AIMS Mathematics, 7(6): 10495–10512.
DOI: 10.3934/math.2022585
Received: 19 December 2021
Revised: 04 March 2022
Accepted: 11 March 2022
Published: 28 March 2022

http://www.aimspress.com/journal/Math

Research article

An intelligent recognition framework of access control system with anti-spoofing function

Dongzhihan Wang[1], Guijin Ma[1] and Xiaorui Liu[1,2,*]

[1] Automation School of Qingdao University, Qingdao 266071, China
[2] Shandong Key Laboratory of Industrial Control Technology, Qingdao 266071, China

* **Correspondence:** Email: liuxiaorui1979@163.com; Tel: +8618765979857.

Abstract: Under the background that Covid-19 is spreading across the world, the lifestyle of people has to confront a series of changes and challenges. This also presents new problems and requirements to automation facilities. For example, nowadays masks have almost become necessities for people in public places. However, most access control systems (ACS) cannot recognize people wearing masks and authenticate their identities to deal with increasingly serious epidemic pressure. Consequently, many public entries have turned to an attendant mode that brings low efficiency, infection potential, and high possibility of negligence. In this paper, a new security classification framework based on face recognition is proposed. This framework uses mask detection algorithm and face authentication algorithm with anti-spoofing function. In order to evaluate the performance of the framework, this paper employs the Chinese Academy of Science Institute of Automation-Face Anti-spoofing Datasets (CASIA-FASD) and Reply-Attack datasets as benchmarks. Performance evaluation indicates that the Half Total Error Rate (HTER) is 9.7%, the Equal Error Rate (EER) is 5.5%. The average process time of a single frame is 0.12 seconds. The results demonstrate that this framework has a high anti-spoofing capability and can be employed on the embedded system to complete the mask detection and face authentication task in real-time.

Keywords: face recognition; mask detection; spoofing detection; face authentication
Mathematics Subject Classification: 68U10

1. Introduction

With Covid-19 spreading across the world, epidemic prevention has become an essential part of public life and more and more people begin to be concerned about their health of themselves and turn to adapt to a new lifestyle. Covid-19 is a kind of respiratory infectious disease that spreads mainly through droplets and close contact. To protect people in public places from infection, one practical suggestion for people is to wear masks as possible, keep social distance, and so on. However, most public facilities cannot be directly compatible with these new normal actions under epidemic. Taking the entry scene for example, most public entries are equipped with ACS to effectively prevent idle personnel, recognize people and check their identities. In order to prevent entry procedures causing the spread of the epidemic, it is not suggested people to taking off masks or sojourn on the ACS. Therefore, it needs to enable ACS with high-level biometric recognition capability and passage efficiency. As one popular direction of computer vision research, biometric recognition is widely used in security systems due to its safety and convenience. Nowadays common biometric recognition methods include face recognition, fingerprint recognition, iris recognition, voice recognition, etc. Among these methods, face recognition is supposed to be the most convenient compared with others, especially its non-contact characteristics match the requirement of Covid-19 epidemic prevention. In most public entry conditions, a face recognition module is implemented based on a monocular camera. In practice, this design has to confront several challenges, such as mask-wearing reorganization and Face Anti-Spoofing Detection (FASD). The article will discuss the solutions we think of for the above two problems.

In terms of mask-wearing recognition, this paper focuses on the fusion between mask detection and ACS based on face recognition in order to remind users to wear masks correctly. Qin et al. [1] proposed a facemask-wearing condition identification method, which is divided into three conditions for face mask detection and includes a face with masks correctly, faces with masks incorrectly and faces without masks. This method achieves relatively satisfying results in experiments but lacks robustness under complex background. Ejaz et al. [2] proposed another face masks detection method based on Principal Component Analysis (PCA). This method meets the application requirement on the embedded devices, but the accuracy of face-mask detection is not good. In order to deal with the lack of datasets, Jiang et al. [3] designed one kind of mask detector which uses transfer learning to enhance networks robustness which increases the capability of image feature extraction in the Backone. Through using feature pyramid networks (FPN) [4] and Convolutional Block Attention Module (CBAM) [5], this method increases the accuracy of classification results and enables the frame compatible with embedded systems too.

