Face Recognition under Mask Occlusion via Multi-Scale PCAnet and Progressive GAN-based Face Restoration

Nan Zhang*, Fangqin Wang
School of Control Engineering, Wuxi Institute of Technology, Wuxi 214121, China
E-mail: zhangn303313@163.com
*Corresponding author

Keywords: mask occlusion, face recognition, principal component analysis network, multi-scale feature fusion, progressive training, face reproduction

Received: May 13, 2025

With the widespread application of face recognition technology in public safety, identity authentication, and other fields, face recognition under mask cover has gradually become a research hotspot. Traditional methods rely on manual feature extraction and template matching, making it difficult to deal with the loss of face features caused by occlusion. To accurately recognize faces under mask occlusion, an improved principal component analysis network based on multi-scale feature fusion (Multi-Scale PCANet, MPCANet) and a progressively trained multi-scale face restoration generative adversarial network (Multi-Scale Face Restoration Generative Adversarial Network, MFR-GAN) are proposed for face biometric recognition under masks. MPCANet takes multi-layer convolution to extract features, concatenates different levels of feature information through channels, and divides the feature map into multi-scale blocks through pyramid pooling layers to form multi-scale feature vectors for matching and recognizing occluded faces. MFR-GAN utilizes a progressive training strategy to gradually increase the resolution from low resolution images for training, and introduces discrete wavelet transform to extract high-frequency information and calculate high-frequency loss, optimizing the quality of image generation. The experiment was conducted on the Labeled Faces in the Wild (LFW) dataset, with an average accuracy of 94.24% and a computation time of 0.87 seconds for MPCANet. The MFR-GAN had a realism of 0.95 and a peak signal-to-noise ratio of 30.25dB at a 30% occlusion ratio. In practical applications, the new model achieved a recognition accuracy of 98.68%, a false recognition rate of 1.32%, and a recognition rate of 97.58% under occlusion conditions, demonstrating high efficiency and accuracy. This method provides an efficient solution for biometric recognition in mask occlusion scenarios, which is helpful for identity authentication.

Povzetek: Predlagana sta modela za prepoznavanje obrazov z masko, ki z večmerilnimi značilkami in obnovo obraza dosežeta visoko natančnost (do 98,68 %) tudi pri zakritih obrazih.

1 Introduction

Face recognition technology, as a biometric technology, has been largely applied in public safety, identity authentication, and financial payments in recent years [1]. In the context of normalized public health incidents, wearing masks has become the norm [2]. Mask occlusion can result in over 40% of key face features (such as nose and mouth) being unable to be effectively extracted, making it difficult for traditional methods to accurately recognize them [3]. In public places such as hospitals, airports, and train stations, mask face recognition technology can monitor individuals who are not wearing masks in real time, improving the efficiency of epidemic prevention and control [4]. In scenarios such as financial payments and access control systems, face recognition technology under masks can achieve contactless identity verification, enhancing security and convenience [5]. Face recognition technology can help enhance public safety and emergency response capabilities in smart transportation and smart communities [6]. Some scholars have explored to improve the accuracy of face recognition under masks. Song et al. built a Spartan face detection system with stacked ensemble deep learning algorithm to accurately detect faces under masks. The system provided a cost-effective face detection solution for enterprises, which could be applied to edge devices [7]. Neto et al. proposed a Masked Face Recognition (MFR) model to separate masked face from various occlusion strategies. The MFR model could be applied to occluded face recognition under reasonable occlusion sizes, and its performance was better than traditional methods, while having lower computational complexity [8]. Kaur et al. built a mask face recognition method based on Convolutional Neural Network (CNN) to existing face recognition systems being difficult to accurately recognize mask wearing faces. This method could maintain high accuracy even in side view or multi-person scenes [9].

Table1: Comparison table of related work methods.

Research Method	Research Content	Performance	Limitations
Spartan System [7]	Proposed a Spartan face detection and recognition system with stacked integrated deep learning algorithms for mask detection and recognition.	Provides a cost - effective face detection and recognition solution for enterprises and schools, applicable to edge devices.	Demands high - end hardware, making it unsuitable for resource - constrained devices.
MFR Model [8]	Presented a MFR model to distinguish masked face recognition from other occluded face recognition methods.	Effective for occluded face recognition within reasonable occlusion sizes, outperforms traditional methods, and has lower computational complexity.	Performance degrades under extreme occlusions and possesses relatively high computational complexity.
CNN - Based Method [9]	Introduced a CNN - based masked face recognition method to address the inaccuracy of existing face recognition systems in recognizing masked faces.	Maintains high accuracy in side - view and multi - person scenarios.	Insufficient accuracy in complex situations and strongly dependent on datasets.
FRGFT Method [10]	Proposed a Face Reconstruction using FRGFT method for face reconstruction.	Achieves 98.5% accuracy in reconstructed face images.	Prone to structural distortion when handling large - area occlusions.
GAN - Based Framework [11]	Presented a GAN framework algorithm combining self - attention mechanism and global - local discriminators.	Superior to other methods in repair effect and training efficiency.	Likely to suffer from structural distortion and semantic blur when dealing with large - area occlusions.
DCGAN Method [12]	Proposed a face restoration method based on DCGAN.	Offers favorable performance in reproducing occluded parts.	Likely to encounter structural distortion and semantic blur when handling large - area occlusions.

In the field of face occlusion repair, Meena et al. proposed a Face Reconstruction using Graph Fourier Transform (FRGFT) method to optimize the face recognition accuracy. The face image reconstructed by FRGFT had an accuracy of 98.5% [10]. She et al. built a Generative Adversarial Network (GAN) that combined self-attention mechanism and global local discriminator to address the structural distortion and semantic blur that existing face image restoration algorithms were prone to when dealing with large areas of occlusion. The experimental results showed that the algorithm outperformed other methods on repair effectiveness and training efficiency [11]. Dai et al. proposed a face restoration method based on Deep Convolutional Generative Adversarial Network (DCGAN) to address the reduced accuracy in face detection algorithms caused by local occlusion factors. The experimental results showed that this method could achieve good results in reproducing the occluded parts [12].

However, there are some problems with face recognition under existing masks. The Spartan system has high hardware requirements and is difficult to apply to resource constrained devices. The MFR model performs poorly in extreme occlusion situations and has high computational complexity. The CNN has insufficient accuracy in complex scenarios such as multiple people and side view, which is highly dependent on the dataset. In the face occlusion restoration, FRGFT is prone to structural deformation when dealing with large areas of occlusion, while the GAN framework still has room for improvement in training efficiency and image detail preservation.

This article aims to propose a Face Reconstruction Model Under Mask (FRMUM) based on MPCANet and MFR-GAN to address the shortcomings of existing mask face recognition technologies on accuracy, robustness, and adaptability.

