IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 24 June 2022, accepted 10 July 2022, date of publication 15 July 2022, date of current version 22 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3191113

RESEARCH ARTICLE

Face Recognition With Masks Based on Spatial Fine-Grained Frequency Domain Broadening

HUA-QUAN CHEN[1,2,3], KAI XIE[1,2,3], MEI-RAN LI[1,2,3], CHANG WEN[3,4], AND JIAN-BIAO HE[5]

[1]School of Electronic Information, Yangtze University, Jingzhou 434023, China
[2]National Demonstration Center for Experimental Electrical and Electronic Education, Yangtze University, Jingzhou 434023, China
[3]Western Institute, Yangtze University, Karamay 834000, China
[4]School of Computer Science, Yangtze University, Jingzhou 434023, China
[5]School of Computer Science and Engineering, Central South University, Changsha 410083, China

Corresponding author: Kai Xie (500646@yangtzeu.edu.cn)

This work was supported in part by the Natural Science Foundation of Xinjiang Uygur Autonomous Region under Grant 2020D01A131, in part by the Fund of Hubei Ministry of Education under Grant B2019039, in part by the Graduate Teaching and Research Fund of Yangtze University under Grant YJY201909, in part by the Teaching and Research Fund of Yangtze University under Grant JY2019011, in part by the Undergraduate Training Programs for Innovation and Entrepreneurship of Yangtze University under Grant Yz2021040, and in part by the National College Student Innovation and Entrepreneurship Training Program under Grant 202110489003.

ABSTRACT Along with social distancing, wearing masks is an effective method of preventing the transmission of COVID-19 in the ongoing pandemic. However, masks occlude a large number of facial features, preventing facial recognition. The recognition rate of existing methods may be significantly reduced by the presence of masks. In this paper, we propose a method to effectively solve the problem of the lack of facial feature information needed to perform facial recognition on people wearing masks. The proposed approach uses image super-resolution technology to perform image preprocessing along with a deep bilinear module to improve EfficientNet. It also combines feature enhancement with frequency domain broadening, fuses the spatial features and frequency domain features of the unoccluded areas of the face, and classifies the fused features. The features of the unoccluded area are increased to improve the accuracy of recognition of masked faces. The results of a cross-validation show that the proposed approach achieved an accuracy of 98% on the RMFRD dataset, as well as a higher recognition rate and faster speed than previous methods. In addition, we also performed an experimental evaluation in an actual facial recognition system and achieved an accuracy of 99%, which demonstrates the effectiveness and practicability of the proposed method.

INDEX TERMS Face recognition with mask, convolutional neural network, frequency domain widening, bilinear module, RMFRD dataset.

I. INTRODUCTION

The COVID-19 pandemic has substantially disrupted everyday life worldwide, and wearing masks is now generally recommended when traveling to prevent the spread of the virus. However, opaque masks occlude important facial feature information such as the nose and the mouth, leading to a reduction of the available facial feature information. Consequently, the amount of information that can be extracted by neural network models is reduced. However, because masks are worn to reduce the propagation of an infectious disease, they cannot be removed to enable facial recognition systems to perform feature extraction.

Therefore, the enrichment of the effective features of face wearing masks has become an important issue in the field of face recognition, along with improving the ability of neural network models to extract non-occluded features. Companies such as Baidu and Sense Time have carried out relevant research and achieved certain results. For example, more feature information may be extracted from the eyes and eyebrows of a person by determining the occluded position, removing the occlusion, and reconstructing the facial image to increase the effective information of the face wearing the mask, resulting in improved facial recognition accuracy.

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

We divided the current facial recognition methods designed for people wearing masks into three groups, including local feature-, image restoration-, and deep learning-based methods.

Regarding the local feature method, the traditional principal component analysis method relies too heavily on the global features of the face; therefore, the global features are destroyed for occluded facial images. Currently, global features are not as robust as local features and are not suitable for facial recognition problems under occlusion. Cheng *et al.* [1] proposed a local nonnegative matrix factorization algorithm to enhance the clarity of local features and represent the non-occluded area of a face better. However, the error caused by such occlusions does not conform to the Gaussian distribution assumed by this algorithm, as it was found that robustness to the occluded area was insufficient. From a different perspective, Min *et al.* [2] proposed the possibility of using local Gabor binary pattern features and occlusion, which solved the problem of facial recognition under conditions of occlusion to a certain extent. Wei *et al.* [3] proposed an occluded-face recognition algorithm using PCANet by intercepting facial feature blocks, performing feature extraction on each type of feature block, and then performing feature extraction based on most common occlusion types. Feature splicing and zero-filling operations input the feature matrix after operation into the traditional support-vector machine (SVM) algorithm for classification. However, in the case of occlusion with this algorithm, the accuracy of the feature point positioning directly affects the final recognition result. To accurately locate feature points under occlusion, Xing *et al.* [4] proposed a facial feature point-location algorithm based on adaptive features for conditions of occlusions. This algorithm used traditional logistic regression algorithms to detect the occlusion state of each feature point. The probability value of each feature point was estimated, and the weight of the feature point was adjusted according to a probability value, thereby reducing the impact of occlusion on facial features and improving the accuracy of feature-point positioning. However, the local feature method has random, unstable, and poor real-time performance in cases of multi-posture and multi-angle occlusion.

In recent years, as occlusion has become a major problem in facial recognition technology, researchers have begun to use image restoration methods to fill and repair damaged images, including occluded face images. The primary assumption of this method is that occlusion accounts for a small proportion of the entire face. Using the redundancy of image information, the gray value of the unoccluded area is used to fill or smooth the occluded area, and the restored image is then used for face recognition. To deal with the occlusion problem, Chen *et al.* [5] developed an occlusion-aware GAN. A pre-trained GAN was used to recover the associated occluded areas. They used original non-occluded faces to train the GAN. Duan and Zhang [6] suggested an end-to-end BoostGAN model for occluded facial recognition, in which the occluded picture was first synthesized, and the non-occluded image was then utilized to perform refined face recognition. In contrast, GAN-based approaches have thus far been unable to reproduce the intricacies of critical locations on the face, especially for large-area occlusions such as facemasks. Ge *et al.* [7] used a GAN to improve the ability of a trained face recognizer to perceive occluded faces. The recognizer used a set of identity-centric features as source data for supervised learning, such that the repaired faces were gathered in an identity center to distinguish the diversity in a given identity class, which resulted in an accuracy of 92.26% on a popular benchmark dataset. Yuan and Park [8] proposed a GAN framework conditioned on a 3D deformable model and consisting of a generator and two discriminators. It used 3DMM before a GAN network, and combined a global and local GAN network to learn a face de-occlusion model, remove the face occlusion, restore the occluded area, and reconstruct a 3D face model used to perform identification. In facial recognition based on image restoration algorithms, occlusion recognition is more effective for small areas of the face, and the generalization ability of recognition models improves. However, when major elements of a face are hidden, such as the eyes and lips, it remains difficult to determine the identity of the individual matching with a given face, and recognition performance remains relatively poor.

Mundial *et al.* [9] and others proposed enhancing the ability of existing face mask recognition technology by using supervised learning methods to recognize faces wearing masks and used deep neural networks (ZF-Net) to extract facial features. After collecting a dataset of faces wearing masks, a support-vector machine classifier was trained based on the extracted facial recognition feature vectors. This method exhibited an accuracy of 97% for facial recognition of people wearing masks. However, their results showed a large deviation in robustness for images taken from the side. Golwalkar and Mehendale [10] proposed the FaceMaskNet-21 deep learning network to extract 128-d codes from static images, real-time video streams, and video files to aid in facial recognition. HOG technology was used to rapidly identify faces wearing masks. When the execution time was less than 10 ms, an accuracy of 88.92% was achieved. Wan and Chen [11] proposed a trainable module called MaskNet, which learned an appropriate method to adaptively generate different feature-mapping masks suitable for different occluded facial images. MaskNet automatically assigned higher weights to hidden units activated by non-occluded facial parts, whereas lower weights were assigned to hidden units activated by occluded facial parts. An accuracy of 93.8% was obtained on a self-made dataset. The occluded facial recognition system based on deep learning proposed by Wu *et al.* [12] used a triple loss function to measure the prediction performance of the model. First, the input face image was cut into fixed-size samples, which were processed with an Inception-ResNet-vl network. The features of all images were mapped to the hypersphere through the L2 function, which finally passed through the embedding layer and used the triple loss function to perform learning.