Completing face recognition is divided into two tasks: Face identification task and face verification task. The face identification task compares one face with face datasets from which the most possible identity will be found. The face verification task is a binary classifier, it makes use of a similarity metric algorithm to get confidence whether this identity indeed matches the tested face. Thanks to the recent developments of Convolutional Neural Networks (CNNs), face recognition has achieved great progress, but there are many challenges, such as illumination, pixel resolution, and so on. There are two ideas for face recognition in the era of deep learning: Metric learning [6] and margin-based classification [9–12]. Metric learning is also called distance metric learning, which classifies input images by calculating the similarity between two input images. However, the model training relies on so large data that the efficiency of the fitting is very slow. Margin-based classification is

margin limited to feature layer by modifying softmax formula indirectly. This method makes the network obtain a feature that is more discriminative than the feature obtained after training. But the model has the instrict of overfitting because of harsh constraints. Yi Sun et al. [7] proposed a DeepID model that achieves 99.15% face verification accuracy on the LFW dataset by designing deep CNNs and using both face identification and verification signals as supervision. The advantage of DeepID is its high accuracy, but its training process is also prone to be overfitted. Dung Nguyen et al. [8] proposed a framework called FaceNet that directly learns a mapping through transferring face images to a compact Euclidean distance which directly corresponds to face similarity. This method achieves an accuracy of 99.63% on the LFW and accuracy of 95.12% on YouTube FacesDB respectively. However, there are many facial features lost in the process, which makes it easy to focus on the local feature rather than the global feature. Besides, another algorithm called ArcFace Jiankang was proposed by Deng et al [9]. It has a geometric interpretation to obtain highly discriminative features for face recognition. The verification accuracy of this method is 99.83% on the LFW dataset and 98.02% on the YTF dataset. ArcFace 9has high performance, but its weight model is so large that it is much more difficult to apply in transfer learning. MV-softmax [10,11] integrates the feature margin and feature mining into one unified loss function, it implements the discriminative learning by adaptively coding the feature vector of false classification. The main disadvantage of MV-softmax is to stress hard samples or half-hard samples, this may cause a negative influence on the model coverage. CurricularFace [12] adopts one kind of adaptive strategy to train loss function using margin samples and hard samples mining, which has great value in practice.

Generally speaking, although face recognition technology has been widely studied, there still exists a balance among the network size, computation cost and generalization. In addition, new face recognition also needs to complete tasks when people wear masks as possible, this is also the focus in the research field.

Compared with mask-wearing recognition, FASD is a long-term and widely-discussed problem because nowadays the biometric information of the face is easier to capture and be utilized for spoofing. Usually, Face spoofing attacks method includes photo attacks [13,14], video attacks [15,16] driven by reinforcement learning model, and 3D mask attacks [17–19] within low cost, hyper-real 3D color masks. Compared with widely studied 2D face presentation attacks, 3D face spoofing attacks are more challenging because present 3D face recognition systems are difficult to balance the effectiveness and cost. A typical 3D anti-spoofing system based on holographic imaging has little availability for most public entry scenes due to its high expense. Consequently, conventional face recognition systems are vulnerable to the above proofing attacks, most widely-applied face anti-spoofing methods up to now are still Human-computer interaction (HCI) methods that verify the user through some behavior interaction, such as winking, head moving, and lip language. Therefore, it is necessary to endow ACS with the capability of verifying the authenticity of facial information, and reject illegal users who present an imitation or fake face of the enrolled user.

The core task of the FASD is to find the intrinsic difference between the sampled information and fake spoofing information. With the development of the reinforcement learning technique, it has not been difficult to produce dynamic face movement completely driven by data, which presents a challenge to the anti-spoofing method based on human-machine interaction. Up to now, the main attention is increasingly drawn on the research involved in color texture analyzing [13,14,20,21] and adaptive recognition based on deep learning [22,23], etc. Määttä et al. [19] utilized multi-scale LBP to analyze the texture of face images. Bouklenafe et al. [13,14] proposed a face anti-spoofing algorithm

based on color texture. This algorithm considered the three-color space, RGB, YCbCr and HSV, which improves the model reliability in various conditions. After that, Bouklenafe et al. [13,14] started to focus on the luminance and chrominance channels and utilized the fusion mechanism of color and texture. Although FASD based on texture has drawn much attention, its performance is easy to be affected by illumination and image resolution. When the above factor decreases in quality, this method may be difficult to maintain self-stability. Another category of FASD is to monitor face movement. Usually, a genuine face has lots of movement characteristics, such as blinking, mouth opening and head-turning, while fake faces are difficult to formulate similar movements. Pan et al. [24] proposed a real-time liveness detection approach by recognizing spontaneous eyeblinks to prevent photos attack in face recognition. Zadeh et al. [25] also proposed a method utilizing eye and mouth movements to conduct liveness detection. Despite these ways tend to have high accuracy, it is still necessary to highly interact with users. On the other hand, the time cost of this detection is relatively long and not friendly to users. In recent years, there has appeared simulated dynamic faces driven by the reinforcement learning model [26], which presents challenges to FASD research. It is also found that the CNNs have great potential for learning faced with various spoofing ways and improved the robustness of anti-spoofing algorithms. Atoum et al. [22] proposed a two-stream CNNs-based approach for face anti-spoofing. Liu et al. [23] assumed liveness detection as a binary classification problem to be directly learned by Deep Neural Networks (DNNs). Despite CNNs [27,28] having potential for face liveness detection, it has to face the problem of overfitting and needs further studies.

