Facial Recognition Technology for Scenic Spot Monitoring Based on U^{2}Net and FFC

Yiyi Liu
Tourism Management Department, Zhengzhou Tourism College, Zhengzhou 450000, China
E-mail: lyylxflqh@163.com

Keywords: scenic spot monitoring, facial recognition, U^{2}Net, fast fourier convolution, global convolution, mask learning

Received: June 30, 2025

To ensure the safety of scenic spots and achieve intelligent management of scenic spots, a face recognition method based on U^{2}Net and FFC is proposed to achieve monitoring face recognition under different occlusion conditions. It consists of a small area regular occlusion face recognition model and a large area irregular occlusion face recognition model. Firstly, a face recognition model grounded on an improved residual network-U^{2}Net is raised to address the problem of small area rule occlusion. This model combines a global convolution module, a feature pyramid network, and a mask learning unit. When evaluating facial recognition methods, multiple evaluation metrics were used, including recognition accuracy, F1-score, recognition rate, structural similarity index, peak signal-to-noise ratio, learning perceptual image block similarity, and Frecht approximation distance. These indicators evaluate the performance of the model under small and large area irregular occlusion conditions from different perspectives, ensuring the comprehensiveness and reliability of the evaluation. The findings denote that the average recognition accuracy of the enhanced residual network-U^{2}Net is as high as 98.7%, the average F1-score is 0.983, and the average recognition rate is 99.5%. Secondly, in response to the problem of large-scale irregular occlusion in facial recognition, a fast Fourier convolution generative adversarial network is proposed, which combines generative adversarial network and Fourier feature convolution to repair and recognize facial images. The outcomes denote that the average structural similarity index and peak signal-to-noise ratio of the model are 0.878 and 34.7dB, respectively, and the average accuracy and recognition rate are 91.0% and 92.6%, respectively. The above results denote that the proposed facial recognition method exhibits superior performance under different occlusion conditions and can effectively promote the intelligent development of scenic area management.

Povzetek: Predstavljena je metoda prepoznavanja obrazov na podlagi U2Net in FFC za inteligentno nadzorovanje v turističnih krajih, ki omogoča prepoznavanje tudi pri zakritih obrazih (maske, klobuki).

1 Introduction

As the global tourism industry quickly develops in recent years, the number of tourists in scenic spots has shown explosive growth, which has put forward higher requirements for the management and service of scenic spots. Due to the low efficiency of traditional manual management models and their inability to cope with the complex and changing challenges of scenic areas, coupled with the dense population and high mobility of people in scenic areas, which significantly increase the difficulty of safety management, it is particularly important to introduce advanced safety management technologies to improve the level of safety management and management efficiency in scenic areas [1-2]. Among numerous security management technologies, facial recognition technology has gradually become an important tool for scenic spot security management due to its high efficiency, convenience, and accuracy. Through facial recognition systems, scenic spots can achieve real-time monitoring, rapid identification, and precise management of personnel, effectively enhancing emergency response capabilities and ensuring the safety of tourists [3-4]. However, the complex environment of scenic spots, frequent personnel flow, and often the presence of obstructions greatly increases the difficulty of facial recognition. However, existing facial recognition technologies have low recognition accuracy when dealing with occlusion problems, making it difficult to meet practical needs.

Qin et al. proposed a multi-purpose algorithm called SwinFace-based on Swin Transformer to address the issue of neglecting task collaboration during the training process of facial recognition models. This method integrated multi-level channel attention modules in each task-specific analysis subnet with the objective of achieving adaptive feature selection. The findings demonstrated that the facial expression recognition and age estimation performance of this method surpassed that of existing methods [5]. Al-Dabbas et al. developed a facial recognition method that utilized classification, machine learning and deep learning models to address the issue of rising counterfeit crime rates. The methodology

employed involved the utilization of Viola Jones, linear discriminant analysis, mutual information, and analysis of variance techniques to construct two facial classification systems. The findings showed that the classification accuracy of both facial classification systems was above 96%, indicating that the proposed model performed well in both accuracy and processing time [6]. Gao et al. proposed the first privacy preserving facial recognition protocol for recognition stage computation in intelligent security systems to address privacy protection and identity recognition efficiency issues. This method introduced a Householder matrix into blind user data, enabling the protocol to support privacy protected facial recognition on semi trusted edge servers. The results showed that the protocol not only protected the privacy of user data, but also could achieve rapid response of large-scale face recognition (FR) through edge computing, effectively improving the efficiency of FR in intelligent security systems [7]. Xie et al. proposed a general privacy protection framework for FR systems that is grounded on edge computing. The purpose of this framework was to address the issue of data privacy leakage that has been identified in such systems. The overarching objective of the proposed framework is to safeguard the confidentiality of facial images and training models by employing a local differential privacy algorithm. The algorithm under discussion is founded upon a comparison of the proportion of feature information. As previously stated, the aforementioned text is concerned with the implementation of identity authentication and hashing techniques, with a view to confirming the legitimacy of terminal devices. The results showed that in numerical experiments, this scheme could ensure the optimal balance between the usability and privacy protection of the facial recognition system [8].

U^2Net, as a deep learning model for image segmentation, combines an encoder and decoder, and introduces a cyclic squeezing unit, which can effectively extract image features of different scales. Therefore, it has significant advantages in image feature extraction. Feng et al. designed a detection method based on crack-U^2Netto address the accuracy issue of road crack detection. This method utilized the U^2Net architecture for feature learning and introduced a geometry-based data augmentation strategy to address the issue of insufficient training data. The results showed that the accuracy of Crack-U^2Net in highway crack detection reached 95.8%, which is superior to existing methods [9]. Shi et al. proposed the U^2CrackNet detection method for road crack detection. The proposed methodology involved the extraction of crack features through the encoding layer, followed by the establishment of a connection between the encoder and decoder via the atrous spatial pyramid pool model, with the objective of capturing multi-scale crack information. The results showed that U^2CrackNet could obtain clearer and more continuous highway cracks, with a detection accuracy of 98.95% [10]. Li et al. proposed a U^2Net-based analysis method to address the issue of low efficiency caused by manual operation in microscope image analysis. This method enhanced the model's ability to extract key information by introducing a convolutional block attention module, and achieved model lightweighting by introducing Ghost convolution. The outcomes denoted that the prediction accuracy of the method model increased from 92.24% to 97.13% [11]. Zheng Z and Yang K proposed a detection method that integrates You Only Look Once version 5 (YOLOv5) and U^2Net for wall crack detection. This method utilized the GhostNet module to optimize YOLOv to improve its training speed, while introducing U^2Net to perform binary classification on the region input extracted by YOLOv5 to enhance the final classification performance. The results denoted that this method could effectively address the issue of poor segmentation of crack targets in large environmental backgrounds [12].

In summary, although the current facial recognition models have high recognition accuracy, they are difficult to cope with the problem of obstructed facial recognition in complex environments of scenic spots. Therefore, to address the above issues, an FR model for different occlusion conditions has been proposed, which consists of two parts: a small area regular occlusion FR model and a large area irregular occlusion FR model. The innovation of the research lies in the combination of Residual Network (ResNet) and U^2Net, and the introduction of Global Convolution Module, Feature Pyramid Network (FPN), and Mask Learning Unit to improve the recognition accuracy of the model for small area regularly occluded faces. Specifically, a global convolution module consisting of two symmetric convolution layers is used to capture global features in both horizontal and vertical directions. At the same time, a mask learning unit is introduced in FPN to remove the features of occluded areas by generating multi-level masks to enhance feature representation. Secondly, by combining Fast Fourier Convolution (FFC) and Generative Adversarial Network (GAN), the problem of low recognition accuracy in large-area irregularly occluded face images can be solved by repairing them.

2 Methods and materials

Due to the influence of facial coverings such as hats, masks, and glasses, as well as changes in facial expressions, the success rate of existing surveillance facial recognition is low, making it difficult to effectively ensure the safety of scenic spots. Therefore, a monitoring FR model based on U^2Net and FFC is proposed to address the recognition problem under face occlusion. It consists of two parts: a small area occlusion FR model and a large area irregular occlusion FR model. Firstly, an FR method based on U^2Net and global convolution is constructed to address the problem of small-scale rule-based occlusion. For the problem of large-scale irregular occlusion, a FR model based on FFC and GAN is proposed.