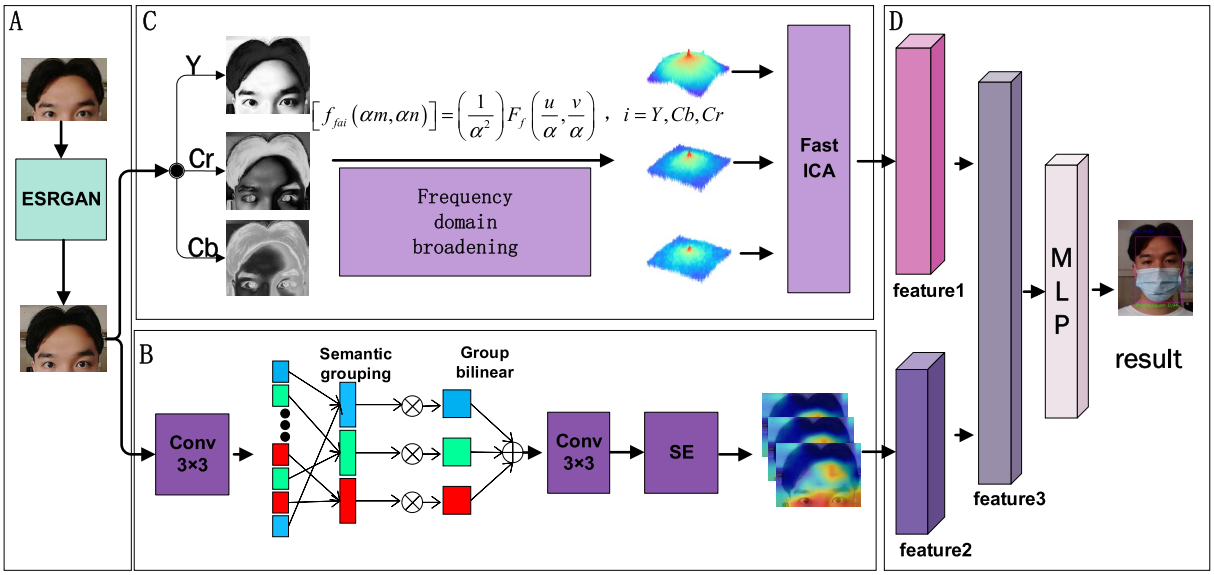


FIGURE 1. Masked face recognition algorithm. A:ESRGAN is used to perform image preprocessing; B: This is part of DBT-MBConvBlock, in which a SG semantic grouping layer and GB group bilinear layer are used to perform spatial fine-grained feature extraction. C: The FBB module, in which we perform frequency-domain broadening of the input image and Fast ICA feature reduction. D: MLP is used to perform feature classification.

Therefore, the most fundamental problem of facial recognition of masked faces is to increase the effective feature information of the face to solve the lack of facial feature information. Therefore, we extracted features from the frequency and spatial domains to enrich the facial feature information and improve the accuracy of facial recognition with masks. Figure 1 shows a flowchart of the proposed algorithm.The main contributions of this study are summarized as follows.

1) We found that preprocessing images of mask-wearing faces with an enhanced super-resolution generative adversarial network (ESRGAN) helped improve the recognition rate.
2) We improved EfficientNet with a deep bilinear module, which was applied to perform feature extraction of the non-occluded region of masked faces, effectively improving the rate of recognition.
3) To the best of our knowledge, this study is the first to apply a frequency-domain broadening method to the task of recognizing masked faces, and we present results that demonstrate that this approach improves on the performance of existing methods.

II. METHOD

A. IMAGE PREPROCESSING

1) SET THE REGION OF INTEREST (ROI)

Image processing is indispensable in all facial recognition systems. We used RetinaFace as a face detector to crop the detected faces and save them in alignment, and to crop out the unoccluded face areas. These were then used as input for color-space fixed bounding box (FBB) and neural network methods.

2) ESRGAN DEBLURRING

The proposed method uses OpenCV to perform region of interest (ROI) cropping on facial images to recover an image of the non-occluded area. To avoid pixelation of the facial image after ROI processing, in which the image loses detail [13], we adopt an ESRGAN model [14] to enhance the details of the non-occluded image. ESRGAN can generate real textures and enhance detailed features through image super-resolution. ESRGAN also benefits from dense connections (as proposed by DenseNet), which not only increases the depth of the network, but also enables more complex structures, enabling the network to learn finer details. Learning to normalize the data distribution between layers is a common practice in several deep neural network models; however, ESRGAN does not use batch normalization (BN). A BN layer normalizes the testing data by using the mean and variance normalization features of a batch of data during training, and normalizes the testing data by using the mean and variance estimated on the entire training set during testing. When the statistical results of the training and test sets differ significantly, the BN layer may limit the generalization ability of the model. Deleting batch standardization can improve stability and reduce computational costs (reduce learning parameters). The discriminator estimates the probability that a real image is relatively more realistic than a fake image. That is, the standard discriminator is replaced by a relativistic average discriminator (RaD); therefore, the loss function of the discriminator is defined as

$$L_G^{Ra} = -E_{x_r}\left[\log(H_1)\right] - E_{x_f}\left[\log(1-H_2)\right] \quad (1)$$

The corresponding loss function of the generator is

$$L_G^{Ra} = -E_{x_r}\left[\log(1-H_1)\right] - E_{x_f}\left[\log(H_2)\right] \quad (2)$$

where $H_1 = D_{Ra}(x_r, x_f)$, $H_2 = D_{Ra}(x_f, x_r)$. The loss of the confrontation includes x_r and x_f; therefore, a low-resolution image passes through the generator, which benefits from the gradient of the generated data and the actual data in the

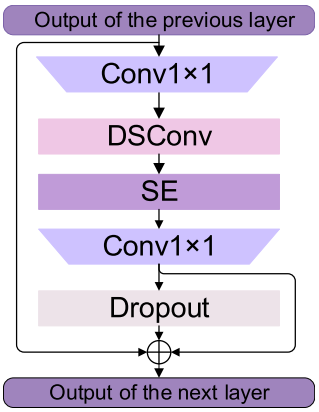


FIGURE 2. MBConvBlock structure diagram.

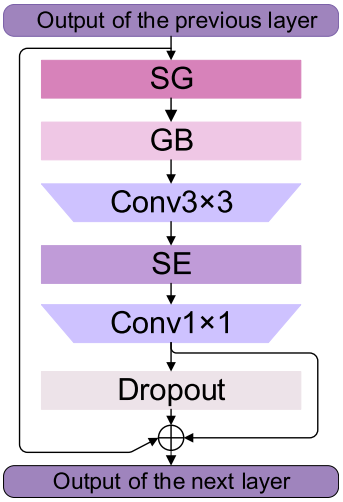


FIGURE 3. DBT-MBConvBlock structure diagram.

adversarial training. This adjustment enables the network to learn sharper edges and more detailed textures to enhance the details of low-pixel ROI face images.

B. BACKBONE: COMBINATION OF EFFICIENTNET AND DBT BLOCK

The backbone network is primarily used to extract features of cropped face ROI images. Based on the EfficientNet [15] network, we replaced the Conv1 × 1 and DSConv parts of MBConvBlock with a semantic grouping (SG) layer and a (GB) group bilinear layer in a depth bilinear transformation (DBT) [16], as shown in Figure 2 and Figure 3. In the original MBConvBlock structure, Conv1 × 1 was used to extract the feature graph output by the previous layer. We chose SG and GB to replace Conv1 × 1 and remove DSConv to improve the ability of deep feature extraction and speed up feature extraction.