Based on the above discussion, it is not difficult to find that the liveness detection research has competed with a variety of spoofing approaches all the time. In this paper, we focus on these issues and present an intelligent access control framework with two improvements:

1) A novel authentication system fusing the mask detection algorithm and face authentication algorithm with face anti-spoofing algorithm is proposed. The mask detection algorithm employs the mask detector called RetinaFaceMask, and uses MobileNetV3 as the feature extractor; the face authentication algorithm makes use of Tripletloss for training the neural network model;

2) In order to realize the anti-spoofing function, this paper makes use of a lightweight neural network and uniform Local Binary Patterns (LBP) method to extract the color texture of the image frame. The accuracy and training efficiency of the above model is tested on the CASIA-FASD dataset and Replay-Attack dataset.

This paper is organized as follows. Sections II and III discuss the design of the network framework in detail, the selection of benchmark datasets, and experiment setups respectively. Then the analysis and discussion of experimental results are presented in Section IV.

2. Design of intelligent access control framework

The proposed framework of the intelligent access control framework is composed of three parts: Mask detection module, face detection and recognition module and face anti-spoofing module (as shown in Figure 1). During the operation process, the passenger is sampled using videos and flagged the mask-wearing status by the mask detection module. Then it makes use of a multi-task convolutional neural network (MTCNN) [29] to process captured video frames and align them within the user's face. Through MTCNN the detected face is centrally located in alignment pictures with a size of 160*160 pixel). The alignment pictures are then filled into the face anti-spoofing module in which the liveness detection algorithm can judge its authenticity. If one face is judged as real, it will enter the face

recognition module for identity matching. In the face recognition module, if the real face matches one in the database, this face and its identity will be flagged by this system. Otherwise, the face recognition will end and present an alarm message.