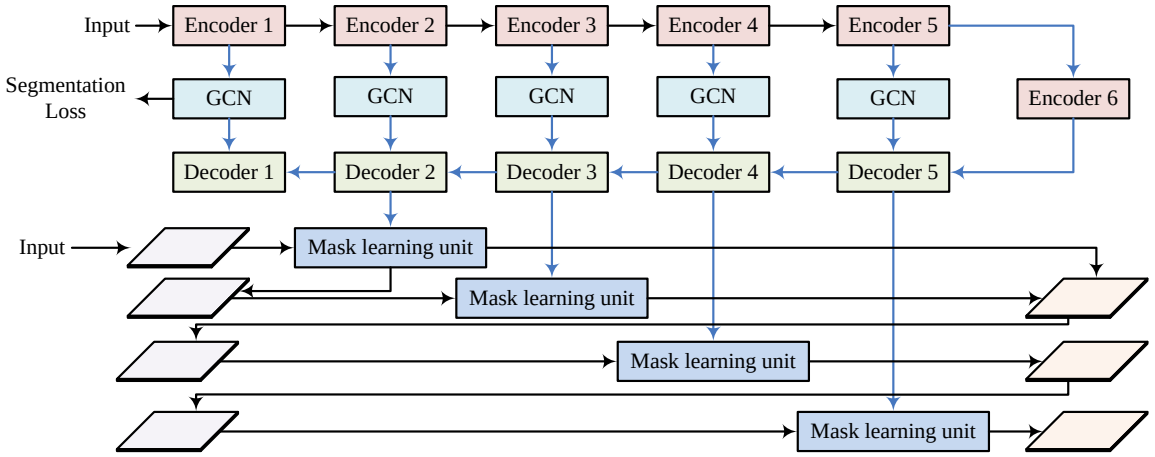


Figure 1: Face recognition model based on improved ResNet-U^{2}Net. (Source from: Author's self drawn)

2.1 Regular occlusion facial recognition model based on U^{2}Net and global convolution

It is difficult to extract facial features from surveillance cameras in conditions of obstruction by regular objects, such as masks, glasses and hats. The accuracy of facial recognition systems is consequently adversely affected. Due to the use of two parallel encoder decoder paths and the introduction of cyclic squeezing units, U^{2}Net is able to simultaneously process global and local features, thereby improving segmentation accuracy. Although the cyclic squeezing unit can capture multi-scale features, its receptive field size is small, which makes it impossible to fully cover all scale features [13-14]. Therefore, to expand the receptive field of U^{2}Net, global convolution is introduced and improved. The FR model based on improved ResNet-U^{2}Net is denoted in Figure 1.

In Figure 1, the FR model based on improved U^{2}Net consists of two parts: occlusion detection segmentation module and feature detection module. The model first generates a multi-level occlusion segmentation map through the occlusion detection module, then extracts image features through the FR module, and removes the influence of occlusion on facial features through the mask learning unit. Finally, FPN is used to fuse the features of each stage. In the feature extraction module, the selected backbone network is ResNet, which can achieve feature reuse through skip connections. However, due to the poor ability of ResNet to extract multi-scale features, it will reduce the accuracy of the model's FR. Therefore, to enhance the multi-scale feature extraction capability and model generalization performance, the FPN module is introduced in the study. FPN upsamples high-level feature maps to the resolution of low-level feature maps through a top-down path, thereby generating a multi-scale feature pyramid. Moreover, feature maps of different scales are fused through horizontal connections to enhance the richness of feature representation. At the same time, the generated feature pyramid can capture both global and local information, improving the model's ability to extract multi-scale features [15-16]. For improved ResNet-U^{2}Net, the input image needs to be first detected and aligned by the method based on the cascaded multi task framework, and then the image size is adjusted to 112 * 112. Next, facial recognition can be performed using the improved ResNet-U^{2}Net. The Batchsize of the model is 128, the initial learning rate is 0.1, the hypersphere radius s is 64, and the spacing m is 0.48. During the training, the learning rate is adjusted to 1/10 of its original level at the 11th, 20th, and 30th epochs. However, FPN is prone to information loss, which can lead to feature damage. Therefore, to improve the above problems, the structure of FPN is optimized by introducing mask learning units to avoid the influence of damaged features. The formula for calculating the feature pyramid is denoted in equation (1).

$$\begin{cases} P_3 = M_3 + ds(M_2) \\ P_4 = M_4 + ds(P_3) \\ P_5 = M_5 + ds(P_4) \end{cases} \tag{1}$$

In equation (1), P_i means the feature pyramid; M_i represents the features obtained after mask operation; ds represents downsampling operation. The calculation formula for M_i is denoted in equation (2).

$$M_i = X_i + Mask_i \otimes X_i \tag{2}$$

In equation (2), $Mask_i$ represents the mask; X_i represents the original input. For the occlusion detection and segmentation module, its backbone network is U^{2}Net, which consists of U-shaped residual modules. Unlike ordinary residual modules, the U-shaped residual module replaces the convolutional layers in the original residual module with U-blocks and replaces the original features with local features to achieve multi-scale feature extraction. The so-called U-block refers to the U-shaped encoder decoder structure. Considering the complexity of the model, the amount of U-shaped residual modules is 6 [17]. Due to the small receptive field of U^{2}Net, a global convolution module is introduced to expand its receptive field size. The structure of the global convolution module is denoted in Figure 2.

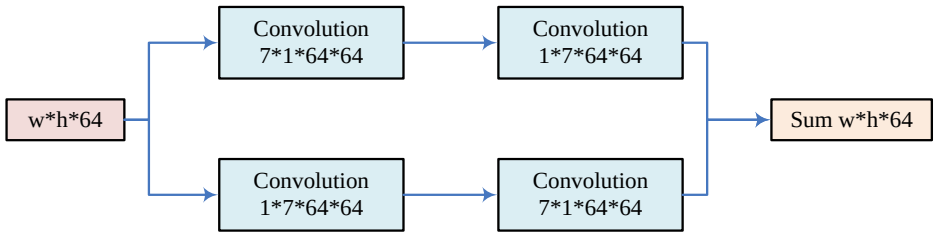


Figure 2: Structure of the global convolution module. (Source from: Author's self drawn)

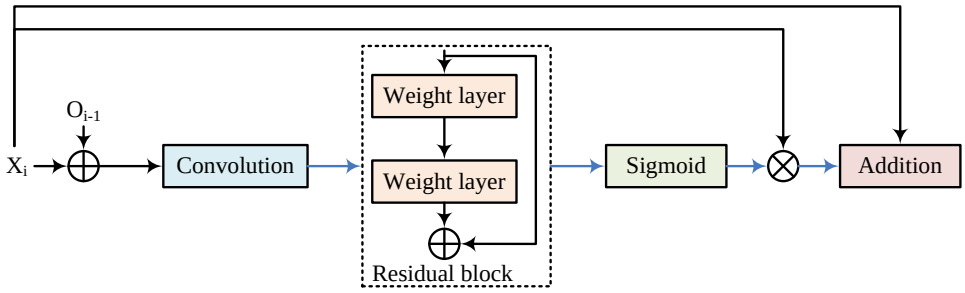


Figure 3: Structure of the mask learning unit. (Source from: Author's self drawn)

In Figure 2, the global convolution module mainly contains four symmetrically distributed convolution layers. This structure enables the global convolution module to capture features of different scales and enhance the model's understanding of the global information of the image through feature fusion. In addition, to avoid damaging the features and affecting the performance of the occlusion detection segmentation module and feature detection module, a mask learning unit is introduced in the study. Although generating masks can remove the features of occluded areas, these methods usually only generate masks at a single scale and cannot effectively handle multi-scale features. Moreover, the mask learning unit can generate multi-level masks, corresponding to feature maps of different scales, thus more comprehensively handling occlusion problems. It suppresses the features of occluded areas through masking while preserving the features of unobstructed areas, enhancing the robustness of feature representation. The structure of the mask learning unit is denoted in Figure 3.