The main objective of the research is to improve the accuracy of face recognition under mask occlusion, reduce the demand for hardware resources, and enhance the applicability of the model in complex scenes.

The expected results include developing an MPCANet that can efficiently extract and fuse multi-scale features, as well as an MFR-GAN that can generate high-quality and realistic face images, thereby achieving accurate recognition and realistic reproduction of masked faces.

The research innovatively enhances the robustness of local occlusion through multi-scale feature fusion and spatial pyramid pooling to fully utilize global and local feature information. Progressive training and multi-scale loss function are used to optimize GAN to generate high-quality and realistic face images, which can significantly reduce the computational complexity and improve the adaptability while maintaining high accuracy.

2 Method

2.1 Improved PCAnet algorithm based on multi-scale feature fusion

With the widespread application of face recognition technology in public safety, identity authentication, and other fields, how to achieve accurate face recognition in complex scenarios (such as wearing masks) has become a research hotspot [13]. The traditional face recognition methods under masks mainly rely on manual feature extraction and template matching, which are difficult to cope with feature loss and face structure changes caused by occlusion, and the recognition accuracy and robustness are significantly limited. Principal Component Analysis Network (PCAnet) is a lightweight feature extraction algorithm based on deep learning [14]. PCAnet achieves efficient feature extraction and dimensionality reduction through hierarchical principal component analysis and binary hash coding, which has low computational complexity and strong robustness [15]. The reason for using PCAnet is that it can effectively extract deep

features of high-dimensional data, while avoiding the dependence of traditional CNN on large-scale training data, especially suitable for processing local features in occluded face images. However, PCAnet suffers from spatial information loss when dealing with occluded faces, especially when there are insufficient training samples, resulting in a significant decrease in recognition accuracy.

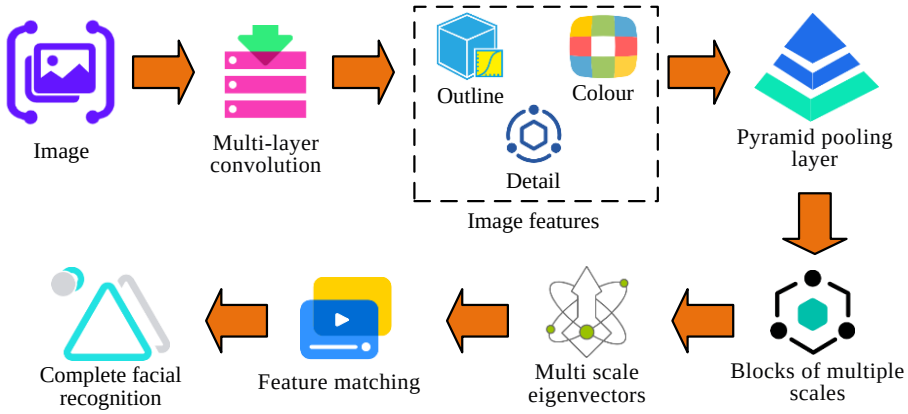

Figure 1: MPCAnet recognition process of face features under mask occlusion.

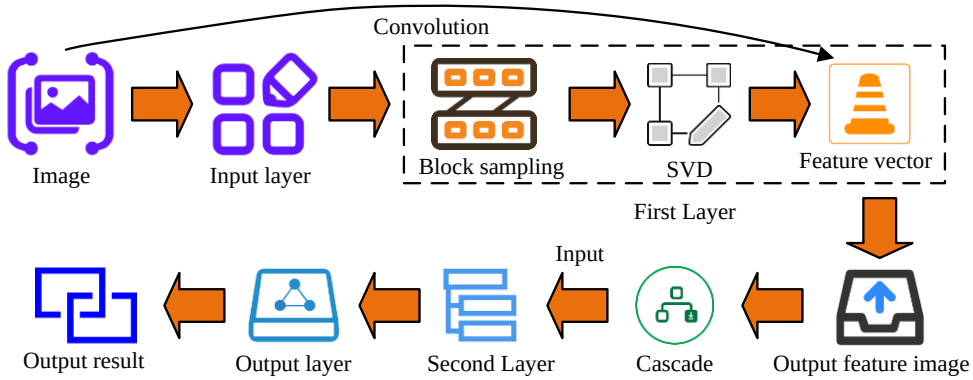

Figure 2: Structure of MPCAnet.

Therefore, a multi-scale feature fusion strategy is introduced to form the MPCAnet. It enhances the robustness of PCAnet to local occlusion by combining feature information from different scales, while preserving global structural information. MPCANet can utilize the feature expression capabilities at different levels to make up for the deficiency of single-scale feature extraction in traditional PCANet, thereby effectively recognizing occluded faces. The process of MPCAnet recognizing face features under mask occlusion is presented in Figure 1.

In Figure 1, firstly, MPCAnet takes multi-layer convolution to extract image features. Different levels of feature information are fused through cascading channels to enhance the ability of feature expression. Next, MPCAnet divides the feature map into blocks with multiple scales through pyramid pooling layers, forming multi-scale feature vectors. Finally, these features are matched with the features of the training data to recognize occluded faces. The structure of MPCAnet is presented in Figure 2.

In Figure 2, the MPCAnet has multiple parts. The input layer receives training images and performs block sampling on them. The input image matrix is shown in equation (1) [16].

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,t} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1k2,1} & x_{k1k2,2} & \cdots & x_{k1k2,t} \end{bmatrix} \tag{1}$$

In equation (1), X represents the image matrix. $x_{k1k2,t}$ represents the pixel value of column $k2$ in row $k1$. t represents the number of blocks. In the first layer of convolution, MPCAnet obtains feature vectors through Singular Value Decomposition (SVD). These feature vectors are represented as matrices and used as filters in the first layer. Each filter convolves with the input image to produce the output feature map of the first layer. The filter matrix of the first layer is shown in equation (2).

$$W_1 = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,L1} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,L1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{k1k2,1} & w_{k1k2,2} & \cdots & w_{k1k2,L1} \end{bmatrix} \tag{2}$$

In equation (2), W_1 represents the filter matrix of the first layer. $w_{k1k2,L1}$ represents the filter weight of column $k2$ in row $k1$. $L1$ represents the number of filters. A key improvement of MPCAnet is to cascade the original image

features and the output features of the first layer at the channel, and take the cascade result as input to the second layer. This enables MPCAnet to utilize more feature information, thereby improving its feature expression ability. The input feature matrix of the second layer is shown in equation (3).

$$Y=\begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,t+L1} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,t+L1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k1k2,1} & y_{k1k2,2} & \cdots & y_{k1k2,t+L1} \end{bmatrix} \tag{3}$$