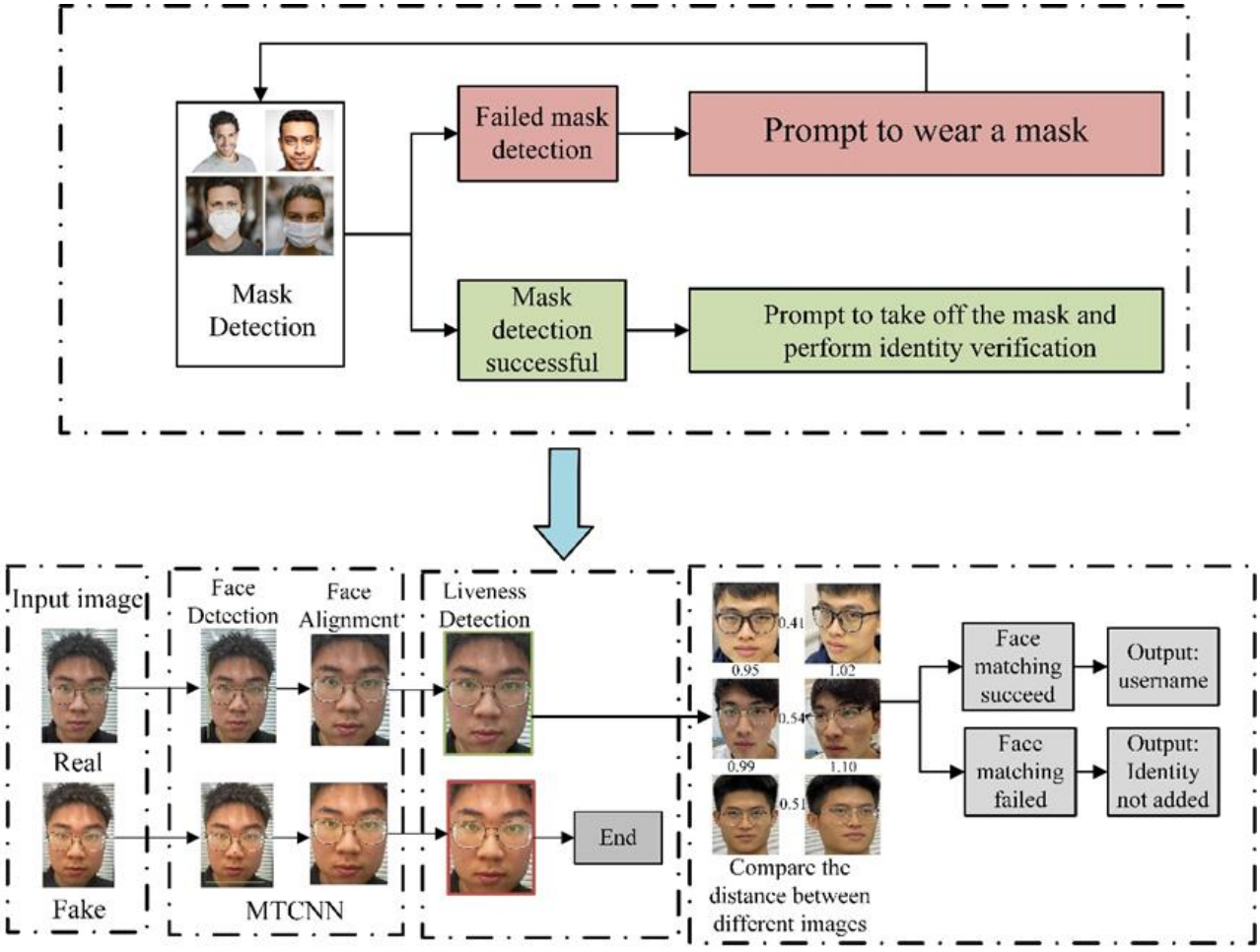

Figure 1. Schematic diagram of the intelligent access control framework.

2.1. Mask and face detection

This paper designs mask and face detection module, which includes mask detection unit and face detection unit. The mask detection unit is designed based on a kind of face-mask detector called RetinaFaceMask (as shown in Figure 2). In order to apply to embedded applications, the feature extractor is replaced with MobileNetV3 which is featured with its low parameter quantity and lightweight architecture. On the other hand, the high-level semantic information is extracted by FPN from the feature maps of different sizes while fusing this information into previous layers' feature maps by adding operation with a coefficient. In Figure 2, the head part of mask detection architecture represents classifiers, predictors, estimators which complete inference function and indicate the mask-wearing status. The test image input into networks model in the testing, which outputs two results are face confidence and mask confidence by removing proposal which is less than the set threshold. The remaining results are filtered using Non-Maximum Suppression (NMS). The prediction results of faces

and masks are retained, whose weighted logistic loss is defined as (1). In the Eq (1), d_i denoted a detection, $y_i \in \{-1,1\}$ indicate whether or not d_i was successfully matched to an object, and let f denote the scoring function (RetinaFaceMask) that jointly scores all detections on an image $f([d_i]_{i=1}^n) = [s_i]_{i=1}^n$. For a variety potential detection results, here loss per detections is coupled to the other detections through the marching produces y_i. By this way, the prediction results of faces and masks are retained.

$$L(s_i, y_i) = \sum_{i=1}^{N} \omega_{y_i} \cdot \log(1 + \exp(-s_i \cdot y_i)). \tag{1}$$