In Figure 3, the mask learning unit mainly contains convolutional layers, residual modules, and sigmoid functions. This module first concatenates the feature and occlusion segmentation representations of each stage, and then processes the concatenated images using convolutional layers and activation functions to generate multi-level masks. Finally, the generated mask is used to remove the features of the occluded area and added to the original input features to enhance the feature representation. The formula for mask learning calculation is shown in equation (3).

$$Mask_i = \text{Sigmoid}\left(\gamma\left(c(concat[F_i, S_{i-1}])\right)\right) \quad (3)$$

In equation (3), γ represents residual operation; F_i represents the characteristics of each stage; S_i represents occlusion segmentation representation. The loss function (LF) of the model is denoted in equation (4).

$$L = L_{fc} + L_{seg} \quad (4)$$

In equation (4), L means the overall LF of the model; L_{fc} denotes the face classification LF; L_{seg} denotes the face segmentation LF. The calculation formula for the face classification LF is denoted in equation (5).

$$L_{fc} = -\frac{1}{B}\sum_{i=1}^{B}\ln\frac{e^{\|x_i\|\left(\cos(\theta_{y_i}+s)\right)}}{e^{\|x_i\|\left(\cos(\theta_{y_i}+s)\right)} + \sum_{j=1, j\neq y_i}^{N} e^{\|x_i\|\cos\theta_j}} \quad (5)$$

In equation (5), B represents batch size; x_i represents the feature vector; s represents spacing; N means the number of categories; θ_j means the angle between the weight and the feature vector. The face segmentation LF is shown in equation (6).

$$L_{seg} = \frac{1}{|N|}\sum_{c\in C}\left[\varepsilon D_{KL}\left(y\|\hat{p}_c\right) + \frac{\delta}{|c|}\sum_{s\in c} D_{KL}\left(\hat{p}_c\|p(s)_c\right)\right] \quad (6)$$

In equation (6), ε and δ both represent hyperparameters; D_{KL} represents Kullback-Leibler divergence; $\hat{p}_c$ means the probability distribution of category c; $p(s)$ represents the probability distribution of the category c of real data.

2.2 Irregular occlusion face recognition model based on FFC and GAN

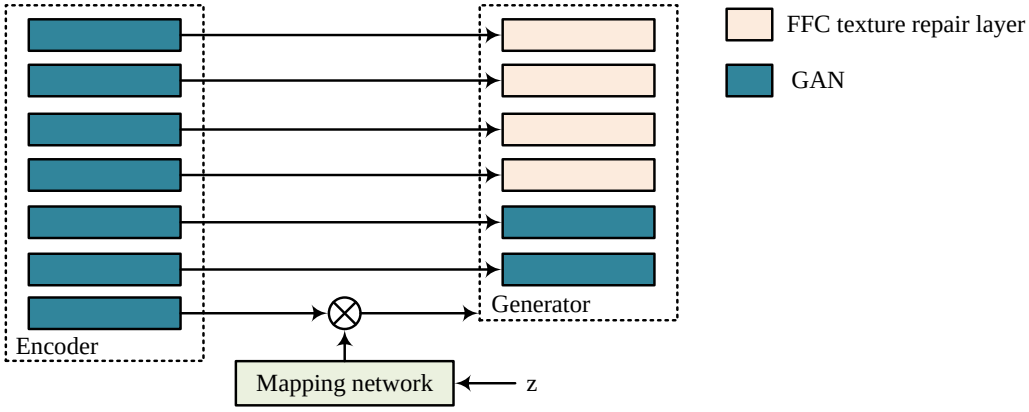


Figure 4: Monitoring face recognition model based on FFC-GAN. (Source from: Author's self drawn)

Although the above method can achieve FR with small area regular occlusion, due to the large flow of people and complex environment in scenic spots, there are cases where faces are obstructed by large irregular objects, which further increases the difficulty of FR. Therefore, to effectively ensure the safety of scenic spots, a large-scale irregular occlusion FR model based on FFC and GAN is proposed. Compared to other convolution methods, FFC can effectively accelerate the training and inference process of networks when processing large image or video data [18-19]. GAN can generate high-quality synthetic data to achieve the restoration of large areas of irregularly occluded faces. The monitoring FR model based on FFC-GAN is denoted in Figure 4.

As shown in Figure 4, the model first uses GAN to repair irregularly occluded facial images, and generates facial image structures using encoding and hidden layer noise vectors. Then FFC is utilized to generate texture details of the image to raise the quality of facial image restoration. Finally, the model is jointly trained using an identity preservation LF to raise the accuracy of FR. For GANs, the input is a random noise vector. This is mapped to the data space through a series of neural network layers to generate fake data. The required style parameters are then generated based on affine changes. The formula for generating style parameters is shown in equation (7).

$$s = A\left(M\left(h\right)\right) \tag{7}$$

In equation (7), s represents the style parameter; A represents affine transformation; M stands for Mapping Network; h stands for hidden layer vector. Although the above method can achieve the restoration of occluded images, it may result in inconsistency between the restored image and the original image. Therefore, to solve the above problems, collaborative modulation methods are introduced in the research. The formula for generating style parameters for collaborative modulation is shown in equation (8).

$$s = A\left(E\left(x\right), M\left(h\right)\right) \tag{8}$$

In equation (8), E represents the image conditional encoder; x represents the input image. It is worth noting that the generator and discriminator of GAN need to be trained alternately. The goal of the generator is to generate restored images that are as close to the real image as possible, while the goal of the discriminator is to distinguish between the generated image and the real image. Therefore, in each iteration, the generator and discriminator update their parameters separately to minimize the adversarial LF. The Batch size of GAN is 24, with an initial learning rate of 0.002, and the learning rate is adjusted to 0.001 after 650000 iterations. The weight of reconstruction loss is 10, and the weight of identity preservation loss is 10. The above method can achieve the restoration of large-area occluded images, but due to the loss of texture details in the restored images, it seriously affects the success rate of FR. Therefore, to achieve the restoration of image texture details, the FFC module is introduced in the study. Although existing texture restoration methods, such as convolution-based restoration methods, can generate certain texture details, they have low efficiency in processing large-scale images and are difficult to effectively capture global features. FFC, through the fusion of global and local features, can generate higher quality texture details and improve the quality of restored images. The structure of FFC is shown in Figure 5.

As shown in Figure 5, FFC consists of global branches and local branches. FFC first splits the input features into global features and local features, where global features are processed through convolutional layers and Spectral Transformers, and local features are processed using two convolutional layers [20-21]. Next, the processed local features and global features are fused, and after batch normalization and ReLU processing, the output features can be obtained. The formula for calculating the local output features of FFC is denoted in equation (9).

$$Y^{l} = Y^{l\to l} + Y^{g\to l} = Fou_{l}\left(\mathrm{X}^{l}\right) + Fou_{g\to l}\left(\mathrm{X}^{g}\right) \tag{9}$$

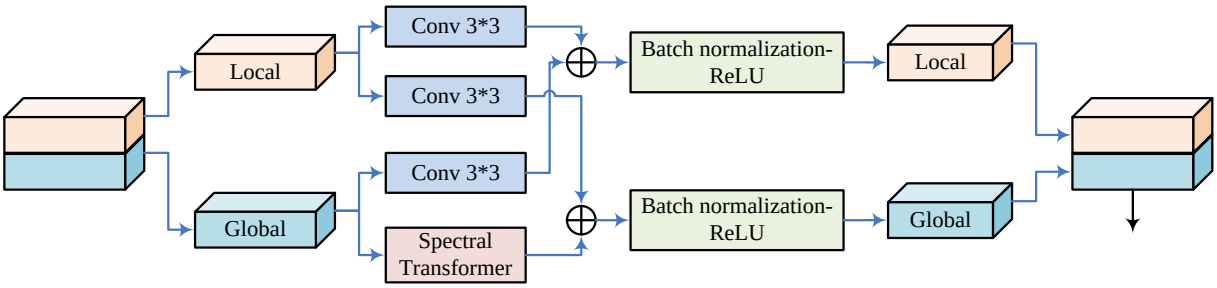


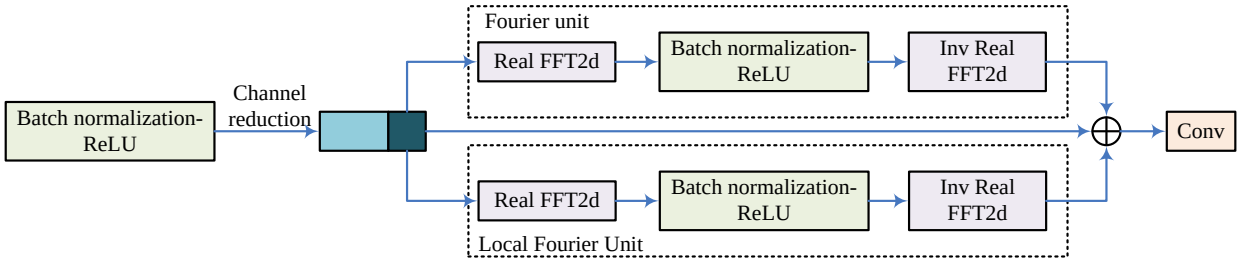
Figure 5: Structure of FFC. (Source from: Author's self drawn)