1) DBT BLOCK

The main part of the DBT is shown in Figure.5. In Figure 5, the semantic grouping steps arrange the related features according to the corresponding region. Red, green and purple represent eyebrows, eyes, and other features, respectively. The group bilinear step performs linear operations on the features after interaction between local features and components, and finally aggregates the interaction features in the group to obtain the aggregation features. The constraint of semantic grouping arranges the channels of the convolution feature. Because each channel of the convolution feature exhibits a high response to a specific semantic image pattern, the response of all the convolution channels describing a certain semantic is be concentrated in the corresponding spatial region. The convolution channels in the same semantic group overlap as much as possible, but the channels of different components do not overlap as much as possible owing to the semantic grouping constraint. This paired grouping constraint enables subsequent grouping bilinear operations to capture the detailed characteristics of each semantic block better. Feature learning based on bilinear transformation can obtain a detailed information expression through a cross-product operation on feature vectors and realize pairwise interactions between channel information. DBT is shown in Figure 5. Its input is the convolution feature of the image. Firstly, the convolutional feature channels are grouped according to the semantic information of parts, and the convolutional features are obtained after grouping. For the feature vectors of each part, the features are further divided into the same group, and the bilinear operation is carried out to obtain the rich detailed expressions of this part. Then, the bilinear expressions between different groups are added to obtain the features as shown in Formula (3). We vectorize the expression as follows to recover the original dimension of the convolution feature.

$$y_i = \Gamma_B(Ax_i) = vec\left(\sum_{j=1}^{G}\left(\left(I_j A x_i + p_j\right)\left(I_j A x_i + p_j\right)^T\right)\right) \tag{3}$$

2) DBT-MBCONVBLOCK

Because we found DBT to be an excellent way to extract rich details from images, we used EfficientNet MBConvBlock's image detail representation to improve the efficiency of DBT t in recognizing faces wearing masks. In addition, the use of depth-separable convolution in the shallow layer of the network slows down the feature extraction speed [17]. Although the deeply detachable volume structure uses fewer parameters than ordinary convolution and smaller FLOPs, some existing accelerators are usually not fully utilized. Although its theoretical computation is small, this approach does not perform well in practice. To accelerate the convolution speed of a shallow network, depthwise convolutions (DSConv) were removed from MBConvBlock [18]. To enrich the expression of image details and ensure the operation efficiency of the network without adding more FLOPs, we replaced the extended Conv1 ×1 and DSConv convolutions in the main branch of MBConvBlock with the

SG semantic grouping layer and the GB group bilinear layer in DBT. We named the MBConvBlock after the above processing dbT-MBConvblock, and its structure is shown in Figure 3. The input of DBT-MBConvblock is the feature map output of the upper layer. These feature maps perform a 3×3 convolution through SG semantic grouping layer and GB group double-line layer, and use SE channel focus [19] to enhance feature extraction of useful information and suppress useless features. After passing the dimension of the 1×1 convolutional layer, the obtained feature map is restored to the original feature dimension of the layer through upgrading, and some parameters are discarded at the dropout layer to prevent network overfitting and speed up network convergence. Next, the feature map is added in parts to that of the previous layer. The obtained feature map is superimposed with that of the 1×1 convolutional layer as the output of DBT-MBConvblock.

In this study, except for the first 3×3 convolutional layer of the network, the last 1×1 convolutional layer, pooling layer, and full connection layer, other parts of the network are stacked by MBConvBlock and DBT-MBConvblock. The network structure is shown in Figure 4. The input image is convolved with a normal size of 3×3, and the extracted convolution features pass through the DBT-MB Convblock layer with an expansion ratio of 1. To obtain more convolution images, we repeatedly stacked four layers of DBT-MBConvblocks with a spread ratio of 4 and a convolution step of 2. To obtain deeper features, we again stacked four layers of DBT-MBConvblocks with a spread ratio of 4 and a convolution step of 2. We also used MBConv4 and MBConv6 blocks in the network structure of the proposed approach. The convolution kernel of these two types of MBConv was 3×3. The structure of the MBConvBlock is shown in Figure 2.

C. FREQUENCY DOMAIN FEATURE EXTRACTION

We also use a wrapping discrete curvelet transform for face images in the non-occluded part of face images, and FastICA [20] for de-correlation and feature reduction. We refer to these processes as an FBB module.

1) FREQUENCY DOMAIN BROADENING FEATURE

However, using only convolutional neural network to extract features of non-occluded areas in the spatial domain is obviously insufficient. Mandal *et al.* [21] achieved 47% face recognition accuracy in the masked face dataset by resnet50. Li *et al.* [22] also analyzed the facial recognition accuracy of some convolutional neural network models on datasets of masked faces in an experiment, but all of them had certain limitations and failed to achieve good results. Li *et al.* [23] used the robustness of two-dimensional discrete cosine transform (2DDCT) to convert image signals from the spatial domain to the frequency domain to reduce the impact of noise on the original image and classify face images, and achieved good results. Therefore, we aimed to further improve the accuracy of the convolutional neural network by extracting the information of occluded faces from the frequency domain.

FIGURE 4. Improved EfficientNet.

At present, most face image acquisition methods use color images, and discarding color information significantly reduces the information contained the image. The appearance of images of faces can change dramatically under different lighting conditions. Bright or dark light may alter the RGB information collected, resulting in inaccurate recognition in the RGB [24] color space. Because YCbCr [25] has a high degree of de-correlation, we use YCbCr color space to broaden the frequency domain. First, we convert the RGB channel of the image into YCbCr color space, and the face image after image processing is defined as follows.

$$\mathrm{f}(m,n) = \begin{cases} f_{ff}(m,n)\ (m,n) \in B_f \\ f_{fb}(m,n)\ (m,n) \in B_b \end{cases} \tag{4}$$

$f_{ff}(m,n)$ denotes the face image,and$f_{fb}(m,n)$ denotes background image. We set the background $f_{fb}(m,n)$ to 0. Candes *et al.* [26] proposed two independent discrete curvelet transform (DCT) algorithms, which improved the processing speed of the transform. The first algorithm was an unequal-step fast Fourier transform (FFT), in which the curvelet coefficients are obtained by irregularly sampling the Fourier coefficients of the image. The second algorithm was wrapping discrete curvelet Transform (WDCT), which uses a range of conversion and encapsulation technologies with shorter computation times and higher operational efficiency. Therefore, the proposed approach uses a wrapping discrete curvelet transform on the unoccluded areas of the above three channels. The slope of the same interval is introduced in this algorithm.

$$\tan\theta_l = l \times 2^{-\lfloor j/2 \rfloor}, \quad l = -2^{\lfloor j/2 \rfloor}, \ldots, 2^{\lfloor j/2 \rfloor} - 1 \tag{5}$$

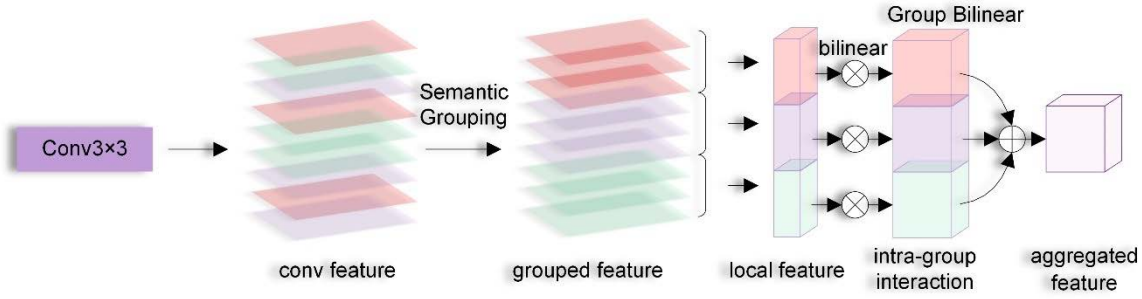


FIGURE 5. Overview of the deep bilinear transformation.

and the following is defined.

$$\hat{U}_{j,l}(\omega) = W_j(\omega)V_j(S_{\theta_l}\omega) \tag{6}$$

where shear matrix S_{θ_l} is defined as:

$$S_{\theta_l} = \begin{bmatrix} 1 & 0 \\ -\tan\theta & 1 \end{bmatrix} \tag{7}$$

The discrete Curvelet transform is defined as:

$$c(j, l, k) = \int \hat{f}(\omega)\hat{U}_j(S_{\theta_l}^{-1}\omega)\exp\left(i\langle b, \omega\rangle\right)d\omega \tag{8}$$

$b \in (k_1 \times 2^{-j}, k_2 \times 2^{-j/2})$.