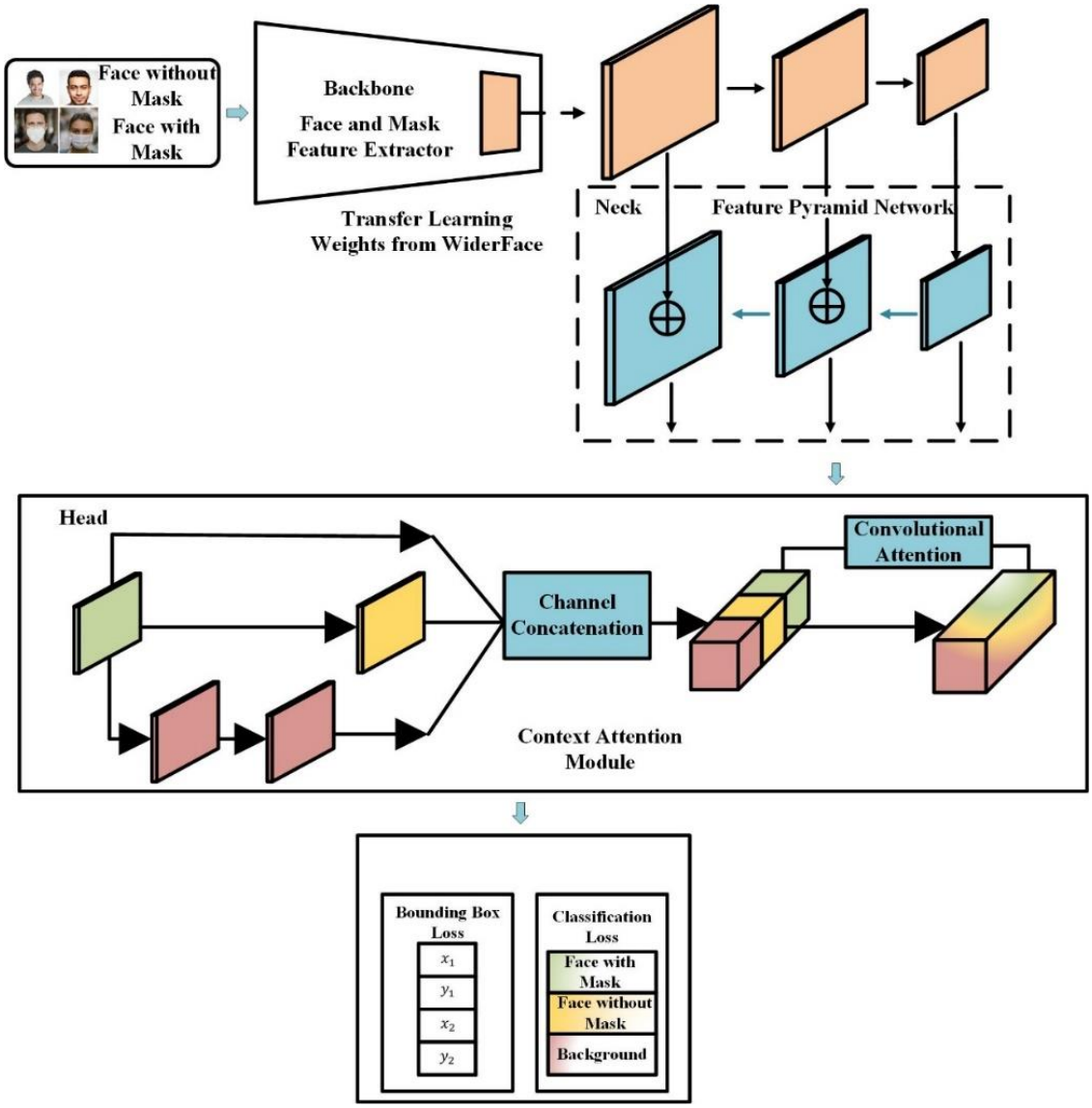

Figure 2. Schematic diagram of the parts in mask detection architecture.

Due to the limited data set of face masks, it is difficult for the network to be trained through the training set to get good results so that we train a more robust network model through transfer learning. In RetinaFaceMask, only the parameters of the backbone and neck are transferred from Wider Face, a dataset consisting of 32,203 images and 393,703 annotated faces. RetinaFaceMask proposed a new

contextual attention module as Heads in order to improve the prediction performance of face masks. The context attention module has three branches, each branch has a 3×3, two 3×3, and three 3×3 convolution kernels, and using different numbers of 3×3 convolution kernels to obtain different receptive fields from the same feature map. The original context module does not consider the face or mask, so it needs to feature an abstraction component put it into retraining. Here the abstraction component is designed to be one simplified convolutional block attention module (CBAM). Given an intermediate feature map from the original RetinaFaceMask architecture, the CBAM can sequentially infer a 1D channel attention map and a 2D spatial attention map. Although this structure is suitable to save the memory cost, the spatial attention module in the CBAM still can occupy a variety of computation costs [30,31]. Therefore, in this paper, we replace the attention module of CBAM with on Maxpool-Convolution-AveragePool component. This simplification is equivalent to cascading the shared spatial perception of the feature abstraction. After it, the RetinaFaceMask is prompted to pay attention to the face and mask features.

Another module of the mask and face detection module is the face detection unit**,** which uses the MTCNN model. This model is built as a three-layer neural network, which is P-Net, R-Net, O-Net. In the operation, the obtained images are processed by these three networks to detect faces and face landmarks. When input images come, they will be resized continuously to produce the image pyramid. The pyramid is filled into P-Net to obtain lots of candidates that are filtered by R-Net more precisely afterward. Finally, the filtered data is filled into O-Net to calculate the landmarks' position of potential faces. This model realizes face detection and alignment functions.