In equation (9), Y^l represents the output characteristics of local branches; $Y^{l\to l}$ represents small-scale feature components of local branches; $Y^{g\to l}$ represents the multi-scale receptive field components exchanged from global branches to local branches; Fou_l and $Fou_{g\leftarrow l}$ both represent fast Fourier transform (FFT); X^l and X^g represent the input features of local and global branches, respectively. The formula for calculating the global output characteristics of FCC is shown in equation (10).

$$Y^g = Y^{g\to g} + Y^{l\to g} = Fou_g\left(X^g\right) + Fou_{l\to g}\left(X^l\right) \quad (10)$$

In equation (10), Y^g represents the output feature of the global branch; $Y^{g\to g}$ represents the small-scale feature components of the global branch; $Y^{g\to g}$ represents the multi-scale receptive field components exchanged from local branches to global branches; Fou_g and $Fou_{l\leftarrow g}$ both represent FFT. The structure of the Spectral Transformer in FCC is shown in Figure 6.

In Figure 6, the Spectral Transformer includes convolutional layers, Fourier units, and local Fourier units. Firstly, Spectral Transformer processes input information through convolutional and batch normalization layers, and then captures global and local features using Fourier units and local Fourier units, respectively, and fuses the features. Finally, the captured features can be output after being convolved again. The Fourier unit and local Fourier unit are both composed of real 2D FFT, convolutional layer, and inverse real two-dimensional FFT. The real 2D FFT is responsible for transforming spatial features into the spectral domain, the convolutional layer is responsible for updating spectral data, and the inverse real 2D FFT is responsible for restoring spatial features [22-23]. By using the above method, FFC is constructed, and after combining it with convolutional layers, a texture restoration module based on FFC can be constructed. The structure of the texture restoration module based on FFC is shown in Figure 7.



Note: Real FFT2d represents Real 2D Fast Fourier Transform, Inv Real FFT2d represents Inverse Real 2D Fast Fourier Transform

Figure 6: Structure of spectral transformer. (Source from: Author's self drawn)

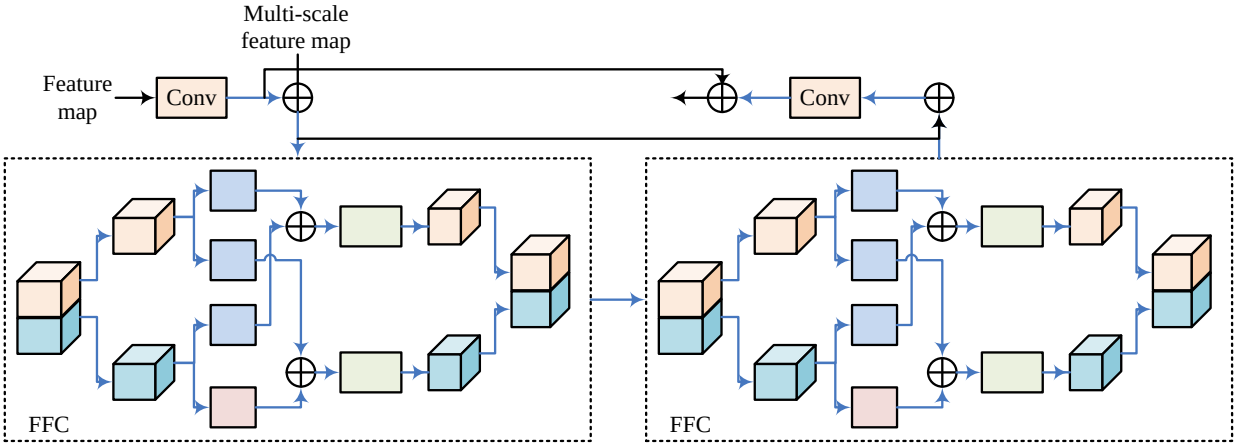


Figure 7: Texture repair module based on FFC. (Source from: Author's self drawn)

As shown in Figure 7, the FFC-based texture restoration module consists of convolutional layers and FFC residual structures. This module first processes the output features of the previous stage through convolutional layers and fuses them with the feature maps extracted by the image condition encoder. Next, the fused feature maps are subjected to contextual information extraction and fusion using the FFC residual structure, and the information is processed using convolutional layers. Finally, the processed information is fused with the features processed by the first convolutional layer to achieve texture restoration of the image. A monitoring FR model based on FFC-GAN is constructed using the above method, and the LF of the model is denoted in equation (11).

$$L_G = L_{gen} + L_{ref} + L_{id} \tag{11}$$

In equation (11), L_G denotes the overall LF of FFC-GAN; L_{gen} represents the adversarial LF of the generator; L_{gen} represents the reconstruction LF; L_{id} stands for identity preservation LF. The calculation formula for the adversarial LF is denoted in equation (12).

$$L_{gen} = -\mathrm{E}_{I_{res}}\left[\log \mathrm{D}\left(I_{res}\right)\right] \tag{12}$$

In equation (12), $E_{I_{res}}$ represents the expected value of the restored image; D stands for discriminator; I_{res} represents the restored image. The calculation formula for the reconstruction LF is denoted in equation (13).

$$L_{ref} = \alpha \left\| I_{res} - I_{ori} \right\|_1 \tag{13}$$

In equation (13), α represents the weight of reconstruction loss; I_{ori} represents the original image. The identity preservation LF is shown in equation (14).

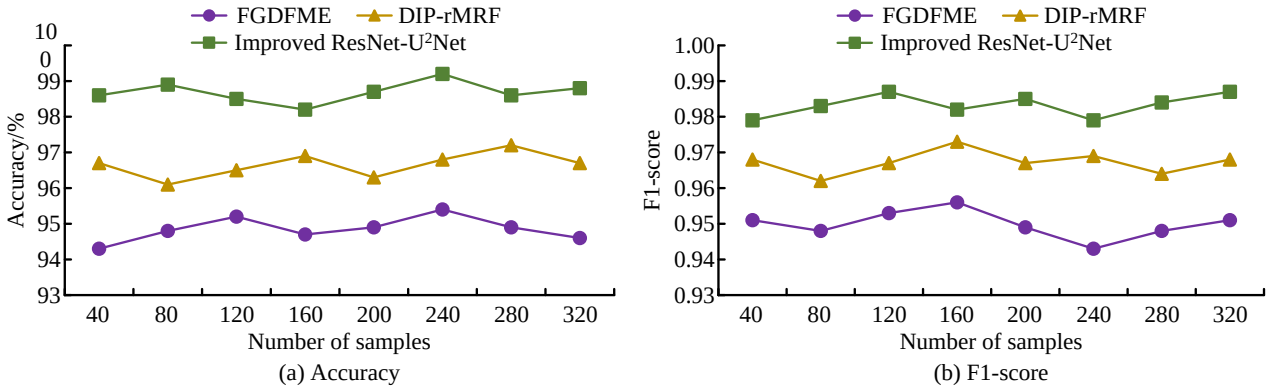
$$L_{id} = \beta \left\| F\left(I_{res}\right) - F\left(I_{ori}\right) \right\|_1 \tag{14}$$

In equation (14), β represents the weight of identity preservation loss; $F(.)$ represents the feature extraction process. The above method can achieve accurate recognition of faces with large areas of irregular occlusion.