In equation (3), Y represents the input feature matrix of the second layer. $y_{k1k2,t+L1}$ represents the characteristic value of column $k2$ in row $k1$. $t+L1$ represents the feature dimension. The output layer includes non-linear processing, block histograms, and feature concatenation. After nonlinear processing, each output feature map is divided into multiple blocks, and the histogram information of each block is statistically analyzed. All histogram features are concatenated into an extended histogram feature vector, which serves as the feature representation of the input image. The histogram feature vector is presented in equation (4).

$$H=\begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,B} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,B} \\ \vdots & \vdots & \ddots & \vdots \\ h_{k1k2,1} & h_{k1k2,2} & \cdots & h_{k1k2,B} \end{bmatrix} \tag{4}$$

In equation (4), H represents the histogram feature vector. $h_{k1k2,B}$ represents the histogram feature values of column $k2$ in row $k1$. B represents the number of blocks. By introducing multi-scale feature fusion, MPCAnet can simultaneously extract high-level semantic information and low-level detail information, exhibiting better robustness and accuracy in occlusion face recognition tasks. In addition, MPCAnet also constructs feature descriptors through spatial pyramid pooling strategy to further utilize image spatial information. Figure 3 displays the pyramid pooling layer structure of MPCAnet.

From Figure 3, this structure aims to enhance the robustness of the model to occlusion and pose changes through multi-scale feature extraction. The structure divides the feature map into multiple levels, and each level corresponds to different spatial resolutions. The number of feature blocks after the feature map is divided in the height direction is shown in equation (5).

$$B_h=\left\lfloor \frac{H}{S} \right\rfloor \tag{5}$$

In equation (5), B_h represents the number of blocks in the height direction of the feature map, and H signifies the height of the feature map. S signifies the edge length of the feature block. $\lfloor\ \rfloor$ signifies rounding down. The number of feature blocks after dividing the feature map in the width direction is shown in equation (6).

$$B_w=\left\lfloor \frac{W}{S} \right\rfloor \tag{6}$$

In equation (6), B_w signifies the quantity of blocks in the width direction of the feature map. W signifies the width of the feature map. The total blocks are shown in equation (7).

$$B_z=B_w\times B_h \tag{7}$$

In equation (7), B_z represents the total blocks. The size of each block is $S\times S$, which facilitates subsequent pooling operations and feature extraction. The feature blocks at each level extract local information through max pooling operation, forming feature representations at different scales. These feature representations are concatenated into a comprehensive feature vector for subsequent classification tasks. MPCAnet inputs the output feature vectors into a random forest classifier for classification to obtain the final classification result. Finally, the research compares the classification features with the features in the database to complete the entire face feature recognition process under the mask. Random forest is chosen as the classifier because it has high efficiency and robustness when dealing with high-dimensional features, and performs well for small sample data. The feature vectors output by MPCANet have a higher dimension, and the random forest can be classified more efficiently. In contrast, support vector machines may face high computational complexity in high-dimensional feature Spaces, while fully connected layers require a large amount of training data to avoid overfitting.

To further validate the effectiveness of the MPCANet structure, particularly the impact of spatial pyramid pooling and multi-layer feature fusion, ablation studies are conducted. When using only basic PCANet, the recognition accuracy is 82.35%. After adding the multi-layer feature fusion, the accuracy increased to 88.76%. After further introducing the spatial pyramid pooling, the accuracy ultimately reaches 94.24%. This indicates that multi-layer feature fusion and spatial pyramid pooling have significant contributions to improving model performance. Specifically, the multi-layer feature fusion enhances the model robustness to local occlusion by combining feature information from different levels. The spatial pyramid pool further improves the model's utilization efficiency of image spatial information through multi-scale feature extraction.

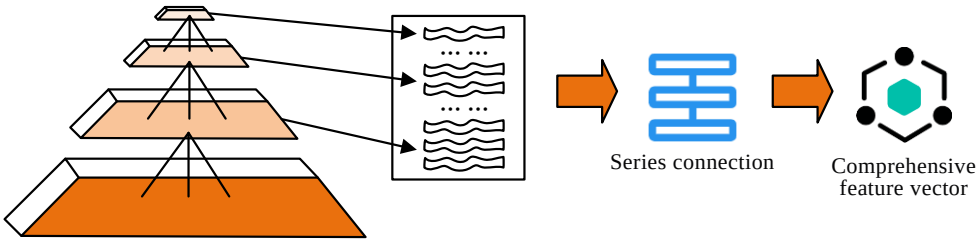

Figure 3: Pyramid pooling layer structure.

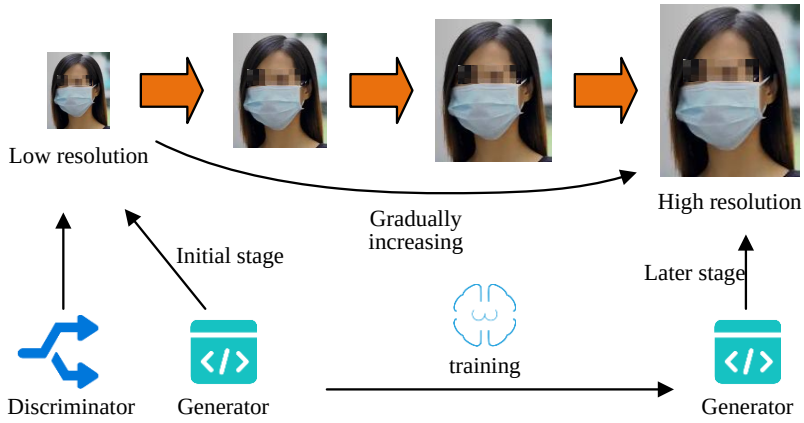

Figure 4: Progressive training method.

2.2 Construction of FRMUM model based on MFR-GAN algorithm

The research improves PCAnet by introducing multi-scale feature fusion methods. The MPCAnet is obtained. MPCAnet can effectively recognize face features under masks. To more intuitively determine the face under the mask, it is necessary to restore and reproduce the face under the mask. GAN is a powerful image generation and restoration tool that can generate realistic images based on existing information. The GAN has excellent performance in image restoration tasks, which can generate high-quality and realistic face images. However, GAN has shortcomings such as blurry generated images and detail loss. To overcome these shortcomings, progressive training and multi-scale loss functions are used to improve GAN, forming the MFR-GAN algorithm. The progressive training method is shown in Figure 4.

In Figure 4, progressive training starts from low resolution images and increases the resolution of the images. In the initial stage, the generator and discriminator are trained on lower resolution images, which help to converge quickly and reduce computational resource consumption. The increased image resolution is shown in equation (8).

$$R_{new} = R_{old} \times 2 \tag{8}$$

In equation (8), R_{new} represents the new image resolution. R_{old} represents the old image resolution. When increasing the resolution, the image resolution doubles. As the training progresses, the image resolution gradually increases, and the generator needs to learn how to maintain the details and structure of the image at higher resolutions. The weighting coefficient during the transition period is shown in equation (9).

$$\alpha = \frac{t}{T} \tag{9}$$

In equation (9), α represents the weighting coefficient of the transition period. t signifies the current training iterations. T signifies the total number of iterations during the transition period. During the transition period, α linearly increases from 0 to 1. The progressive training method helps to improve the ability of MFR-GAN algorithm to process high-resolution images. The structure of MFR-GAN algorithm is presented in Figure 5.