2.2. Face anti-spoofing

Boulkenafet et al. [14] propose that face anti-spoofing could be distinguished by analyzing color texture. The color is analyzed by utilizing uniform LBP in the luminance and chrominance channels. Uniform LBP extracts the histogram from a single image band and concatenates these histograms into the final descriptor which be calculated by Support Vector Machine (SVM) to distinguish real face and fake face.

LBP is a feature descriptor of local texture for gray-scale images proposed by Ojala et al. [32]. The traditional extraction approach of the LBP feature descriptor makes the original image be divided into blocks of n*n and each block is divided into cells of a*a. Then, the feature of each cell is extracted by LBP. Considering the center pixel value as a threshold to compare with 8 pixels from the yield of threshold around. The LBP descriptor is defined as Eq (2):

$$LBP(x_c, y_c) = \sum_{p=0}^{7} 2^p s(i_p - i_c). \tag{2}$$

(x_c,y_c) is the center point in the 3*3 area, which pixel value is i_c. i_p represents the value of other pixels in the area. p is the number of pixel points. $s(x)$ is a symbolic function which is defined as Eq (3):

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}. \tag{3}$$

$s(x) = 1$, if $x \geq 0$, otherwise $s(x) = 0$. The values of each pixel point be calculated by exploiting LBP descriptor. The result is connected to a binary value and converted to a decimal value which is

the value of the LBP descriptor. Generating situations are 256 by this approach, but lots of situations are very complex so it is not good for the extraction of facial features. Thereby, Ojala, et al. [33] proposed Uniform LBP to optimize the traditional LBP descriptor for implementing dimensionality reduction through the transition approach. The uniform LBP descriptor is defined as Eq (4):

$$LBP_{P,R}^{(i)}(x,y) = \begin{cases} \sum_{n=0}^{P-1} \delta\left(r_n^{(i)} - r_c^{(i)}\right) \times 2^n, U^{(i)} \le 2 \\ P(P-1), \qquad\qquad U^{(i)} > 2 \end{cases}. \tag{4}$$

(x,y) is the center point of the neighborhood, $r_n(n = 0, \cdots, P-1)$ is the pixel value of surrounding neighborhood, r_c is the pixel value of center point, where R is the radius of the surrounding neighborhood, P is the number of points in the surrounding neighborhood, where $U^{(i)}$ is defined as Eq (5):

$$U^{(i)} = \left|\delta\left(r_{P-1}^{(i)} - r_c^{(i)}\right) - \delta\left(r_0^{(i)} - r_c^{(i)}\right)\right| + \sum_{n=1}^{P} \left|\delta\left(r_n^{(i)} - r_c^{(i)}\right) - \delta\left(r_{n-1}^{(i)} - r_c^{(i)}\right)\right| \tag{5}$$

where, $\delta(x)$ is the symbolic function which is defined as Eq (6):

$$\delta(x) = \begin{cases} 1, & x \ge 0 \\ 0, & x < 0 \end{cases}. \tag{6}$$

$\delta(x) = 1$, if $x \ge 0$, otherwise $\delta(x) = 0$.

Uniform LBP extracts the histogram from a single image band, and then concatenates these histograms into the final descriptor which be calculated by SVM [34] to distinguish real face and fake face. The whole process is shown as Figure 3. The Figure 3 contains two options to complete liveness detection task. The option 1 is face anti-spoofing based on color texture analysis. It extracts the faces in the dataset and crops them to the unified format of 224 * 224 pixels. These unified face frames are converted into the color spaces of HSV and YCbCr. Then the LBP descriptor extracts the feature histogram from color spaces, and connects them in series. In the end, serialized feature histogram is input in SVM to classify as real or fake face.

Option 2 in Figure 3 completes face anti-spoofing based on a lightweight neural network. The advantages of the lightweight neural network [35–39] are high accuracy, smaller parameter, detection faster and can be well applied to intelligent access control systems. Face liveness detection is defined as a binary classification question. Using the face data set of size 224×224 to train a lightweight neural network, and change the linear activation function of the last layer of the network structure to the Tanh activation function. Because the weights of multi-type image classification and living detection two classifications are different, the two-classification model suitable for living detection is used for pre-training. In terms of hyperparameter selection, Epoch is 150, Batch Size is 64, the cross-entropy loss function is used, the optimizer selects random gradient descent, the weight attenuation is 4×10^{-5}, and the Momentum is 0.9. According to the above conditions, the lightweight neural network is trained.

