differences in personality [34] as well. Therefore, facial expression recognition technology needs to be improved in many ways to achieve better results. Deep learning has recently become a very hot research topic. Using the convolutional layer, pooling layer and full connection layer of a convolutional neural network, we can let this network structure study and detect relevant features by itself, and put them to use. This feature is very convenient for research, and can omit the very difficult modeling process. In addition, there are various research gaps in image classification, object detection, pose estimation and image segmentation. Furthermore, although deep learning has a wide sphere of applications and strong versatility, we should continue to try to expand it to other applications. In summary, deep learning still has a lot of potential to explore and discover, and we need to develop more algorithms to fill these research gaps. This paper proposes a GAN model based on a dual-discrimination network to realize occlusion facial expression recognition. The dual-discrimination network structure can generate real and coherent high-quality images in the face restoration stage, at which point a high recognition rate can be achieved. At the same time, the Regular Face loss function is introduced in the recognition stage, which effectively increases the distance between different classes and reduces the distance between classes. In the restoration stage, the face image data set Celeb A is used to pre-train its weight parameters.

Experiments have proved that the method in this paper is more realistic and coherent in the repaired face image, and then facial expression recognition is performed on it after the repair, and the recognition rate is higher than other methods. However, the method in this paper also has certain shortcomings. It is not particularly ideal for large-area occlusion, especially for images with more than 50% occlusion. The large error rate of the repair results is not good for random occlusion and daily occlusion.

Because a black shadowing mask was used to occlude the original image during the training method in this paper, the model learned only to repair the black area, which is not suitable for non-manually added occlusion, and the model training is based on closed set implementation, which is not good for generalization of random occlusion. In addition, the training of GAN network is relatively free and the training is difficult to control. This is mainly due to its adversarial function. Therefore, we plan to study and improve the defects of the method in this paper in future.

7. Conclusions

In our research, face recognition using machine learning is based on the convolutional layer, pooling layer and full connection layer of a convolutional neural network. In practice, we can enable this network structure to study and detect relevant features by itself, and perform effective face recognition. This technique allows us to resolve the difficult modeling process that affects face recognition research, which did not perform well in the era before deep learning methods emerged. Furthermore, based on our generative confrontation network, the deep learning face recognition framework in our paper has greatly improved

in image classification. We know that our deep learning technique still has a lot of potential that we can explore to enhance the increased accuracy and efficiency.

Author Contributions: Software, B.W.; Supervision, Z.Z. Writing—original draft, H.G.; Writing—review and editing, Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China (No. 62006102).

Institutional Review Board Statement: Not available.

Informed Consent Statement: No consent is required.