In Figure 5, the MFR-GAN algorithm is a deep learning model used for face restoration under mask occlusion. The MFR-GAN algorithm first obtains the input image and uses progressive training for high-resolution image learning. The generator network consists of multiple convolutional and deconvolution layers, used to extract and generate features at different scales. This multi-scale feature extraction helps the generator network to generate high-quality images at different resolutions.

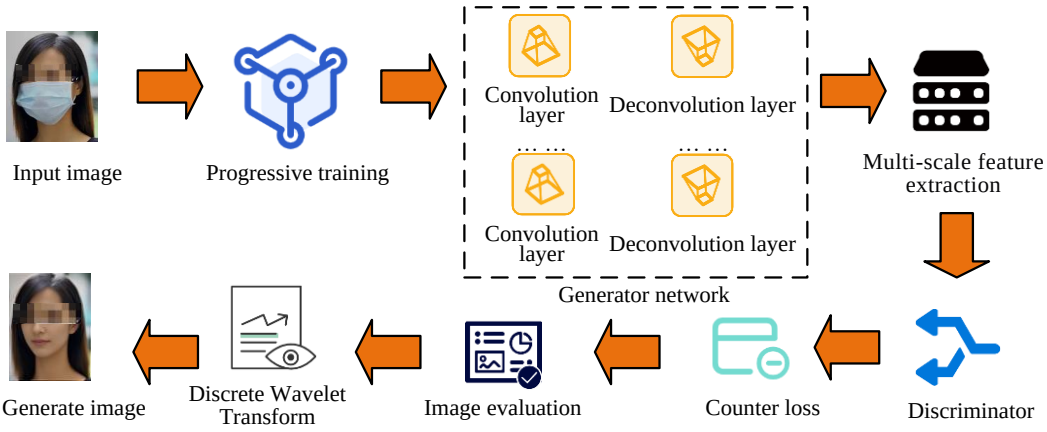

Figure 5: Structure of MFR-GAN algorithm.

The discriminator is used to distinguish between generated images and real images, optimizing the output of the generator through adversarial loss to make it closer to the real image. The adversarial loss function is shown in equation (10).

$$L_{adv} = -E_{x\sim p_{data}(x)}[\log D(x)] - E_{z\sim p_z(z)}[\log(1-D(G(z)))] \quad (10)$$

In equation (10), L_{adv} represents the adversarial loss. $D(x)$ represents the discriminator's discrimination result on the real image x. $G(z)$ signifies the image generated by the generator. $p_{data}(x)$ signifies the distribution of real images. $p_z(z)$ represents the distribution of input noise for the generator. The discriminator evaluates the generated images at multiple scales to ensure their realism at different resolutions. The MFR-GAN introduces a pre-trained Visual Geometry Group (VGG) network for extracting advanced features of images. These features are used to calculate perceptual loss to improve the visual quality of generated images. The perceptual loss function is shown in equation (11).

$$L_{perc} = \frac{1}{N}\sum_{i=1}^{N}\left\|\phi_i(G(z)) - \phi_i(x)\right\| \quad (11)$$

In equation (11), L_{perc} represents the perceived loss. ϕ_i signifies the features extracted from the i-th layer of the pre-trained VGG. N represents the number of characteristic layers. Finally, MFR-GAN algorithm uses discrete wavelet transform to extract the high-frequency information, and optimizes the generator through high-frequency loss to ensure that the generated image retains rich high-frequency details. The high frequency loss function is shown in equation (12).

$$L_{hf} = \left\|DWT(G(z)) - DWT(x)\right\| \quad (12)$$

In equation (12), L_{hf} represents the high-frequency loss. DWT represents the discrete wavelet transform, used to extract high-frequency information from images. The decomposition process of discrete wavelet transform is divided into low-frequency part (approximation coefficients) and high-frequency part (detail coefficients). The low-frequency coefficient is presented in equation (13).

$$c(A_{j,k}) = \sum_n x[n]\cdot\phi_{j,k}(n) \quad (13)$$

In equation (13), $c(A_{j,k})$ represents the low-frequency coefficient. $x[n]$ represents the input image. $\phi_{j,k}(n)$ represents the scale function (low-pass filter). j represents the scale parameter, indicating the decomposition level. k represents the position parameter, indicating the position of the signal in time or space. The high-frequency coefficient is shown in equation (14).

$$cD_{j,k} = \sum_n x[n]\cdot\psi_{j,k}(n) \quad (14)$$

In equation (14), $cD_{j,k}$ represents the high-frequency coefficient. $\psi_{j,k}(n)$ represents the wavelet function (high-pass filter). The reconstructed image is shown in equation (15).

$$x_{re}[n] = \sum_k cA_{j,k}\cdot\phi_{j,k}(n) + \sum_k cD_{j,k}\cdot\psi_{j,k}(n) \quad (15)$$

In equation (15), $x_{re}[n]$ represents the reconstructed image. To accurately recognize the face under mask, FRMUM model is built based on MPCAnet and MFR-GAN. The process of FRMUM model reproducing the face under the mask is shown in Figure 6.

In Figure 6, the FRMUM model first preprocesses the input image, including cropping to retain the face part and normalizing to meet the input requirements of the model. The preprocessed image is input into the MPCANet module. This module extracts features through multi-layer convolution and uses channel cascading to fuse feature information from different levels. Then, the feature map is divided into blocks of multiple scales through the pyramid pooling layer to form multi-scale feature vectors. These feature vectors fuse the local features (such as eyes, forehead, etc.) and global features (such as the overall structure of the face) of the unoccluded areas, which can effectively avoid spatial information loss caused by occlusion. The MPCANet module can determine face recognition based on these features. The extracted feature vectors are then input into the MFR-GAN module. In MFR-GAN, the generator network gradually generates face images of different resolutions through progressive training. The initial stage starts training from low-resolution images. As the training progresses, the image

resolution gradually increases, and the generator learns to maintain the image details and structure at higher resolutions. The discriminator network is used to evaluate the authenticity of the generated image and optimize the output of the generator through adversarial loss to make it closer to the real image. Furthermore, MFR-GAN combines the pre-trained VGG network to extract advanced features for calculating perceptual loss, and takes discrete wavelet transform to extract high-frequency information for calculating high-frequency loss, thereby further improving the quality of the generated images. After multiple iterations and optimizations, the images generated by MFR-GAN are highly consistent with the real human face on local details and global structure. Eventually, an unoccluded human face image is output to retain the identity features of the original human face and eliminate the influence of mask occlusion.