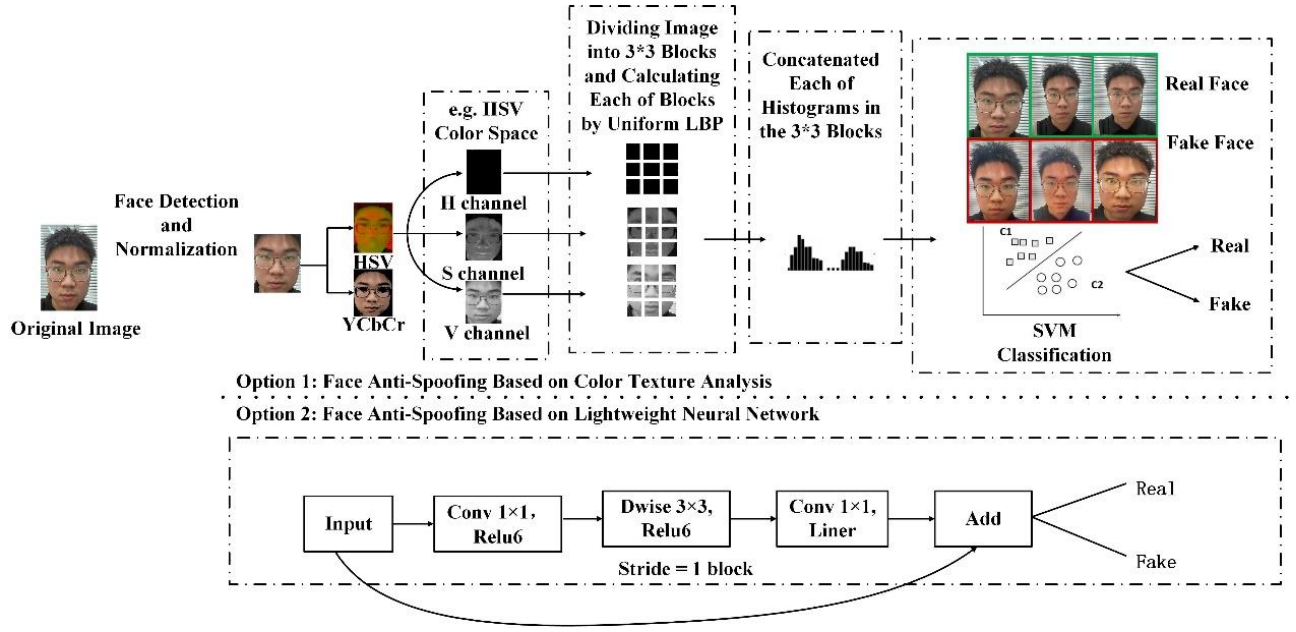

Figure 3. Schematic diagram of face anti-spoofing detection using SVM.

2.3. Identity verification through face recognition

The function of face detection and verification is a hot spot in the field of security. In this paper, we propose a neural network called Face-net. This neural network encodes face images into digital vectors (or called face-net vectors). In order to calculate the confidence between the face image and preset identity, this paper uses Euclidean distance to compare two face-net vectors and determine whether these two vectors belong to the same person. As shown in Figure 4(b) represents distances between images. The distance between the same person's images is basically below one threshold which is demarcated a visa experiment.

As shown in the part (a) of Figure 4 In this paper, Face-net model is designed to realize the face verification and recognition, architecture of Face-net model is shown in Figure 4(a), batch refers to the input face image sample which is found and then is cropped to a fixed size of 160×160 pixel. Deep architecture refers to a deep learning architecture, the deep architecture of Face-net uses the GoogleNet network, which combines the traditional convolutional neural network with Inception structure. This configuration can get the balance between the size and over-fitting characteristics of the deep neural network. In practice, the main disadvantage of this method is the over-high dimensions of output features. So, optimization is conducted by replacing the last layer of the softmax classifier with one layer of L2 feature normalization. Through the normalization of the L2 layer, the revised module will output a 128-dimensional feature vector which is then optimized with triple loss. The triplet loss is one kind of loss function in Figure 4(c). The input of the triplet loss function is three groups of images. These three image groups are called Achor(A), Negative(N) and Positive(P) respectively. The A and P are captured from the same person with different imaging angles. The N is captured by other people. The similarity between image groups is quality by Euclidean distance. When training face-net, the Euclidean distance between A and P will gradually decrease, and the distance between A and N will gradually increase. Through this mechanism, the face-net learns the separability between features, and builds connections between every single image from a specific person. This is the basis for identity

verification. The definition of triplet loss is shown as Eq (7):

$$\left\|x_i^A - x_i^P\right\|_2^2 + \alpha < \left\|x_i^A - x_i^N\right\|_2^2, \forall\left(x_i^A, x_i^P, x_i^N\right) \in T. \tag{7}$$