3 Results

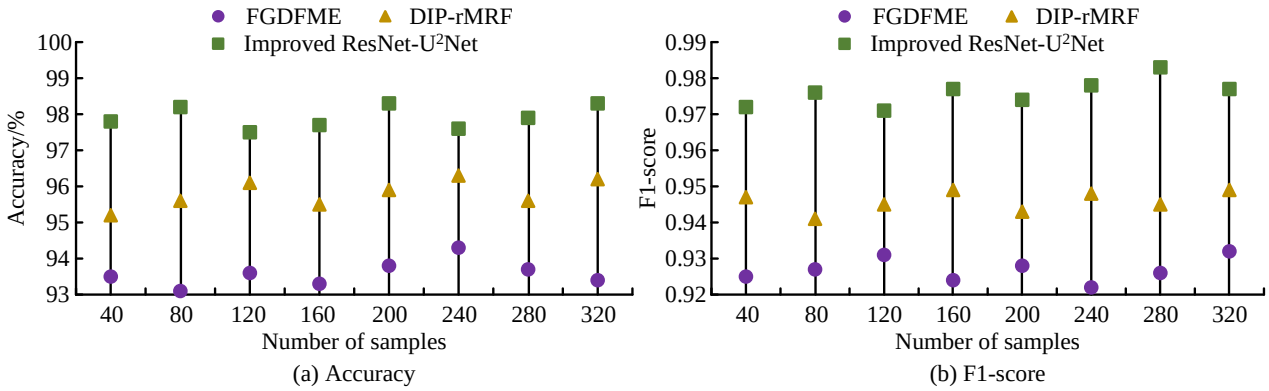
3.1 Small area occlusion face recognition test results

To test the recognition effect of the improved ResNet-U^{2}Net proposed in the study for small area regular occlusion faces, it was tested and compared with the Fine-Grained Deep Feature Mask Estimation (FGDFME) occlusion FR algorithm and the Depth Image Priors and Robust Markov Random Fields (DIP-rMRF) occlusion FR algorithm based on depth image priors and robust Markov random fields. The datasets used in the experiment were the Labeled Faces in the Wild (LFW) dataset and the Masked Faces in Real World for Face Recognition (MFR2) dataset used for FR in the real world. The LFW dataset contains 13233 facial images, covering 5749 individuals of different identities. Each image is labeled with the name of the corresponding person, with 1680 individuals having two or more images. Meanwhile, each image has a size of 250 * 250 pixels, with the majority being color images, but there are also a few black and white facial images. The MFR2 dataset contains the identities of 53 celebrities and politicians, with a total of 269 images. The size of each image is 160 * 160 * 3. To ensure the reliability of the experimental results, a simulated occlusion dataset was constructed using the LFW dataset, which involves adding objects such as masks, sunglasses, and mobile phones to mask facial images. The CPU utilized in the experiment was Intel core i7 4720HQ, with 16GB of memory and GeForce RTX 4060Ti GPU. The Batchsize and initial learning rate of the model were 128 and 0.1, respectively, and the radius and spacing of the hypersphere were 64 and 0.48, respectively. For each evaluation metric, the mean and standard deviation of multiple experimental results was calculated to assess the stability and reliability of the model performance. The 95% confidence interval to evaluate the confidence level of the model performance. The recognition accuracy and F1-score of each model in the simulated occlusion dataset are shown in Figure 8.



(a) Accuracy (b) F1-score

Figure 8: The recognition accuracy and F1-score of each model in the simulated occlusion dataset. (Source from: Author's self drawn)



(a) Accuracy (b) F1-score

Figure 9: Recognition accuracy and F1-score of each model in MFR2 data set. (Source from: Author's self drawn)

In Figure 8 (a), in the simulated occlusion dataset, the facial recognition accuracy of FGDFME and DIP rMRF was the highest at 95.4% and 97.2%, the lowest at 94.3% and 96.1%, and the average accuracy was 94.9% and 96.7%, respectively. The improved ResNet-U^{2}Net had a minimum FR rate of 98.2% and an average accuracy rate of up to 98.7%, which was higher than other algorithms. From Figure 8 (b), in the simulated occlusion dataset, the F1-score of FGDFME and DIP rMRF were the highest at 0.956 and 0.973, and the lowest at 0.943 and 0.962, respectively. The average F1-score was 0.950 and 0.967, respectively. The lowest F1-score of ResNet-U^{2}Net improvement was 0.979, with an average F1-score of 0.983. The above outcomes denoted that the improved ResNet-U^{2}Net had good performance in small area rule-based occlusion FR. The recognition accuracy and F1-score of each model in the MFR2 dataset are shown in Figure 9.

From Figure 9 (a), in the MFR2 dataset, the highest facial recognition accuracy of FGDFME and DIP rMRF was 94.3% and 96.3% respectively, the lowest was 93.1% and 95.2% respectively, and the average accuracy was 93.6% and 95.8% respectively. The improved ResNet-U^{2}Net had a minimum FR rate of 97.6% and an average accuracy rate of 97.9%, which was higher than other algorithms. From Figure 9 (b), in the MFR2 dataset, the highest F1-score for FGDFME and DIP rMRF were 0.956 and 0.973, and the lowest were 0.943 and 0.962, respectively. The average F1-score was 0.950 and 0.967, respectively. The lowest F1-score of ResNet-U^{2}Net improvement was 0.979, with an average F1-score of 0.983. The True Acceptance Rate (TAR) of each model in different datasets is shown in Figure 10.

According to Figure 10 (a), in the simulated occlusion dataset, the highest TAR of FGDFME and DIP rMRF were 96.2% and 98.3%, respectively, and the lowest were 95.3% and 97.1%, respectively. The average TAR was 95.7% and 97.6%, respectively. The TAR of ResNet-U^{2}Net was improved from a mini of 99.2% to a max of 99.9%, with an average TAR of 99.5%. According to Figure 10 (b), in the MFR2 dataset, the highest and lowest TARs for FGDFME and DIP rMRF were 95.3% and 96.8%, respectively, and 94.1% and 95.8%, respectively, with an average TAR of 94.6% and 96.3%. The TAR of the improved ResNet-U^{2}Net ranged from 98.1% to 99.9%, with an average TAR of 98.5%. The above results indicated that the improved ResNet-U^{2}Net had strong facial recognition capabilities and could effectively ensure the safety of scenic spots. To further analyze and improve the performance of ResNet-U^{2}Net, ablation experiments were conducted on it. The outcomes of the ablation experiment are denoted in Table 1.

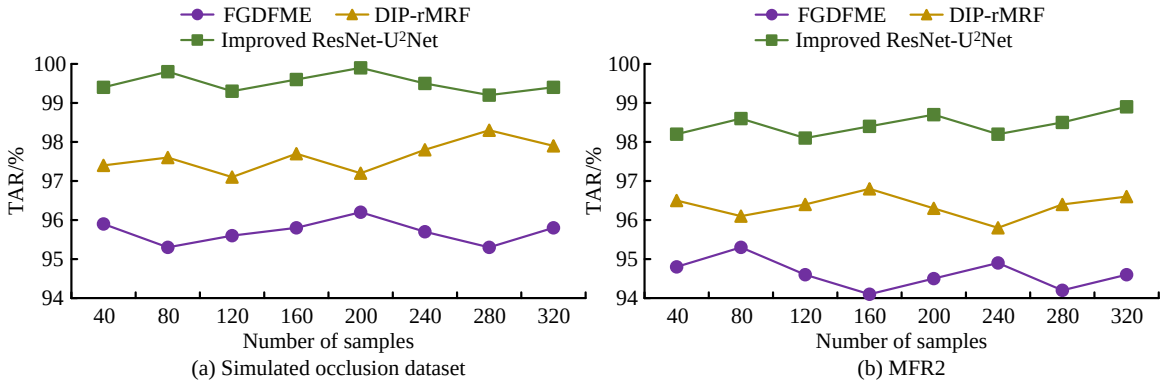


Figure 10: TAR of each model in different data sets. (Source from: Author's self drawn)

Table 1: Results of ablation experiments.

Model	ResNet	FPN	U²Net	Mask learning unit	Accuracy/%
1	√	×	×	×	92.2
2	√	√	×	×	93.1
3	√	×	√	×	94.2
4	√	×	×	√	94.5
5	√	√	√	×	95.6
6	√	√	×	√	96.7
7	√	×	√	√	97.4
8	√	√	√	√	98.7

According to Table 1, the facial recognition accuracy of the backbone network ResNet was only 92.2%. After introducing FPN, U²Net, and mask learning units, the facial recognition accuracy of the model significantly improved. Among them, U²Net and mask learning units had the most significant impact on model performance. After introducing the above two modules, the facial recognition accuracy of the model increased to 94.2% and 94.5%, respectively.