The working steps of wrapping discrete curvelet transform algorithm are as follows.

(1) The Fourier transform of the face image is given as follows.

$$\left[f_{fa}(\alpha m, \alpha n)\right] = \left(\frac{1}{\alpha^2}\right) F_f\left(\frac{u}{\alpha}, \frac{v}{\alpha}\right) \tag{9}$$

According to Formula 7, expand the frequency domain of the three channels in YcbCr [27] is given as follows.

$$f_i = \left[f_{fai}(\alpha m, \alpha n)\right] = \left(\frac{1}{\alpha^2}\right) F_f\left(\frac{u}{\alpha}, \frac{v}{\alpha}\right), \quad i = Y, Cb, Cr \tag{10}$$

(2) The equation given below is used to generate the product of different angles L and scale j.

$$\mathrm{U}_{j,l}[m, n] \cdot f_i[\alpha m, \alpha n] \quad i = Y, Cb, Cr \tag{11}$$

Product results of wrapping wave coefficients [28].

$$\tilde{f}_{j,l}[m, n] = W\left(U_{j,l}\right) \cdot f_i[\alpha m, \alpha n] \quad i = Y, Cb, Cr \tag{12}$$

The characteristic coefficients of the curvilinear transform are obtained by the inverse two-dimensional discrete Fourier transform. The curvature coefficients obtained by discrete curvature transform are multi-scale and multi-directional, and can represent human facial features well. However, direct use of curvilinear coefficients not only exhibits poor recognition performance, but also leads to information redundancy. Therefore, the feature with the maximum curvelet coefficient in each dimension and each scale-corresponding pixel is selected as a candidate feature to reduce the redundancy of information [29]. To reduce the redundancy of the scale features of the fusion curve coefficients, we use fast independent component analysis (FastICA) to process these coefficients to obtain useful correlation information. FastICA can remove correlations between variables while preserving higher-order statistics. We assume that features undergo independent component analysis and remove the correlation between different features as follows.

$$\mathrm{FV}_{fre} = X = \{x_1, x_2, \ldots, x_m\} \tag{13}$$

After the depth feature and frequency domain detail feature are obtained, the parallel fusion method is selected. The features to be merged must be extended and normalized. Assume that CNN feature vector and frequency domain feature vector are respectively FV_{CNN} and FV_{fe}, Then, the normalized combination feature vector is expressed as.

$$FV_i^n = \frac{FV_i - \mu_i}{\sigma_i}, \quad i = CNN, fre \tag{14}$$

$$FV = \left[FV_{CNN}^n FV_{fre}^n\right]^T \tag{15}$$

D. FEATURE CLASSIFICATION

1) MULTILAYER PERCEPTRON FEATURE CLASSIFICATION

When we obtain the spatial and frequency-domain characteristic FV, we classify the FV by using multi-layer perceptual classifier [30]. Each test image is assigned its identity information. Each face is defined by a feature vector, which is defined as $P = [p_1, \ldots, p_k]$. P_k is the input of MLP. The number of hidden layers we set to $\frac{\eta}{2}$, where η is the class number. Finally, the sigmoid function in MLP is used to output the final classification result.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTAL CRITERIA AND SETUP

1) EXPERIMENTAL CRITERIA

We used the accuracy rate as the main indicator to evaluate the recognition performance of the proposed model. Other indicators such as precision, recall, area under the receiver

operating characteristic (ROC) curve (AUC), and ROC were used to further evaluate its performance.

The formulas for accuracy, precision, recall, AUC, and ROC are given as follows.

Accuracy is defined as the proportion of correct classifications to the total number of samples.

$$Accuracy = \frac{Number\ of\ correctly\ detected\ faces}{Total\ number\ of\ validated\ sets} \tag{16}$$

The formula for accuracy can also be expressed as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + TN} \tag{17}$$

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

$$TPR = \frac{TP}{TP + FN} \tag{20}$$

$$FPR = \frac{FP}{FP + TN} \tag{21}$$

where FP denotes false positive, TP, true positive, FN, false negative, and TN stands for true negative.

Recal rate indicates how many positive examples in the sample were predicted correctly. The abscissa of the ROC curve is the false positive rate (FPR), and the ordinate is the true positive rate (TPR). AUC is the area enclosed by the ROC curve and the horizontal axis, which can quantitatively reflect the performance of the model measured based on the ROC curve.

2) EXPERIMENTAL DATASET

The publicly available real mask face recognition dataset (RMFRD) mask-wearing faces was used to perform training and evaluation [31]. This sample set includes three datasets, referred to as the mask occlusion face detection dataset (MFDD), the mask occlusion face recognition dataset (RMFRD), and simulated mask occlusion face recognition dataset (SMFRD). In this study, we used RMFRD and SMFRD. The experiment used a cross-validation method to perform training on 4000 images, then used 800 images to perform training, and 4000 to verify that performance of the proposed approach.

RMFRD is among the most comprehensive datasets of masked faces available globally. It contains 5,000 masked faces and 90,000 unmasked faces from 525 people. A semi-automatic annotation method was used to crop out information-rich facial areas. Fig. 6 shows a number of pairs of images of faces from the RMFRD dataset.

SMFRD features 500,000 simulated faces with masks drawn from the labeled faces in the wild (LFW) [32] and Webface [33] databases. The simulation was performed using the Dlib library [34]. This dataset is balanced but more difficult to handle because the simulated mask is not always added in the right place. Figure7 depicts various examples of generated human faces wearing masks from the SMFRD dataset. It should be noted that in Section C:MASKED FACE RECOGNITION RESULT DISPLAY, we used the dataset we collected in the laboratory, and then used the proposed method to obtain the effect of training.

FIGURE 6. Pair of face images in the RMFRD dataset: face images without masks (top) and face images with masks (bottom).

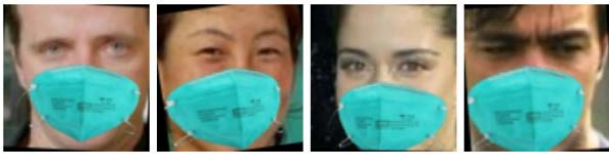


FIGURE 7. SMFRD dataset of faces wearing masks.

TABLE 1. Parameters of experimental platform.

Hardware environment	CPU	Intel Core i7- 10750H
	GPU	GeForceRTX2060
	RAM	16GB
Software environment	Platform	Windows10
		Tensorflow1.13.0
		CUDA10.0+Cudnn7.4
		Pycharm+Python3.7

3) EXPERIMENTAL PLATFORM

Parameters of the experimental platform are listed in Table 1

B. EXPERIMENTAL DETAILS

The maximum number of training batches used in this experiment was 8000. The initial learning rate was set to 0.001, the RMSProp optimizer was used with a decay of 0.96, a momentum of 0.9, and a batch norm momentum of 0.98. Considering the capacity of the GPU, the batch size was set to 32. The best model was saved every 500 iterations from 0 to 8000 iterations of training. After the training was completed, the model with the best performance among the 16 saved models was selected. The key parameters of the proposed network are listed in Table 2.

1) EFFICIENTNET WITH DBT MODEL

In the network parameter setting of the method used in this study, we selected φ =1, α = 1.2, β = 1.1, and

TABLE 2. Model parameter setting.