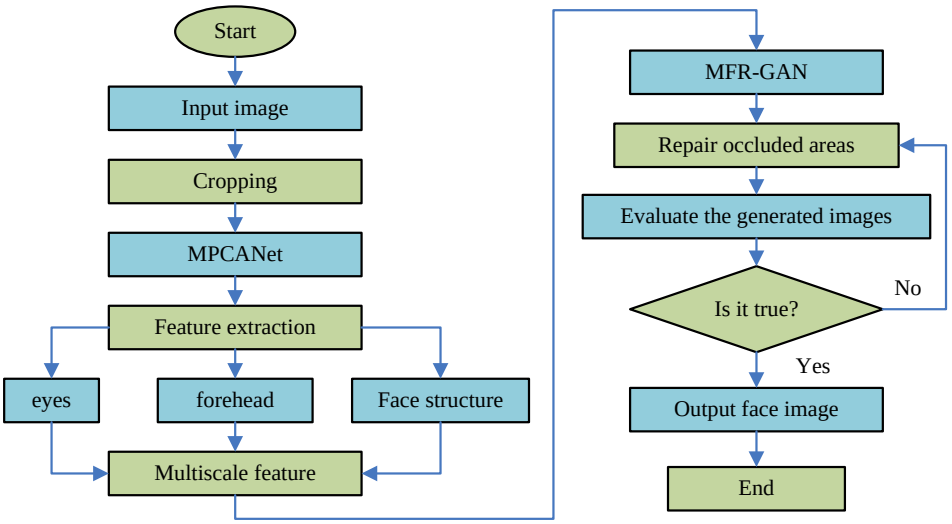

Figure 6: The process of FRMUM model reproducing face under mask.

When applying the FRMUM model in practice, ethical and privacy issues need to be emphasized. Reconstructing faces from occluded data may raise privacy concerns. To ensure ethical compliance, the principle of informed consent must be followed, clearly informing users of the purpose of their data and obtaining their authorization. Meanwhile, technologies such as encryption and access control should be adopted to protect data security and prevent leakage and abuse. In addition, the application of the reconstructed image should be limited to recognize authentication only and not other scenarios that infringe on privacy, so as to achieve effective recognition while protecting privacy.

3 Results

3.1 Performance analysis of MPCAnet algorithm

To verify the MPCAnet algorithm, a high-performance experimental platform is built. The CPU of the experimental platform is Intel Core i7-12700K 5.0GHz, the GPU is NVIDIA GeForce RTX 3080 Ti 12GB GDDR6X, the memory is DDR4 3200Mhz, and the hard disk is 1TB SSD. The operating system is Ubuntu 20.04 LTS, the deep learning framework is PyTorch 1.9.0, and the development environment is Python 3.8. Table 2 presents the parameters of MPCAnet.

The experimental dataset is Labeled Faces in the Wild (LFW) dataset. LFW is a public face recognition dataset, which has 13,233 images, involving 5,749 people. It is suitable for medium-sized experiments. The dataset size is moderate, which can provide enough samples for training and testing without significantly increasing the computational burden due to the large amount of data. LFW dataset contains a large number of occluded face images (such as wearing masks, sunglasses, etc.), which can effectively simulate the occlusion problem in real scenes. The comparison algorithm adopts PCAnet and DeepMasknet [17]. DeepMasknet employs a ten-layer deep neural network architecture that consists of six convolutional layers and four fully connected layers. The training conditions of DeepMasknet are consistent with those of PCAnet and MPCAnet. The accuracy and operation time of each algorithm to recognize the face under the mask are shown in Figure 7.

Table 2: Parameters of MPCAnet algorithm.

Parameter Name	Meaning	Parameter
Input Image Size	The three dimensions of the input image, representing height, width, and number of channels	80×50×3
Tensor Block Size	The size of the tensor blocks used for extracting local features	3×3×20
Number of Projection Matrices	The number of projection matrices for each mode	3
Projection Matrix Dimension	The dimension of each projection matrix	8
Number of Convolutional Layers	The number of convolutional layers	2
Pooling Window Size	The size of the pooling window	8×5
Pooling Overlap Ratio	The overlap ratio between pooling windows	50%
Energy Retention Ratio	The proportion of energy retained in the projection dictionary layer	97%

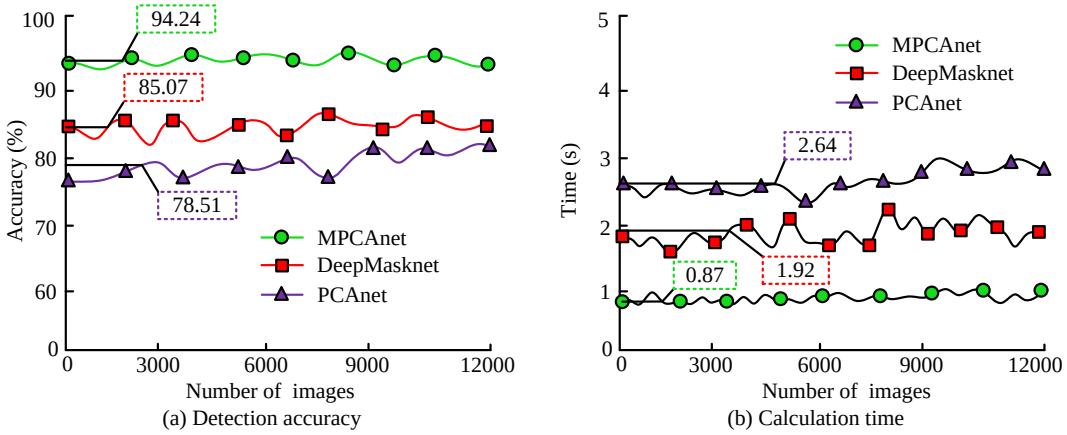

Figure 7: Comparison of face recognition accuracy and operation time of each algorithm.