3.2 Large area occlusion face recognition test results

To test the effect of the proposed FFC-GAN in repairing and recognizing large-area irregularly occluded faces, it was tested and compared with the Partial Convolution and Multiscale Feature Fusion (PCMSF) facial image restoration model, Multiscale Feature Fusion U-Net (MSFFU-Net), Involution facial Feature Correction Network (IFFR-Net), and Depth Separable Convolution and Hypersphere Loss (DSCHL) occlusion model. The software and hardware settings of the experiment are consistent with the above experiment and will not be repeated. The dataset utilized in the experiment was the CelebA HQ dataset, which contains 30000 facial images with a resolution of 1024 × 1024. To simulate irregular occlusion situations, various shapes were randomly used to occlude facial images, with an occlusion rate of over 50%. In the experiment, the reconstruction loss weight and identity preservation loss weight were both 10, and the initial learning rate and Batchsize were 0.002 and 24, respectively. Firstly, the facial image restoration performance of FFC-GAN was tested. The Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) of different models are shown in Figure 11.

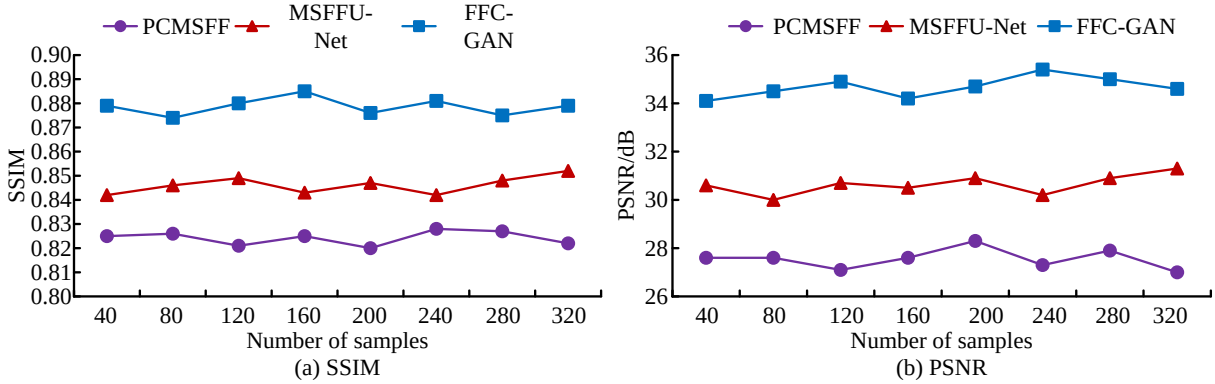


Figure 11: SSIM and PSNR of different models. (Source from: Author's self drawn)

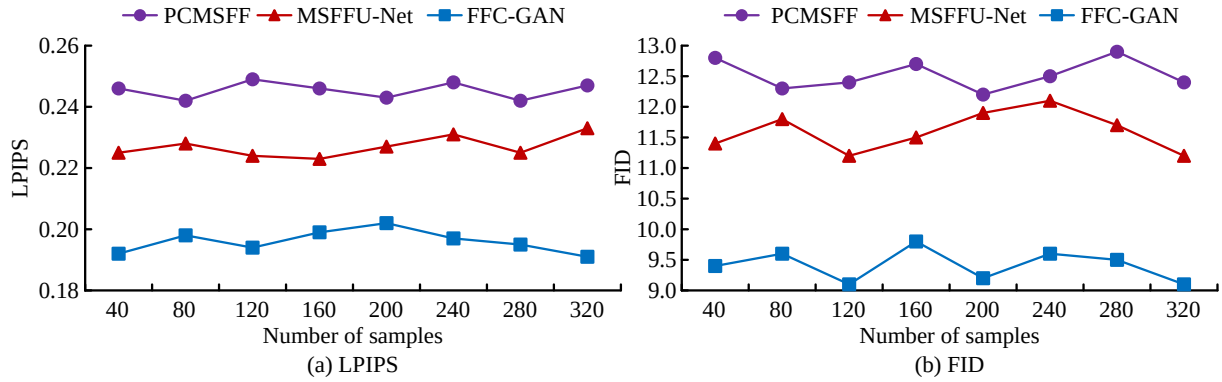


Figure 12: LPIPS and FID of different models. (Source from: Author's self drawn)

From Figure 11 (a), the SSIMs of PCMSFF and MSFFU-Net were the highest at 0.828 and 0.852, the lowest at 0.820 and 0.842, and the average SSIMs were 0.824 and 0.846, respectively. The SSIM of FFC-GAN was the lowest at 0.874, with an average SSIM of 0.878, which was higher than other methods. From Figure 11 (b), the PSNRs of PCMSFF and MSFFU-Net were the highest at 28.3dB and 31.3dB, and the lowest at 27.0dB and 30.0dB, respectively, with average PSNRs of 27.6dB and 30.6dB, respectively. The PSNR of FFC-GAN was the lowest at 34.1dB, with an average PSNR of 34.7dB, which was also higher than other algorithms. The above results indicated that the facial image restoration quality of FFC-GAN was superior to other algorithms. The Learned Perceptual Image Patch Similarity (LPIPS) and Frechet Inception Distance (FID) of different models are shown in Figure 12.

According to Figure 12 (a), the minimum and maximum LPIPS of PCMSFF and MSFFU-Net are 0.242 and 0.223, respectively, and 0.249 and 0.233, respectively. The average LPIPS was 0.245 and 0.227. The maximum LPIPS of FFC-GAN was 0.202, and the average LPIPS was 0.196, which was much lower than other methods. According to Figure 12 (b), the maximum FID of PCMSFF and MSFFU-Net were 12.9 and 12.1, and the minimum FID was 12.2 and 11.2. The average FID was 12.5 and 11.6, respectively. The maximum FID of FFC-GAN was 9.8, and the average FID was 9.4, which was also lower than other algorithms. The above results indicated that FFC-GAN could achieve high-quality restoration of large-area irregularly occluded face images. The facial recognition accuracy and TAR of different models are shown in Figure 13.

According to Figure 13 (a), the highest and lowest facial recognition accuracies of IFFR Net and DSCHL were 86.4% and 88.5%, respectively, and 84.7% and 87.3%, respectively. The average accuracies were 85.6% and 87.9%, respectively. The recognition accuracy of FFC-GAN was the lowest at 90.5%, with an average accuracy of 91.0%, which was higher than other algorithms. From Figure 13 (b), the TAR of IFF-Net and DSCHL were the highest at 88.3% and 90.3% respectively, the lowest at 87.2% and 89.1% respectively, and the average TAR was 87.7% and 89.6% respectively. The lowest TAR of FFC-GAN was 92.1, with an average TAR of 92.6%, which was also higher than other algorithms. The above results indicated that FFC-GAN could achieve accurate recognition of faces with large areas of irregular occlusion. To further analyze the performance of FFC-GAN, ablation experiments were conducted on it. The findings of the ablation experiment are denoted in Table 2.

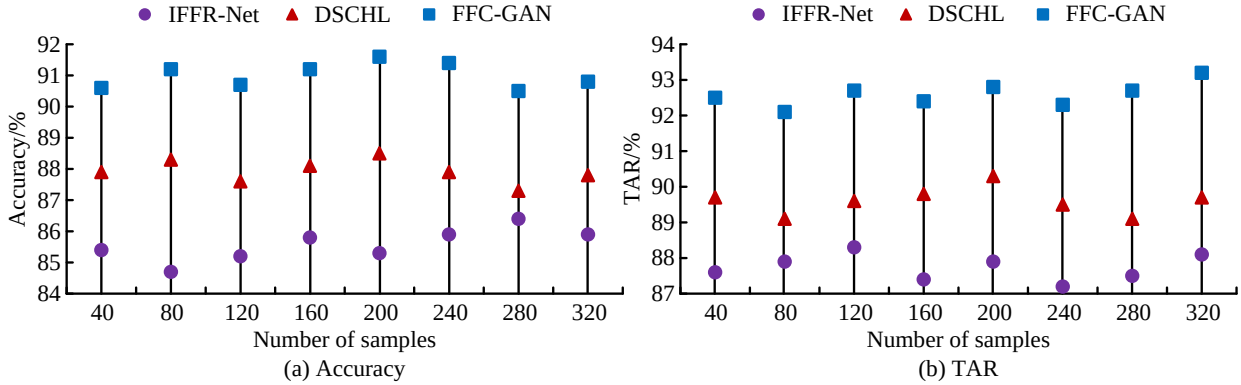


Figure 13: Face recognition accuracy and TAR of different models. (Source from: Author's self drawn)

Table 2: Ablation results.