Parameters	Value
expansion ratio of DBT-MBConv_1 layer	1
expansion ratio of DBT-MBConv_2 layer	4
expansion ratio of DBT-MBConv_3 layer	4
Number of DBT-MBConv_1 layer	2
Number of DBT-MBConv_2 layer	4
Number of DBT-MBConv_3 layer	4
Number of MBConv_1 layer	6
Number of MBConv_2 layer	9
Number of MBConv_3 layer	15
se_ratio of MBConv_1	0.25
se_ratio of MBConv_2	0.25
se_ratio of MBConv_3	0.25

$\gamma = 1.15$ as ideal parameters, and maintained the optimal value of the performance of our method by using the search architecture. The performance of this model is shown in Figure 10. As shown in Figure 10, the accuracy of some common baseline models, ResNet50 [35] and Densenet169 [36], on RMFRD datasets was indeed improved to a certain extent after DBT was integrated. Our method was proposed based on EfficentNetv2-S, and ultimately surpassed that it in terms of accuracy, which is consistent with the conclusion discussed above. The parameters of our model were largely the same as those of the lightweight Efficentnet-B4 network, but our proposed approach achieved a higher accuracy, which was significant compared to the performance of current mainstream classification models on the RMFRD dataset. Therefore, removing the MBConvBlock of first-level and depthwise convolutions does not reduce the efficiency of the network. In contrast, when we used SG semantic grouping layer and GB group dual-line layer to replace them, we were able to improve the accuracy of face occlusion without adding an excessive amount of FLOPs.

2) ABLATION EXPERIMENT OF DBT AND FBB

To verify the effectiveness of the proposed method, we conducted ablation experiments on a deep bilinear module and a frequency domain broadening module (FBB), and the experimental results are shown in Table 3. We conducted experiments using the RMFRD and SMFRD datasets. As may be observed from the first row of the table, the accuracy shown on both datasets was lowest when only efficentNetv2-s was used to perform feature extraction. We retained the feature extraction network EfficentNetv2-S and added a deep bilinear module and a frequency-domain extension module before and after. The accuracy of this model on the RMFRD and SMFRD datasets was improved by 11.8% and 10.8% and 5.9% and 4.4%, respectively. When the depth bilinear (DBT) and frequency domain extension (FBB) modules were added in parallel to EfficentNetv2-S feature extraction network for feature extraction, as shown in Table 3, the accuracy of the model on the RMFRD and SMFRD datasets was improved

TABLE 3. Influence of different components of feature enhancement modules on network recognition rate.

EfficentNetV2-S	DBT	FBB	Dataset	Acc(%)
✓			RMFRD	69.1
			SMFRD	68.5
✓	✓		RMFRD	80.9
			SMFRD	79.3
✓		✓	RMFRD	75.4
			SMFRD	72.9
✓	✓	✓	RMFRD	91.8
			SMFRD	90.3

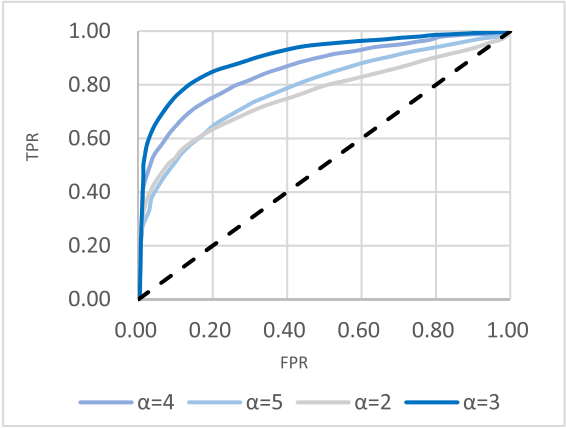


FIGURE 8. ROC curves corresponding to different compressibility coefficients α in time domain.

to 22.1% and 21.8%, respectively, when the two modules were not integrated. Experiments show that the depth bilinear module and frequency expansion module were able to effectively enhance the spatial and frequency domain features, respectively, and improved accuracy on the datasets of mask-wearing faces.

3) DIFFERENT TIME DOMAIN COMPRESSION COEFFICIENTS

α on FBB: The spatial compression coefficient [37] α determines the multiple expansion of the spatial signal in the frequency domain. We selected four compression coefficients $\alpha = 2$, $\alpha = 3$, $\alpha = 4$, and $\alpha = 5$ for the experiments; when the time-domain compression coefficient α was between 3 and 4, the value of AUC only differed by 1%, which is the ROC value when the feature is amplified in the frequency domain by a factor of 3. Therefore, different time-domain compression coefficients have a significant influence on the number of features in the non-occluded area of the facial image in the Y, Cb, and Cr channels. The number of features affects the recognition of the facial image by the entire model. Figure 8 shows the results of the experiments.

4) INFLUENCE OF DIFFERENT COLOR SPACES ON FBB

Color is the result of the perception of frequencies differences along the spectrum of visible light by the human eye.

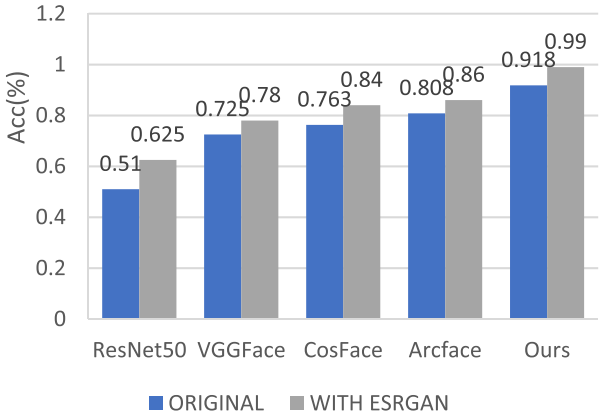


FIGURE 9. Performance on masked faces with and without ESRGAN image preprocessing on RMFRD dataset.

The difference of the same color in different color spaces has different degrees of influence on the two-dimensional Fourier transform of the image [38]. All colors can be obtained by mixing RGB according to the principles of colorimetry and optics. Therefore, we expanded the frequency domain of the input ROI facial image in the three common color spaces of YCBCR, RGB, and HSV [39]. High-frequency components correspond to detailed information of the human face, and low-frequency components correspond to contour information of the human face. Figure 14 shows a three-dimensional spectrogram of the face image after the two-dimensional Fourier transform. The selected frequency-domain broadening multiple was 3. We found that after a two-dimensional discrete Fourier transform of a non-occluded face in the RGB color space, the image retains concentrated low-frequency information to a certain extent, with a larger low-frequency radius and less high-frequency information. The low-frequency information of the spectrogram after the two-dimensional Fourier transform of the unoccluded face in the HSV color space was mainly concentrated in the horizontal and vertical directions, whereas the color of the area corresponding to the high-frequency information was relatively dim, indicating that there was less high-frequency information, and the HSV color space was unable to retain high-frequency information adequately. In the YCBCR color space, the non-occluded face spectrogram was able to retain more high-frequency detail information. Therefore, it the YCBCR color space may be considered more suitable for feature extraction from non-occluded faces [40].

5) ABLATION EXPERIMENT OF ESRGAN

We used ESRGAN to perform image preprocessing to improve the quality of cropped ROI images. As shown in Figure 9, by comparing the image recognition rate without ESRGAN pretreatment and after ESRGAN pretreatment of image recognition rate, it may be observed that compared with the original image direct input, images exhibited more abundant details of image features after ESRGAN pretreatment, which provides a good foundation for feature extraction by backbone networks. Richer features help classifiers to perform better, so the recognition rate of the proposed approach was improved over that of its original basis. Not only that, but other methods of facial recognition have improved to differing degrees.

C. ANALYSIS OF RESULTS

1) MODEL TRAINING TIME

We used 32 batches to train the proposed network model, and the RMFRD dataset was used to perform training.