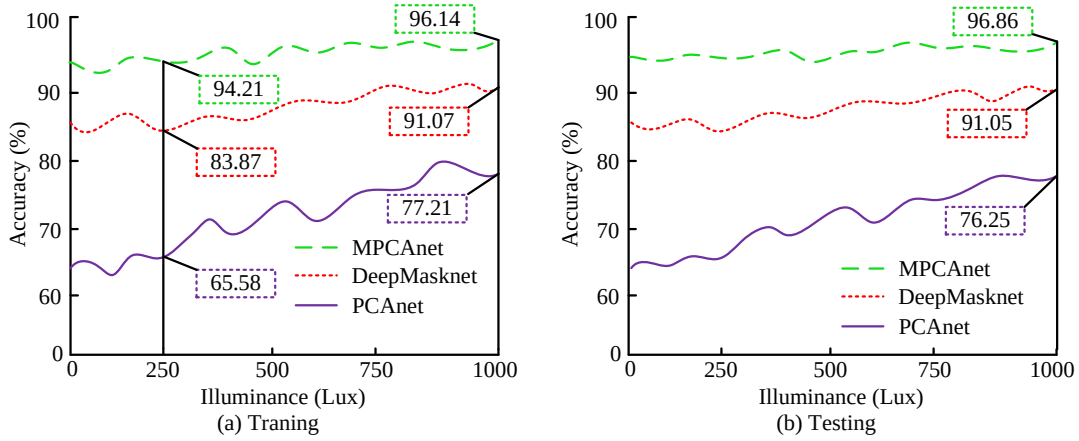

Figure 8: Face recognition accuracy under different lighting intensities.

In Figure 7 (a), the face recognition accuracy of MPCAnet under mask occlusion is significantly better than that of PCAnet and DeepMasknet. The average accuracy of MPCAnet, PCAnet and DeepMasknet was 94.24%, 78.51% and 85.07%, respectively. This shows that MPCAnet has higher robustness and accuracy in processing occluded faces. MPCAnet can extract more abundant feature information through multi-layer feature fusion and channel level online system, so it can perform well in complex occlusion conditions. In Figure 7 (b), MPCAnet also had better operation time than PCAnet and DeepMasknet. The average recognition time of MPCAnet for face image was 0.87 seconds, while that of PCAnet was 2.64 seconds, and that of DeepMasknet was 1.92 seconds. The high efficiency of MPCAnet benefits from its lightweight structure and the high performance of random forest classifier, which makes it maintain high accuracy and significantly reduce the calculation time. The face recognition accuracy under different illumination intensities is presented in Figure 8.

In Figure 8 (a), during the training process, the accuracy of each algorithm for face recognition increased with the increase of illumination. When the illumination intensity was 250lux, the face recognition accuracy of MPCAnet, PCAnet and DeepMasknet were 94.21%, 65.58% and 83.87%, respectively. When the illumination intensity increased to 1,000lux, the face recognition accuracy of MPCAnet, PCAnet, and DeepMasknet were 96.14%, 77.21% and 91.07%, respectively. The accuracy of face recognition of MPCAnet was least affected by illumination. In Figure 8 (b), the performance of each algorithm was basically consistent with the training process during the test. When the illumination intensity increased to 1,000lux, the face recognition accuracy of MPCAnet, PCAnet and DeepMasknet was 96.86%, 76.25% and 91.05%, respectively.

3.2 Analysis of the practical application effect of MFR-GAN algorithm and FRMUM model

The performance of MPCAnet for fast and accurate face recognition under mask provides the basis for MFR-GAN to reproduce the face under mask. To analyze the performance of MFR-GAN in reproducing face images under masks, face images with different occlusion ratios are selected for testing. Authenticity (consistency between generated images and real face images), Structural Similarity (SSIM), and Peak Signal-to-Noise Ratio (PSNR) are taken as image performance parameters. GAN and FMODR [18] are taken as comparison algorithms. The face image reproduction results across varying occlusion ratios are shown in Figure 9.

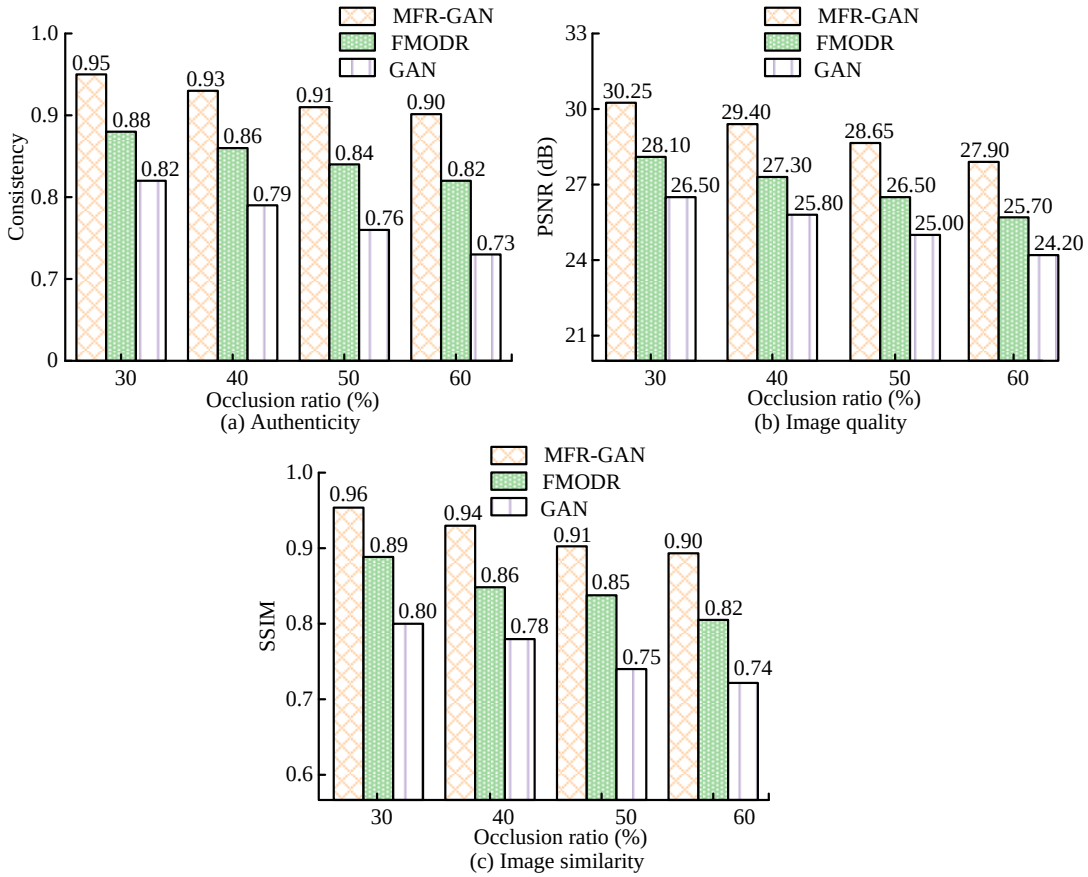

Figure 9: The face image reproduction results across varying occlusion ratios.