Model	FFC	Spectral Transformer	Residual block	SSIM	Accuracy/%
1	×	×	×	0.725	84.4
2	√	×	×	0.796	88.2
3	×	√	×	0.771	87.5
4	×	×	√	0.795	87.9
5	√	√	×	0.827	88.9
6	√	×	√	0.846	89.2
7	×	√	√	0.859	89.8
8	√	√	√	0.878	91.0

According to Table 2, after introducing FCC, Spectral Transformer, and residual blocks, the SSIM and accuracy of the model significantly increased, reaching 0.878 and 91.0%, respectively. Among them, FFC and residual blocks had the most significant impact on model performance. After introducing FFC and residual blocks, the SSIM of the model increased to 0.796 and 0.795, respectively, and the accuracy increased to 88.2% and 87.9%, respectively.

4 Discussion

In recent years, with the booming development of the tourism industry, the number of tourists in scenic spots has been continuously increasing, which has brought many challenges to scenic spot management. The traditional management method of scenic spots has problems such as low efficiency, easy errors, and inability to monitor in real time, which not only affects the tourist experience but may also lead to safety hazards. The advent of artificial intelligence, computer vision, and deep learning technologies has precipitated a substantial enhancement in the security, convenience, and accuracy of facial recognition technology [24-25]. Real-time monitoring of personnel within the scenic area can be achieved through facial recognition technology, detecting abnormal behavior in a timely manner and issuing alerts. In addition, facial recognition systems can quickly locate missing persons or lost items, enhancing the emergency response capabilities of scenic spots. However, due to the complex environment and huge pedestrian flow in scenic spots, facial recognition is difficult [26]. Therefore, to achieve accurate recognition of faces in scenic area monitoring, an FR method based on improved ResNet-U^{2}Net was proposed to address the problem of FR under small area rule occlusion such as sunglasses and masks. A recognition method based on FFC-GAN was proposed for the FR problem of large irregular occlusion.

For the improved ResNet-U^{2}Net, experimental results showed that its average recognition accuracy and F1-score in simulated occlusion datasets were 98.7% and 0.983, respectively, with an average TAR of 99.5%, both higher than FGDFME and DIP rMRF. In the MFR2 dataset, the average recognition accuracy and F1-score of the improved ResNet-U^{2}Net were 97.9% and 0.967, respectively, with an average TAR of 98.5%, which was also higher than other algorithms. Haider et al. designed a variational invariant FR method based on multi-task learning, which redefines FR by combining temporal dependence and temporal independence to decompose the face into age and residual features. The experimental results showed that this method could achieve accurate recognition of faces of different races [27]. However, the above methods had low accuracy in recognizing occluded faces, while the proposed method could achieve accurate recognition of faces under objects such as masks and sunglasses. Akheel T S et al. proposed using optimized projection matrices in linear collaborative regression classification to improve recognition accuracy, and introduced a whale lion combination model to optimize the projection matrix. The findings denoted that the facial recognition accuracy of the model could reach 91.2% [28]. Compared to the above algorithms, the improved ResNet-U^{2}Net proposed in the study had higher facial recognition accuracy. This is because the improved ResNet-U^{2}Net introduces a global convolution module, allowing the model to capture a larger range of global information. Meanwhile, the model also introduced FPN, effectively enhancing its multi-scale feature extraction capability. In addition, the study also introduced a mask learning unit, which removes the features of occluded areas by generating multi-level masks to enhance feature representation.

For FFC-GAN, its average SSIM and average PSNR were 0.878 and 34.7 dB, respectively, which were higher than PCMSF and MSFFU-Net. The average LPIPS and FID were 0.196 and 9.4, respectively, which were lower than other algorithms. FFC-GAN could achieve accurate restoration of large-area irregularly occluded facial images. In terms of facial recognition performance, the average accuracy and TAR of FFC-GAN were 91.0% and 92.6%, respectively, both higher than existing advanced algorithms. Yan L. et al. proposed a methodology for optimizing image feature compensation coefficients. This methodology is based on an enhanced simulated annealing algorithm, the purpose of which is to enhance the recognition rate of facial recognition systems. The findings indicated that when the training image was designated as 6, the recognition rate attained a maximum of 100% [29]. Compared to the above methods, although the proposed method had lower recognition accuracy, it could effectively address the problem of large-scale irregular facial occlusion. Zaaraoui et al. put forward an FR method based on the mini value string, utilizing the mini value string as the face feature extractor for face representation. The findings demonstrated that the method exhibited high recognition accuracy and efficiency [30]. However, compared to the methods proposed in the research, the above methods significantly reduced the

accuracy of FR under large-scale irregular occlusion conditions. The reason why the proposed FFC-GAN can achieve accurate recognition of large-area irregularly occluded faces is that this method can accurately repair occluded images through GAN and accurately restore image texture details through FFC.

In summary, the improved ResNet-U^{2}Net and FFC-GAN can achieve accurate recognition of occluded faces, among which the improved ResNet-U^{2}Net has high recognition accuracy for small area regularly occluded faces. FFC-GAN can effectively repair large areas of irregularly occluded facial images, thereby achieving accurate facial recognition. The above two methods provide strong support for the development of facial recognition technology for scenic spot monitoring, which helps to achieve intelligent management of scenic spots. However, due to the high number of parameters and computational complexity of the proposed model, it requires high computing power from the server, making the deployment of the model difficult. Therefore, in the future, the model structure will be optimized to minimize the number of parameters and computational complexity of the model, so that it can be deployed on platforms with limited processing capabilities such as mobile devices and embedded devices.

5 Conclusion

A small area regular occlusion FR model based on improved ResNet-U^{2}Net and a large area irregular occlusion FR model based on FFC-GAN were proposed to address the issue of FR in scenic spot monitoring. The improved ResNet-U^{2}Net achieved accurate recognition of small area regularly occluded faces by introducing global convolution, FPN, and mask learning units. The findings denoted that the average recognition accuracy and F1-score of the improved ResNet-U^{2}Net reached 98.7% and 0.983, respectively, with an average TAR of 99.5%. The FFC-GAN model utilized GAN and FFC modules to repair and recognize large-area irregularly occluded facial images. The findings denoted that the average SSIM and PSNR of the model were 0.878 and 34.7dB, respectively, and the average accuracy and TAR were 91.0% and 92.6%, respectively, which were better than existing advanced algorithms. The above results indicated that improved ResNet-U^{2}Net and FFC-GAN could achieve accurate recognition of facial images under different occlusion conditions, providing strong support for the development of facial recognition technology for scenic spot monitoring. However, the model has high parameter count and computational complexity, which makes it impossible to deploy on mobile devices, greatly limiting its application scope. Therefore, in the future, the model will be lightweighted to reduce its complexity.

6 Funding

The research is supported by Research and Practice Project on Education and Teaching Reform of Henan Provincial Department of Education in 2024 (Project No. 2024SJGLX0837); The series of development achievements of the 2024 Henan Province Higher Education Teaching Achievement "Practical Research on the Transformation and Application Mode of Tour Guide Service Skills Competition" Competition Teaching Post "Promotion of General Education" (Achievement Number: Yujiao [2024] 49961).