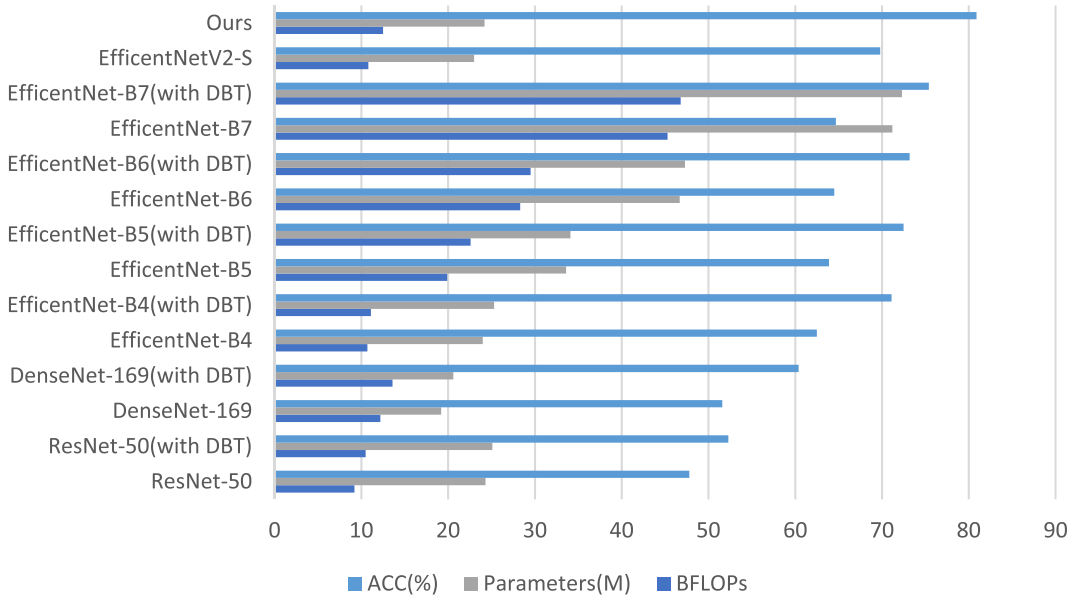


FIGURE 10. Accuracy, Flops, and parameters of data set RMFRD for models with and without DBT.

TABLE 4. Performance comparison of different method presented by Mean ± STD on the RMFRD dataset.

Model	AUC	Accuracy (%)	recall	Precision
colour texture [43]	0.640±0.021	0.503±0.016	0.625±0.015	0.458±0.225
AlexNet	0.658±0.047	0.5173±0.054	0.651±0.016	0.479±0.017
ResNet-50	0.613±0.033	0.531±0.084	0.623±0.019	0.437±0.025
lnception V3	0.672±0.041	0.562±0.013	0.645±0.005	0.518±0.003
LBP[44]	0.693±0.079	0.612±0.043	0.636±0.076	0.636±0.176
Gabor +LBP[45]	0.702±0.018	0.657±0.035	0.679±0.016	0.712±0.022
Luttrell et al. [19]	0.714±0.016	0.703±0.070	0.867±0.025	0.729±0.034
VGGFace	0.763±0.017	0.741±0.016	0.835±0.013	0.714±0.019
CosFace	0.781±0.006	0.763±0.007	0.852±0.009	0.830±0.016
Arcface	0.821±0.012	0.821±0.013	0.833±0.015	0.834±0.010
Our method	**0.901±0.013**	**0.987± 0.011**	**0.915±0.014**	**0.919±0.025**

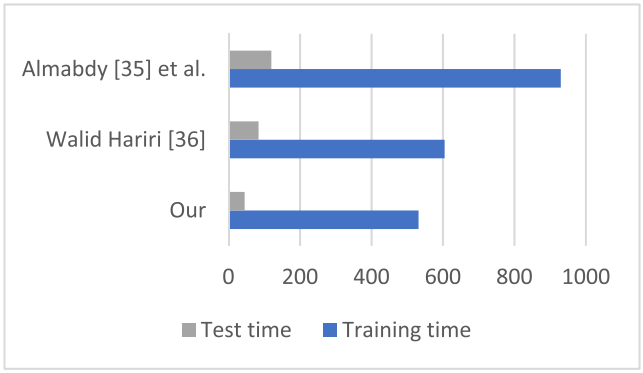


FIGURE 11. Training and testing time of different methods on RMFRD dataset.

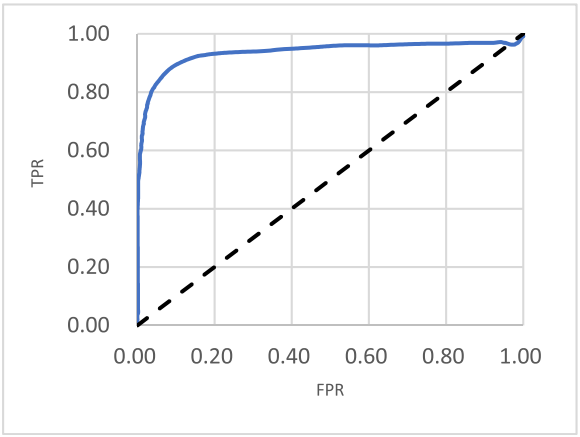


FIGURE 12. ROC curve of our method is tested on the RMFRD dataset.

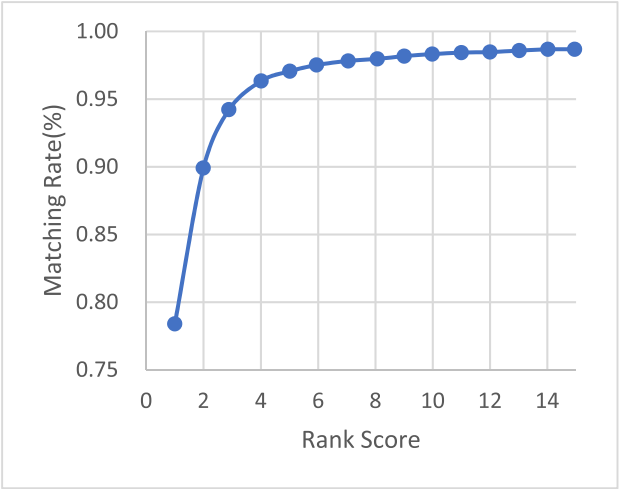


FIGURE 13. The CMC curve of our method is tested on the RMFRD dataset.

The training and testing times are shown in Figure 11; the proposed approach was faster than that of Almabdy and Elrefaei [41] and Hariri [42]. This may be attributed to the fact that they used traditional deep learning network models, specifically VGG16 and AlexNet, and the number of parameters in the training model was relatively large. In contrast, the DBT module has the advantage of adding detailed feature information without adding new parameters. FastICA also removes part of the redundant information, ensuring the effectiveness and independence of features extracted in the frequency domain. From Figure 11, we can also see that the number of parameters in our model is relatively small. Therefore, our method was faster in training and testing models than theirs.

2) MASKED FACE RECOGNITION RESULT DISPLAY

The accuracy of this method was tested in a real environment. As shown in Figure 15, we captured users' faces using a mobile phone with an ordinary imaging function and obtain a video 1 minute and 53 seconds long. Each frame of the image contained 1920 × 1080 pixels. We used the proposed method to perform facial recognition on a mask face. As may be observed from the figure, our model was able to recognize faces wearing masks well. This proves that the proposed method was able to accurately identify people even after their facial features are occluded by wearing masks. These results demonstrate the validity of the proposed method.

3) FINAL RESULT EVALUATION

In Table 4, we compare and evaluate the recall rate of current mainstream facial recognition algorithms and models from

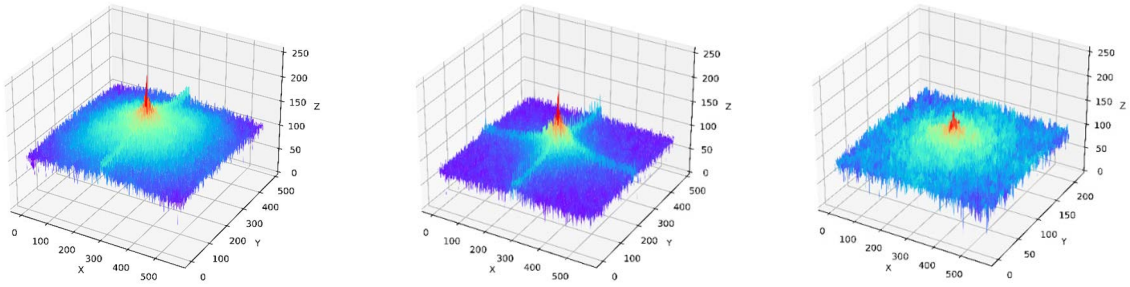


FIGURE 14. Unoccluded face image is a spectrogram of frequency domain widening in different color spaces. From left to right, there are spectrograms in RGB, HSV, and YCBCR color spaces.