Data Availability Statement: To be supplied upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. He, C. Overview of Face Recognition Technology. *Intell. Comput. Appl.* **2016**, *6*, 112–114.
2. Chen, S.; Liu, Y.; Gao, X.; Han, Z. MobileFaceNets: Efficient CNNs for Accurate Real-time Face Verification on Mobile Devices. In *Proceeding of Chinese Conference on Biometric Recognition*; Springer: Cham, Switzerland; Berlin, Germany, 2018; pp. 428–438.
3. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-time Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
4. Zhu, M.H.; Li, S.T.; Ye, H. Facial expression recognition method based on sparse representation. *Pattern Recognit. Artif. Intell.* **2014**, *27*, 708–712.
5. Xue, Y.; Mao, X.; Catalin-Daniel, C.; Lv, S. Robust facial expression recognition method under occlusion conditions. *J. Beijing Univ. Aeronaut. Astronaut.* **2010**, *36*, 429–433.
6. Tang, Z.; Li, Y.; Chai, X.; Zhang, H.; Cao, S. Adaptive Nonlinear Model Predictive Control of NOx Emissions under Load Constraints in Power Plant Boilers. *J. Chem. Eng. Japan.* **2020**, *53*, 36–44. [[CrossRef](#)]
7. Ou, J. Classification Algorithms Research on Facial Expression Recognition. *Phys. Procedia* **2012**, *25*, 1241–1244. [[CrossRef](#)]
8. Wang, W.; Yang, J.; Xiao, J.; Li, S.; Zhou, D. Face Recognition Based on Deep Learning. In *Proceeding of the International Conference on Human Centered Computing*; Springer: Cham, Switzerland; Berlin, Germany, 2015; pp. 812–820.
9. Grm, K.; Štruc, V.; Artiges, A.; Caron, M.; Ekenel, H.K. Strengths and weaknesses of deep learning models for face recognition against image degradations. *IET Biom.* **2018**, *7*, 81–89. [[CrossRef](#)]
10. Zhao, J.; Lv, Y.; Zhou, Z.; Cao, F. A novel deep learning algorithm for incomplete face recognition: Low-rank-recovery network. *Neural Netw.* **2017**, *13*, 94. [[CrossRef](#)]
11. Yu, C.; Pei, H. Face recognition framework based on effective computing and adversarial neural network and its implementation in machine vision for social robots. *Comput. Electr. Eng.* **2021**, *92*, 107128. [[CrossRef](#)]
12. Yu, Y.; Huang, F.H. Face Recognition with Noise Based on Fast PCA and Simplified PSO. *J. Qinghai Univ.* **2020**, *38*, 49–57.
13. Zhang, Z.; Yang, Q. Target recognition method based on BP neural network and improved DS evidence theory. *Comput. Appl. Softw.* **2018**, *35*, 6151–6156.
14. Tang, Z.; Zhao, G.; Ouyang, T. Two-phase deep learning model for short-term wind direction forecasting. *Renew. Energy* **2021**, *173*, 1005–1016. [[CrossRef](#)]
15. Li, X.X.; Liang, R.H. An overview of occluded face recognition: from subspace back to depth learning. *Chin. J. Comput.* **2018**, *41*, 177–207.
16. Goel, T.; Murugan, R. Deep Convolutional—Optimized Kernel Extreme Learning Machine Based Classifier for Face Recognition. *Comput. Electr. Eng.* **2020**, *85*, 106640. [[CrossRef](#)]
17. Jing, C.K.; Song, T.; Zhuang, L. A review of face recognition technology based on deep convolutional neural network. *Comput. Appl. Softw.* **2018**, *35*, 223–231.
18. Yu, T.; Tong, Y.; Cao, X.H. An Occlusion Face Recognition Algorithm Based on Iterative Weighted Low-rank Decorption. *Comput. Technol. Dev.* **2019**, *29*, 42–46.
19. Gou, J.; Song, J.; Ou, W.; Zeng, S.; Yuan, Y.; Du, L. Representation-based classification methods with enhanced linear reconstruction measures for face recognition. *Comput. Electr. Eng.* **2019**, *79*, 106451. [[CrossRef](#)]
20. Wang, W.M.; Tang, Y.; Zhang, J.; Zhang, Y.Q. Face Recognition Algorithm Based on Convolutional Neural Network Feature Fusion. *Comput. Digit. Eng.* **2020**, *1*, 88–92.
21. Sun, X.; Pan, T.; Ren, F.J. Facial Expression Recognition Based on ROI-KNN Convolutional Neural Network. *Acta Autom. Sin.* **2016**, *42*, 883–891.
22. Zhang, F.Y.; Wang, X.; Xu, X.Z. Structured Occlusion Coding and Extreme Learning Machine Local occlusion face recognition. *Comput. Appl.* **2019**, *39*, 2893–2898.
23. Cao, Z.; Yang, Y.; Qi, Y. Image restoration method based on multi-loss constraint and attention block. *J. Shaanxi Univ. Sci. Technol.* **2020**, *38*, 158–165.

24. Chen, J.Z.; Wang, J.; Gong, X. Face map based on cascading generation adversarial network. *J. Univ. Electron. Sci. Technol. China* **2019**, *48*, 910–917.
25. Feng, X.R.; Hui, K.H.; Liu, Z.D. Face recognition based on convolution feature and Bayesian classifier. *J. Intell. Syst.* **2018**, *73*, 101–107.
26. Qiu, D.; Zheng, L.; Zhu, J.; Huang, D. Multiple improved residual networks for medical image super-resolution. *Futur. Gener. Comput. Syst.* **2020**, *116*, 200–208. [[CrossRef](#)]
27. Zhao, S.W.; Zhang, R.X.; Wang, Y.M. Review of biometric identification technology. *China Secur.* **2015**, *7*, 79–86.
28. Hu, L.Q. Research on Feature Extraction Algorithm of Face Recognition under Complex Conditions. Master's Thesis, Donghua University, Shanghai, China, 2016.
29. Lin, Z.M. Research on Improved PSO Face Recognition Algorithm. *J. Nat. Sci. Harbin Norm. Univ.* **2020**, *36*, 51–55.
30. Zhang, D.F.; Gao, N.H.; Wang, H.; Feng, X.H.; Huo, J.W.; Zhang, J. Based on block LBP fusion feature And SVM face recognition algorithm. *Transducers Microsyst.* **2019**, *38*, 154–156.
31. Liu, L.; Li, D. Patents Analysis of Face Recognition Technology in Surveillance System. *Sci. Technol. Innov. Appl.* **2021**, *2*, 46–47.
32. Duan, J.W. Research on face recognition technology in big data environment. *Electron. World* **2019**, *1*, 185–187.
33. Chen, K.; Xing, X.Y.; Tian, X.Y.; Wang, S. Design of face recognition access control system based on CNN. *J. Xuzhou Inst. Technol. (Nat. Sci. Ed.)* **2018**, *33*, 89–92.
34. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [[CrossRef](#)]