In Figure 9 (a), the authenticity of MFR-GAN under different occlusion ratios was significantly higher than that of FMODR and GAN. For example, under the 30% occlusion ratio, the authenticity of MFR-GAN was 0.95, FMODR was 0.88, and GAN was 0.82. This shows that the image generated by MFR-GAN is closer to the real face image and has higher visual quality. In Figure 9 (b), the PSNR value of MFR-GAN under different occlusion ratios was also higher than that of FMODR and GAN. For example, at 30% occlusion ratio, the PSNR of MFR-GAN was 30.25 dB, FMODR was 28.10 dB, and GAN was 26.50 dB. Because MFR-GAN adopts progressive training method, it ensures that the generator can generate high-quality images. In Figure 9 (c), the SSIM value of MFR-GAN remained above 90% under different occlusion ratios. Therefore, the image quality generated by MFR-GAN is superior to other algorithms. Taking the conventional 50% occluded face image as an example (to avoid the copyright problem, the non-occluded face image adopts mosaic processing), the face reproduction effect under the mask of each algorithm is shown in Figure 10.

In Figure 10, the face image reproduced by MFR-GAN was the most similar to the original face image and could accurately reproduce face features. Although FMODR could reproduce a clear face image, it had poor consistency with the original face. For example, the face length was shorter than the original face, and the lip shape was also different. The quality of the face image reconstructed by GAN is poor, and the generated image resolution is low, so it can not be effectively recognized. To verify the reproduction effect of FRMUM model based on MPCAnet and MFR-GAN, the model is applied in a factory to recognize the identity of employees. The actual application effect is shown in Table 3.

(a) Origin face image (b) Mask image (c) MFR-GAN (d) FMODR (e) GAN

Figure 10: Face reproduction effect under masks for various algorithms.

Table 3: Actual application effect of FRMUM model.

Metric	FRMUM	Traditional Manual Review
Recognition Accuracy	98.68%	85.00%
Misrecognition Rate	1.32%	15.00%
Recognition Time (seconds per person)	0.86	5
Robustness (Recognition Rate under Occlusion)	97.58%	60.00%
System Stability	High	Low
Data Processing Capacity (people per hour)	4,186	720
Cost (Labor Cost per Hour)	Low	High

In Table 3, the recognition accuracy of FRMUM reached 98.68%, which was significantly higher than 85.00% of traditional manual review. This shows that FRMUM model has higher accuracy in automatic recognition. The misrecognition rate of FRMUM model was only 1.32%, which was far lower than 15.00% of the traditional manual review. This shows that FRMUM model performs well in reducing false recognition. The recognition time includes the complete process of preprocessing, feature extraction, and matching, which is a time-consuming end-to-end process. The average recognition time of FRMUM model was 0.86 seconds/person, while the traditional manual review required 5 seconds/person. This shows that FRMUM model has significant advantages in efficiency. The study calculated the ratio of correctly recognized faces to the total occluded face samples based on images of employees wearing masks. The average recognition rate under occlusion was 97.58%, while traditional manual review was only 60.00%. This shows that FRMUM model has higher robustness when dealing with complex scenes. FRMUM model has high system stability and can run continuously and stably, while the traditional manual review is greatly affected by human factors and has low stability. FRMUM model could handle 1,200 people per hour, while the traditional manual review was only 100 people. This shows that FRMUM model has obvious advantages in large-scale data processing. FRMUM model has low operation cost, while traditional manual review requires high labor cost. FRMUM model performs well in practical application, which can significantly optimize the recognition efficiency and accuracy, while reducing the misrecognition rate and labor cost.

3.3 Computational complexity analysis

In terms of computational complexity analysis, MPCANet achieved efficient feature extraction and classification through multi-scale feature fusion and spatial pyramid pooling. According to the experimental results, the operation time of MPCANet was only 0.87 seconds/sample, significantly lower than that of PCANet (2.64 seconds) and DeepMaskNet (1.92 seconds). Its lightweight characteristic stems from the following three aspects:

Model structure optimization: The principal Component Analysis (PCA) filter is adopted to replace the traditional convolution kernel. After combining with the random forest classifier, the parameter size is approximately 1.2MB. Compared with the ten-layer deep network of DeepMaskNet (with a parameter size of approximately 15.6MB), the model size is reduced by 92.31%.

Computational load control: Spatial pyramid pooling divides the feature map into multi-scale blocks (such as 8×5 windows), and replaces fully connected operations with local histogram statistics, reducing the number of Floating-point Operations Per Second (FLOPs) to 4.8×104. It is 60% less than the traditional PCANet (1.2×105).

Hardware adaptability: The channel cascade fusion strategy only increases memory usage by 15%, but enhances the feature expression ability. Meanwhile, the progressive trained MFR-GAN is generated through hierarchical resolution, controlling the video memory consumption within 8GB and adapting to mainstream GPU devices. MPCANet has low computational complexity and can meet the deployment requirements of edge devices.

4 Discussion

The performance of MPCANet and MFR-GAN was evaluated in experiments. The results showed that the research method had significant advantages in mask face recognition tasks. Compared with PCANet and DeepMaskNet, MPCANet had an average accuracy of 94.24%, which was 15.73% and 9.17% higher than the other two methods, respectively. Its 0.87-second computation time was also better than other methods. This indicates that MPCANet has faster processing speed while maintaining high accuracy. MFR-GAN also performs

outstandingly on image realism and PSNR indicators, achieving a realism of 0.95 and a PSNR value of 30.25 dB at a 30% occlusion ratio, both superior to FMODR and GAN. Statistical significance analysis showed that the performance difference between this research method and other methods was statistically significant ($p<0.05$), indicating that this method had significant advantages. In terms of realism and PSNR improvement, the progressive training and multi-scale loss function of MFR-GAN can better restore image details and structures, resulting in more realistic and high-quality generated images. In specific scenarios, such as partial occlusion and low lighting conditions, the multi-scale feature fusion and channel level linking of MPCANet can extract richer feature information, making it more robust under complex occlusion and lighting changes. MFR-GAN can more accurately restore the details of occluded areas through progressive training and multi-scale loss functions, thereby achieving better recognition and repair effects in specific scenes.

5 Conclusion

In this study, an improved PCAnet (MPCAnet) based on multi-scale feature fusion and MFR-GAN algorithm based on progressive training were proposed for face recognition under mask occlusion, and FRMUM model was constructed for face feature recognition and reproduction. MPCAnet significantly improved the recognition performance of occluded faces through multi-layer feature fusion and spatial pyramid pooling. MFR-GAN generated high-quality and realistic face images through progressive training and multi-scale loss function. The average accuracy of MPCAnet in face recognition under mask occlusion was 94.24%, and the operation time was 0.87 seconds, which was better than that of PCAnet (78.51%) and DeepMasknet (85.07%). To verify the statistical significance of this improvement, independent sample t-tests were conducted on the same test set to evaluate the accuracy of different methods. The test results indicated that there was a statistically significant difference ($p<0.05$) in performance between MPCANet, PCANet, and DeepMaskNet, thus confirming the significant advantage of MPCANet in improving the performance of occluded face recognition. In addition, the 95% confidence interval for the accuracy of MPCANet was [93.56%, 94.92%], indicating the high stability. The authenticity of MFR-GAN under 30% occlusion ratio was 0.95, and the PSNR value was 30.25 dB, which was better than GAN and FMODR. FRMUM model performed well in practical application, the recognition accuracy was 98.68%, the misrecognition rate was 1.32%, and the recognition rate in the occlusion was still 97.58%. To sum up, the FRMUM solves the face recognition and reproduction under the mask, significantly improves the recognition efficiency and accuracy, and reduces the misrecognition rate and labor cost. However, FRMUM model has poor recognition effect in extreme occlusion such as double occlusion of goggles and masks. In the future, cross-scene transfer learning will be explored to improve the adaptability of FRMUM model under different occlusion types.