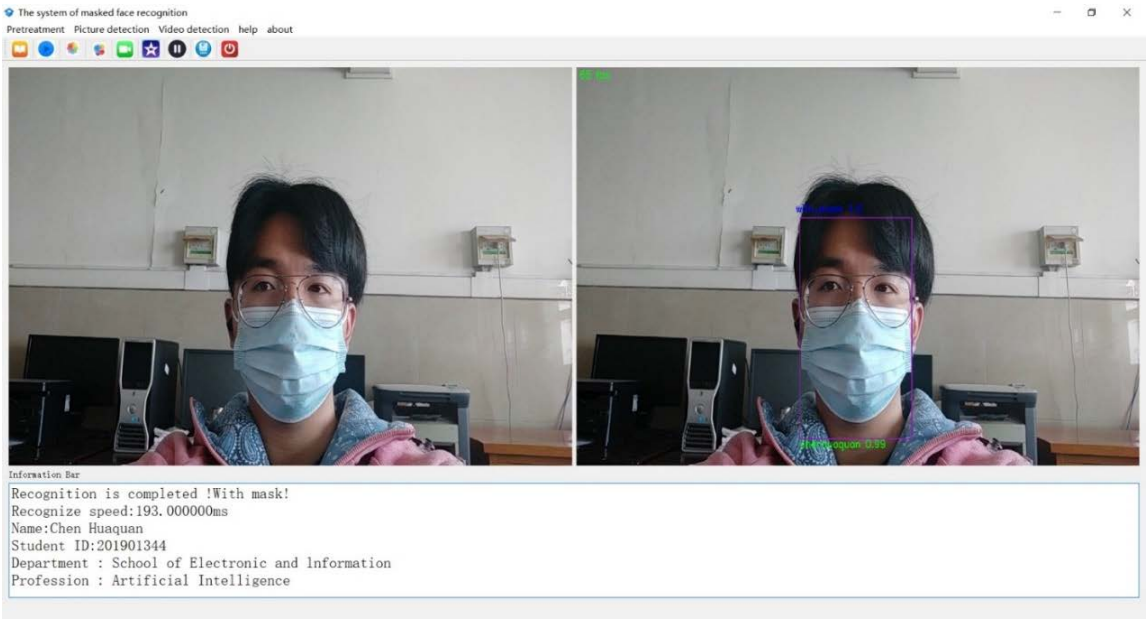


FIGURE 15. Our method is tested in a face recognition system that shows wearing a mask.

the perspective of the test samples. To evaluate the overall performance of the proposed approach, we performed tests on accuracy and recall. As shown in Table 4, the proposed face recognition model for masks is compared with state-of-the-art facial recognition methods. The Colour Texture [43], LBP [44], and Gabor +LBP [45] methods feature human face images from the perspective of texture features. Although VGGFace,CosFace and Arcface achieved good accuracy in face recognition without occlusion, they involve some limitations in terms of the accuracy in cases where few features can be extracted. However, they all share a common limitation in terms of feature extraction, and they all operate from a single direction to shade the face when extracting features. In contrast, our method uses the angle of the space and frequency domains to extract features, we use the same bilinear transformation in groups and between the different semantic group aggregation and generate effective characteristics of fine-grained information. Simultaneously, the FBB increases the frequency domain details of the non-occluded area of the face, which improves the recognition accuracy of our algorithm to a certain extent. Figure 12 shows the ROC curve of this approach. Figure 13 shows the CMC curve of our method, which is divided into 525 classes and 15 ranks. Among them, RANK1 equals 78.4%, RANK2 equals 89.9%, RANK3 equals 94.2%, and RANk15 equals 98.6%. About 412 people achieved the highest accuracy in the first test, i.e. hit the target. The second and third tests had 471 and 495 hits, respectively. As the number of tests increases, the accuracy of identification increases.

IV. CONCLUSION AND FUTURE WORK

Due to the novel coronavirus pandemic, people must wear face masks in public, but existing facial recognition systems do not operate properly on faces wearing masks. In this study, we have proposed three measures to solve this problem. We used DBT to improve efficentNetv2-S and make it more suitable for face recognition tasks with occlusions by masks. The results of experiments have been presented to show that the DBT module was able to improve recognition performance on RMFRD datasets. They also showed that

when the FBB and DBT module were used to extract features from the frequency and space domains in parallel, recognition accuracy was further improved for faces wearing masks. We also further optimized our method from the perspective of image preprocessing. The results show that using ESRGAN technology can improve the accuracy of the proposed method on RMFRD datasets. Finally, we tested the effectiveness of the proposed method in a real environment. However, our model has some limitations. Although the features extracted from the frequency domain and spatial domain by our model can be well classified in the final classifier, there are still some redundant features in these classified features, resulting in some errors.

We plan to continue to improve on the proposed methods in the future. For example, we will add a feature screening network before the classifier to further screen more effective facial features. In addition, EfficentNetv2-S was selected based on our existing equipment and the actual running speed of the system used to perform the calculations. In the future, we will compare the effects of other EfficentNetV2 sizes with the proposed approach. We also plan to continue to use the NAS grid search technology to find an optimal combination of MBConv and DBT-MBCONV or use a GAN to generate more spectral images containing facial feature information to further improve the generalization of this method so as to solve the problem of facial recognition on faces wearing masks in more complex environments. In addition, using generative model to generate face hidden features is also a direction worthy of further exploration. For example, a generated model is used to reconstruct faces obscured by masks to increase the number of training sets and reduce the hiding effect. Or use the generated model to mix the features of the face with the original features in order to check the matching accuracy of the proposed method.

AUTHOR CONTRIBUTIONS

Hua-Quan Chen conceived algorithms of the paper and write the manuscript, Kai Xie reviewed the paper, Mei-Ran Li designed experiments and conducted comparative experiments, Chang Wen collected data and conducted ablation study, and Jian-Biao He checked spelling and grammar and made suggestions.

REFERENCES

[1] X. Cheng, R. Zhang, J. Zhou, and W. Xu, "DeepTransport: Learning spatial-temporal dependency for traffic condition forecasting," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[2] R. Min, A. Hadid, and J.-L. Dugelay, "Efficient detection of occlusion prior to robust face recognition," *Sci. World J.*, vol. 2014, pp. 1–10, Jan. 2014.

[3] G. W. Wenshuo and Q. U. Haicheng, "Face recognition algorithm of occlusion location based on PCANet," *J. Frontiers Comput. Sci. Technol.*, vol. 13, no. 12, p. 2149, 2019.

[4] J. Xing, Z. Niu, J. Huang, W. Hu, X. Zhou, and S. Yan, "Towards robust and accurate multi-view and partially-occluded face alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 987–1001, Apr. 2018.

[5] Y.-A. Chen, W.-C. Chen, C.-P. Wei, and Y.-C.-F. Wang, "Occlusion-aware face inpainting via generative adversarial networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1202–1206.

[6] Q. Duan and L. Zhang, "Look more into occlusion: Realistic face frontalization and recognition with BoostGAN," *IEEE Trans. Netw. Learn. Syst.*, vol. 32, no. 1, pp. 214–228, Jan. 2021.

[7] S. Ge, C. Li, S. Zhao, and D. Zeng, "Occluded face recognition in the wild by identity-diversity inpainting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 10, pp. 3387–3397, Oct. 2020.

[8] X. Yuan and I. K. Park, "Face de-occlusion using 3D morphable model and generative adversarial network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10062–10071.

[9] I. Q. Mundial, M. S. U. Hassan, M. I. Tiwana, W. S. Qureshi, and E. Alanazi, "Towards facial recognition problem in COVID-19 pandemic," in *Proc. 4rd Int. Conf. Electr., Telecommun. Comput. Eng. (ELTICOM)*, Sep. 2020, pp. 210–214.