References

[1] Hatef Otroshi Shahreza, and Sébastien Marcel. Template inversion attack using synthetic face images against real face recognition systems. IEEE Transactions on Biometrics, Behavior, and Identity Science, 6(3):374-384, 2024. https://doi.org/10.1109/TBIOM.2024.3391759

[2] Gautam Srivastava, and Surajit Bag. Modern-day marketing concepts based on face recognition and neuro-marketing: A review and future research directions. Benchmarking: An International Journal, 31(2):410-438, 2024. https://doi.org/10.1108/bij-09-2022-0588

[3] Volodymyr Mykolaevich Opanasenko, Shavkat Khayrullaevich Fazilov, Olimjon Nomazovich Mirzaev, and Shukrullo Sa'dullo ugli Kakharov. An ensemble approach to face recognition in access control systems. Journal of Mobile Multimedia, 20(3):749-768, 2024. https://doi.org/10.13052/jmm1550-4646.20310

[4] Thai-Viet Dang. Smart attendance system based on improved facial recognition. Journal of Robotics and Control, 4(1):46-53, 2023. https://doi.org/10.18196/jrc.v4i1.16808

[5] Lixiong Qin, Mei Wang, Chao Deng, Ke Wang, Xi Chen, Jiani Hu, and Weihong Deng. SwinFace: A multi-task transformer for face recognition, expression recognition, age estimation and attribute estimation. IEEE Transactions on Circuits and Systems for Video Technology, 34(4):2223-2234, 2023. https://doi.org/10.1109/TCSVT.2023.3304724

[6] Hind Moutaz Al-Dabbas, R. A. Azeez, and Akbas Ezaldeen Ali. Two proposed models for face recognition: Achieving high accuracy and speed with artificial intelligence. Engineering, Technology & Applied Science Research, 14(2):13706-13713, 2024. https://doi.org/10.48084/etasr.7002

[7] Wenjing Gao, Jia Yu, Rong Hao, Fanyu Kong, and Xiaodong Liu. Privacy-preserving face recognition with multi-edge assistance for intelligent security systems. IEEE Internet of Things Journal, 10(12):10948-10958, 2023. https://doi.org/10.1109/JIOT.2023.3240166

[8] Yun Xie, Peng Li, Nadia Nedjah, Brij B. Gupta, David Taniar, and Jindan Zhang. Privacy protection framework for face recognition in edge-based internet of things. Cluster Computing, 26(5):3017-3035, 2023. https://doi.org/10.1007/s10586-022-03808-8

[9] Huifang Feng, Wen Li, Lingfei Ma, Yiping Chen, Haiyan Guan, and Yongtao Yu. Crack-U^{2}Net:

Multiscale feature learning network for pavement crack detection from large-scale MLS point clouds. IEEE Transactions on Intelligent Transportation Systems, 25(11):17952-17964, 2024. https://doi.org/10.1109/TITS.2024.3436015

[10] Pengfei Shi, Fengting Zhu, Yuanxue Xin, and Shen Shao. U^{2}CrackNet: A deeper architecture with two-level nested U-structure for pavement crack detection. Structural Health Monitoring, 22(4):2910-2921, 2023. https://doi.org/10.1177/14759217221140976

[11] Yunchai Li, Run Fang, Nangang Zhang, Chengsheng Liao, Xiaochang Chen, Xiaoyu Wang, Yunfei Luo, Leheng Li, Min Mao, and Yunlong Zhang. An improved algorithm for salient object detection of microscope based on U^2-Net. Medical & Biological Engineering & Computing, 63(2):383-397, 2025. https://doi.org/10.1007/s11517-024-03205-w

[12] Zujia Zheng, and Kui Yang. Wall crack detection method based on improved YOLOv5 and U^2-Net. International Journal of Wireless and Mobile Computing, 25(4):362-367, 2023. https://doi.org/10.1504/ijwmc.2023.135405

[13] Jie Chen, Yong Kong, Dawei Zhang, Yinghua Fu, and Songlin Zhuang. Two-dimensional phase unwrapping based on U^2-Net in complex noise environment. Optics Express, 31(18):29792-29812, 2023. https://doi.org/10.1364/OE.500139

[14] Huahao Fan, and Yuan Li. Image recognition and reading of single pointer meter based on deep learning. IEEE Sensors Journal, 24(15):25163-25174. 2024. https://doi.org/10.1109/JSEN.2024.3416436

[15] Liangzhe Liao, Zhenkun Lei, Chen Tang, Ruixiang Bai, and Xiaohong Wang. Performance of a U^2-Net model for phase unwrapping. Applied Optics, 62(34):9108-9118, 2023. https://doi.org/10.1364/AO.504482

[16] Zunmei Hu, Yuwen Huang, and Yuzhen Yang. Dual-feature and multi-scale fusion using U^2-Net deep learning model for ECG biometric recognition. Journal of Intelligent & Fuzzy Systems, 45(5):7445-7454, 2023. https://doi.org/10.3233/JIFS-230721

[17] Hebbi Chandravva, and Mamatha hr. Comprehensive dataset building and recognition of isolated handwritten kannada characters using machine learning models. Artificial Intelligence and Applications, 1(3):179-190, 2023. https://doi.org/10.47852/bonviewAIA3202624

[18] Kunhua Liu, Yunqing Zhang, Yuting Xie, Leixin Li, Yutong Wang, and Long Chen. SynerFill: A synergistic RGB-D image inpainting network via fast Fourier convolutions. IEEE Transactions on Intelligent Vehicles, 9(1):69-78, 2023. https://doi.org/10.1109/TIV.2023.3326236

[19] Siavash Jafarzadeh, Farzaneh Mousavi, Longzhen Wang, and Florin Bobaru. PeriFast/Dynamics: A MATLAB code for explicit fast convolution-based peridynamic analysis of deformation and fracture. Journal of Peridynamics and Nonlocal Modeling, 6(1):33-61, 2024. https://doi.org/10.1007/s42102-023-00097-6

[20] Xi Jia, Joseph Bartlett, Wei Chen, Siyang Song, Tianyang Zhang, Xinxing Cheng, Wenqi Lu, Zhaowen Qiu, and Jinming Duan. Fourier-net: Fast image registration with band-limited deformation. Proceedings of the AAAI Conference on Artificial Intelligence, 37(1):1015-1023, 2023. https://doi.org/10.1609/aaai.v37i1.25182

[21] Valeriy A. Buryachenko. Fast Fourier transform method in peridynamic micromechanics of composites. ASME 2023 International Mechanical Engineering Congress and Exposition, 29(9):1844-1878, 2024. https://doi.org/10.1177/10812865241236878

[22] Nicholas F. Marshall, Oscar Mickelin, and Amit Singer. Fast expansion into harmonics on the disk: A steerable basis with fast radial convolutions. Methods and Algorithms for Scientific Computing, 45(5):2431-2457, 2023. https://doi.org/10.1137/22M1542775

[23] Qiaosi Yi, Faming Fang, Guixu Zhang, and Tieyong Zeng. Frequency learning via multi-scale Fourier transformer for MRI reconstruction. IEEE Journal of Biomedical and Health Informatics, 27(11):5506-5517, 2023. https://doi.org/10.1109/JBHI.2023.3311189

[24] Sefik Ilkin Serengil, and Alper Ozpinar."A benchmark of facial recognition pipelines and co-usability performances of modules. Bilişim Teknolojileri Dergisi, 17(2):95-107, 2024. https://doi.org/10.17671/gazibtd.1399077

[25] Shivam Gupta, Sachin Modgil, Choong-Ki Lee, and Uthayasankar Sivarajah. The future is yesterday: Use of AI-driven facial recognition to enhance value in the travel and tourism industry. Information Systems Frontiers, 25(3):1179-1195, 2023. https://doi.org/10.1007/s10796-022-10271-8

[26] Rosalie A. Waelen. The struggle for recognition in the age of facial recognition technology. AI and Ethics, 3(1):215-222, 2023. https://doi.org/10.1007/s43681-022-00146-8

[27] Abbas Haider, Guanfeng Wu, Ivor Spence, and Hui Wang. Residual feature decomposition and multi-task learning-based variation-invariant face recognition. Neural Computing and Applications, 36(32):20147-20166, 2024. https://doi.org/10.1007/s00521-024-10234-x

[28] T. Syed Akheel, V. Usha Shree, and S. Aruna Mastani. Hybrid model for face recognition using optimized linear collaborative discriminant regression classification. Mathematical Statistician and Engineering Applications, 71(4):10916-10924, 2022. https://doi.org/10.1007/s00521-018-3475-4

[29] Lijuan Yan, Yanhu Zhang, and Yanjun Zhang. A fast face recognition system based on annealing algorithm to optimize operator parameters. The Imaging Science Journal, 71(3):323-330, 2023. https://doi.org/10.1080/13682199.2023.2182261

[30] Hicham Zaaraoui, Samir El Kaddouhi, and Mustapha Abarkan. A novel face recognition approach based

on strings of minimum values and several distance metrics. International Journal of Computer Aided Engineering and Technology, 18(1):60-76, 2023. https://doi.org/10.1504/ijcaet.2023.127787