6 Funding

The research is supported by “Research and Practice on the Teaching Mode of Industry-Education Integration for Automation-related Majors in Higher Vocational Colleges under the Trend of Industrial Internet” of Jiangsu Province Higher Education Teaching Reform (Project No.: 2023JSJG624); The research project on “Research on the Construction of High-level Information-based Teaching Capacity of Double-qualified Teacher Teams in Vocational Colleges” of Jiangsu Province Education Science Planning (Project No.: ZJCX/2022/10); The training project of Jiangsu Province’s Sixth Batch “333 High-level Talent Cultivation Project” (Document No.: Su Ren Cai Ban [2022] No. 2).

References

[1] Suo Gao, Rui Wu, Xingyuan Wang, Jiafeng Liu, Qi Li, and Xianglong Tang. EFR-CSTP: Encryption for face recognition based on the chaos and semi-tensor product theory. Information Sciences, 621(1): 766-781, 2023. https://doi.org/10.1016/j.ins.2022.11.121

[2] Adrian Sean Bein, and Alexander Williams. Development of deep learning algorithms for improved facial recognition in security applications. IAIC Transactions on Sustainable Digital Innovation (ITSDI), 5(1): 19-23, 2023. https://doi.org/10.34306/itsdi.v5i1.605

[3] Xintao Li, and Hongyan Guo. Fusion of deep convolutional neural networks and brain visual cognition for enhanced image classification. Informatica, 49(16), 2025. https://doi.org/10.31449/inf.v49i16.7787

[4] Mohammed Taha, Tarek Mostafa, and Tarek Abd El-Hafeez. A novel hybrid approach to masked face recognition using robust PCA and GOA optimizer. Scientific Journal for Damietta Faculty of Science, 13(3): 25-35, 2023. https://doi.org/10.21608/sjdfs.2023.222524.1117

[5] Slim Ben Chaabane, Mohammad Hijji, Rafika Harrabi, and Hassene Seddik. Face recognition based on statistical features and SVM classifier. Multimedia Tools and Applications, 81(6): 8767-8784, 2022. https://doi.org/10.1007/s11042-021-11816-w

[6] Fahima Tabassum, Md. Imdadul Islam, Risala Tasin Khan, and M.R. Amin. Human face recognition with combination of DWT and machine learning. Journal of King Saud University-Computer and Information Sciences, 34(3): 546-556, 2022. https://doi.org/10.1016/j.jksuci.2020.02.002

[7] Ziwei Song, Kristie Nguyen, Tien Nguyen, Catherine Cho, and Jerry Gao. Spartan face mask detection and facial recognition system. Healthcare, 10(1): 87-88, 2022. https://doi.org/10.3390/healthcare10010087

[8] Pedro C. Neto, João Ribeiro Pinto, Fadi Boutros, Naser Damer, Ana F. Sequeira, and Jaime S. Cardoso. Beyond masks: On the generalization of masked face recognition models to occluded face recognition. IEEE Access, 10(1): 86222-86233, 2022. https://doi.org/10.1109/ACCESS.2022.3199014

[9] Gagandeep Kaur, Ritesh Sinha, Puneet Kumar Tiwari, Srijan Kumar Yadav, Prabhash Pandey, Rohit Raj, Anshu Vashisth, and Manik Rakhra. Face mask recognition system using CNN model. Neuroscience Informatics, 2(3): 100035-100036, 2022. https://doi.org/10.1016/j.neuri.2021.100035

[10] Manisha Kumari Meena, Hemant Kumar Meena, and Ramnivas Sharma. Partially occluded face reconstruction using graph-based algorithm. Journal of Electrical Engineering & Technology, 19(6): 3655-3664, 2024. https://doi.org/10.1007/s42835-024-01995-5

[11] Jianchu She, and Ying Liu. Facial image inpainting algorithm based on attention mechanism and dual discriminators. Journal of Image and Graphics, 10(1): 43-49, 2022. https://doi.org/10.18178/joig.10.1.43-49

[12] Changxiu Dai, and Xianfeng Zeng. Occluded face recognition network based on DCGAN and ResNet. Procedia Computer Science, 243(1): 724-733, 2024. https://doi.org/10.1016/j.procs.2024.09.087

[13] Yong Wang, Xiaoan Zhan, Tong Zhan, Jiahao Xu, and Xinzhu Bai. Machine learning-based facial recognition for financial fraud prevention. Journal of Computer Technology and Applied Mathematics, 1(1): 77-84, 2024. https://doi.org/10.5281/zenodo.11004115

[14] Le Sun, Guangrui Zhao, Yuhui Zheng, and Zebin Wu. Spectral–spatial feature tokenization transformer for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 60(1): 1-14, 2022. https://doi.org/10.1109/TGRS.2022.3144158

[15] Ahmed Rajab Kadhim, Raidah S. Khudeyer, and Maytham Alabbas. Facial sentiment analysis using convolutional neural network and fuzzy systems. Informatica, 48(12), 2024. https://doi.org/10.31449/inf.v48i12.6151

[16] Md. Saikat Islam Khan, Nazrul Islam, Jia Uddin, Sifatul Islam, and Mostofa Kamal Nasir. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. Journal of King Saud University-Computer and Information Sciences, 34(8): 4773-4781, 2022. https://doi.org/10.1016/j.jksuci.2021.06.003

[17] Naeem Ullah, Ali Javed, Mustansar Ali Ghazanfar, Abdulmajeed Alsufyani, and Sami Bourouis. A novel DeepMaskNet model for face mask detection and masked facial recognition. Journal of King Saud University-Computer and Information Sciences, 34(10): 9905-9914, 2022. https://doi.org/10.1016/j.jksuci.2021.12.017

[18] Feihong Lu, Hang Chen, Kang Li, Qiliang Deng, Jian Zhao, Kaipeng Zhang, and Hong Han. Toward high-quality face-mask occluded restoration. ACM Transactions on Multimedia Computing, Communications and Applications, 19(1): 1-23, 2023. https://doi.org/10.1145/3524137