[10] R. Golwalkar and N. Mehendale, "Masked-face recognition using deep metric learning and FaceMaskNet-21," *Int. J. Speech Technol.*, vol. 2022, pp. 1–12, Feb. 2022.

[11] W. Wan and J. Chen, "Occlusion robust face recognition based on mask learning," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3795–3799.

[12] G. Wu, J. Tao, and X. Xu, "Occluded face recognition based on the deep learning," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2019, pp. 793–797.

[13] M. T. H. Fuad, A. A. Fime, D. Sikder, M. A. R. Iftee, J. Rabbi, M. S. Al-Rakhami, A. Gumaei, O. Sen, M. Fuad, and M. N. Islam, "Recent advances in deep learning techniques for face recognition," *IEEE Access*, vol. 9, pp. 99112–99142, 2021.

[14] X. Wang, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 63–79.

[15] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[16] H. Zheng, "Learning deep bilinear transformation for fine-grained image representation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–10.

[17] M. Tan and Q. Le, "EfficientNetV2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10096–10106.

[18] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

[19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[20] A. Aldhahab and W. B. Mikhael, "Face recognition employing DMWT followed by FastICA," *Circuits, Syst., Signal Process.*, vol. 37, no. 5, pp. 2045–2073, May 2018.

[21] B. Mandal, A. Okeukwu, and Y. Theis, "Masked face recognition using ResNet-50," 2021, *arXiv:2104.08997*.

[22] Y. Li, K. Guo, Y. Lu, and L. Liu, "Cropping and attention based approach for masked face recognition," *Appl. Intell.*, vol. 51, no. 5, pp. 3012–3025, 2021.

[23] Z. Li, Q. Zhang, X. Duan, and F. Zhao, "Face recognition based on regression analysis using frequency features," in *Proc. 4th IEEE Int. Conf. Inf. Sci. Technol.*, Apr. 2014, pp. 192–195.

[24] F. Z. Chelali, N. Cherabit, and A. Djeradi, "Face recognition system using skin detection in RGB and YCbCr color space," in *Proc. 2nd World Symp. Web Appl. Netw. (WSWAN)*, Mar. 2015, pp. 1–7.

[25] J. Advith, K. R. Varun, and K. Manikantan, "Novel digital image watermarking using DWT-DFT-SVD in YCbCr color space," in *Proc. Int. Conf. Emerg. Trends Eng., Technol. Sci. (ICETETS)*, Feb. 2016, pp. 1–6.

[26] E. Candès, L. Demanet, D. Donoho, and X. Ying, "Fast discrete curvelet transforms," *Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861–899, Sep. 2006.

[27] T. Qiu, C. Wen, K. Xie, F. Wen, G. Sheng, and X. Tang, "Efficient medical image enhancement based on CNN-FBB model," *IET Image Process.*, vol. 13, no. 10, pp. 1736–1744, Aug. 2019.

[28] L. Zhou, W. Liu, Z.-M. Lu, and T. Nie, "Face recognition based on curvelets and local binary pattern features via using local property preservation," *J. Syst. Softw.*, vol. 95, pp. 209–216, Sep. 2014.

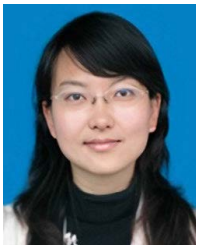
[29] J. Kong, M. Chen, M. Jiang, J. Sun, and J. Hou, "Face recognition based on CSGF(2D)^{2}PCANet," *IEEE Access*, vol. 6, pp. 45153–45165, 2018.

[30] C. Ferhaoui-Cherifi and M. Deriche, "On the performance of wavelet families in face recognition using a multilayer perceptron neural network classifier," in *Proc. 4th IEEE Int. Conf. Eng. Technol. Appl. Sci. (ICETAS)*, Nov. 2017, pp. 1–6.
[31] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, H. Chen, Y. Miao, Z. Huang, and J. Liang, "Masked face recognition dataset and application," 2020, *arXiv:2003.09093*.
[32] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Proc. Workshop Faces 'Real-Life' Images, Detection, Alignment, Recognit.*, 2008, pp. 1–15.
[33] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," 2014, *arXiv:1411.7923*.
[34] N. Boyko, O. Basystiuk, and N. Shakhovska, "Performance evaluation and comparison of software for face recognition, based on dlib and OpenCV library," in *Proc. IEEE 2nd Int. Conf. Data Stream Mining Process. (DSMP)*, Aug. 2018, pp. 478–482.
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[36] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
[37] F. Z. Zhou, G. C. Wan, and M. S. Tong, "Accurate image recognition in convolutional neural networks based on two-dimensional discrete Fourier transform," in *Proc. Photon. Electromagn. Res. Symp.-Spring (PIERS-Spring)*, Jun. 2019, pp. 1975–1978.
[38] H. Sun, E. Chen, and L. Qi, "Face recognition based on the feature fusion in fractional Fourier domain," in *Proc. 12th Int. Conf. Signal Process. (ICSP)*, Oct. 2014, pp. 1210–1214.
[39] S. Li, X. Ning, L. Yu, L. Zhang, X. Dong, Y. Shi, and W. He, "Multi-angle head pose classification when wearing the mask for face recognition under the COVID-19 coronavirus epidemic," in *Proc. Int. Conf. High Perform. Big Data Intell. Syst.*, May 2020, pp. 1–5.
[40] J. Luttrell, Z. Zhou, Y. Zhang, C. Zhang, P. Gong, B. Yang, and R. Li, "A deep transfer learning approach to fine-tuning facial recognition models," in *Proc. 13th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, May 2018, pp. 2671–2676.
[41] S. Almabdy and L. Elrefaei, "Deep convolutional neural network-based approaches for face recognition," *Appl. Sci.*, vol. 9, no. 20, p. 4397, Oct. 2019.
[42] W. Hariri, "Efficient masked face recognition method during the COVID-19 pandemic," *Signal, Image Video Process.*, vol. 16, no. 3, pp. 605–612, Apr. 2022.
[43] J. Y. Choi, K. N. Plataniotis, and Y. M. Ro, "Using colour local binary pattern features for face recognition," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 4541–4544.
[44] J. Zhang and X. Xiao, "Face recognition algorithm based on multi-layer weighted LBP," in *Proc. Int. Symp. Comput. Intell. Design (ISCID)*, Dec. 2015, pp. 196–199.
[45] P. V. Bankar and A. C. Pise, "Face recognition by using Gabor and LBP," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2015, pp. 45–48.

HUA-QUAN CHEN was born in Guangxi, China, in 2000. In 2020, he joined the National Electrical and Electronic Education Experimental Demonstration Center, Yangtze University, to study image processing and deep learning. He is committed to research in the laboratory image recognition and image classification and other scientific research projects. His research interests include artificial intelligence, machine learning, object detection, and face recognition.

KAI XIE received the M.S. degree in electronic engineering from the National University of Defense Technology, Changsha, China, in 2003, and the Ph.D. degree in pattern recognition and intelligent system from Shanghai Jiao Tong University, Shanghai, China, in 2006. He is currently a Professor with the School of Electronic Information, Yangtze University, Jingzhou, China. He works in the field of image processing and signal processing.

MEI-RAN LI was born in Shijiazhuang, in 2002. In 2020, she joined the National Demonstration Center for Experimental Electrical and Electronic Education to study image processing and deep learning. She has been doing research in medical image processing and artificial intelligence and has a good academic record.



CHANG WEN received the B.S. degree in computer science from the Naval University of Engineering, Wuhan, China, in 2002, and the M.S. degree in computer science from Yangtze University, Jingzhou, China, in 2008. She is currently an Assistant Professor with the School of Computer Science, Yangtze University. She works in the field of image processing and signal processing.

JIAN-BIAO HE received the B.S. and M.S. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 1986 and 1989, respectively. He is currently an Associate Professor with the School of Computer Science and Engineering, Central South University. His research interests include artificial intelligence, the Internet of Things, pattern recognition, mobile robots, and cloud computing.