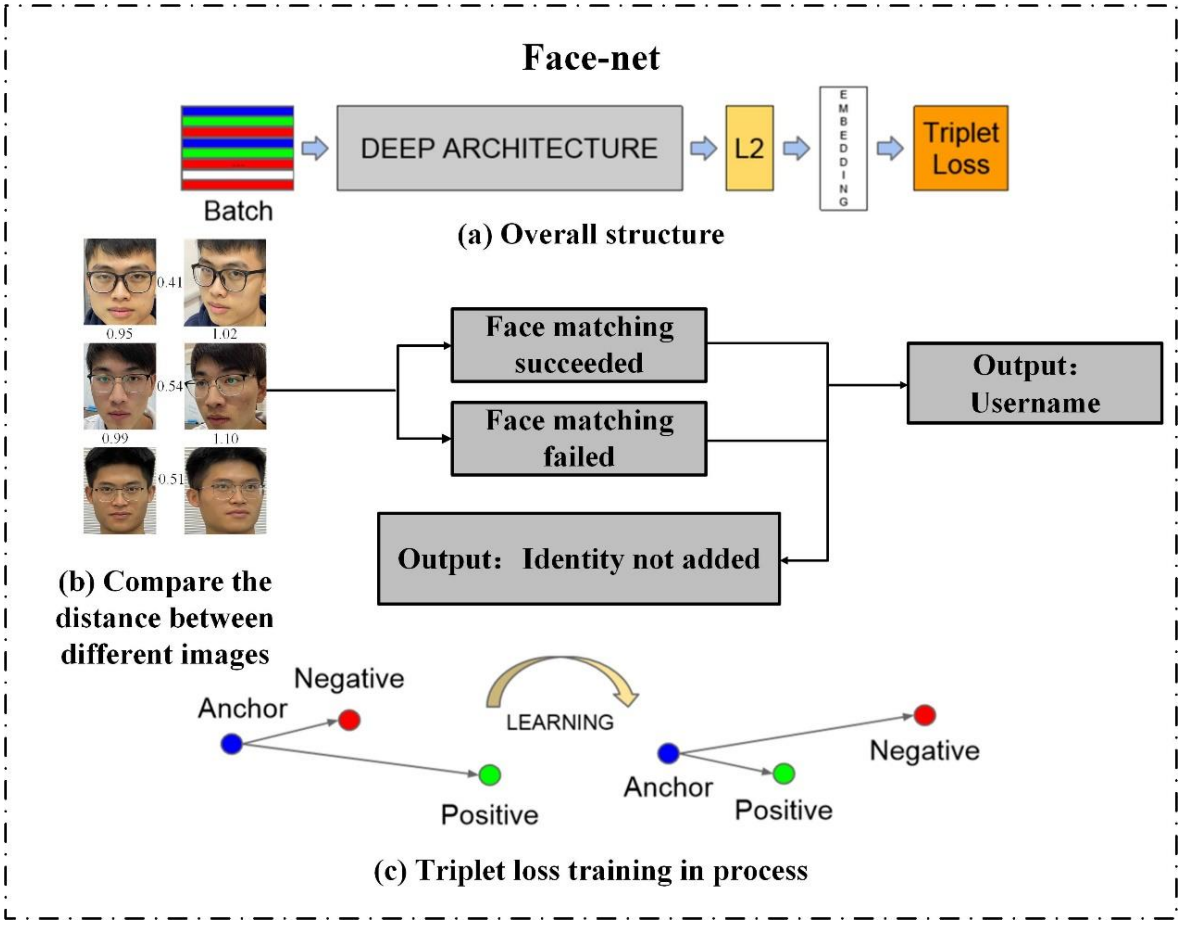

Figure 4. Schematic diagram of how Facenet identifies a person.

In the Eq (6), x_i^A (Anchor) represents the image of A, x_i^P (Positive) represents the different image of A, x_i^N (Negative) represents the image of B. α represents constant, and T represents the set of triplets in the training set. The loss function minimization L is defined as Eq (8).

$$L = \sum_i^M \left[\left\|f\left(x_i^A\right) - f\left(x_i^P\right)\right\|_2^2 - \left\|f\left(x_i^A\right) - f\left(x_i^N\right)\right\|_2^2 + \alpha\right]_+. \tag{8}$$

In the Eq (7), $f(x) \in \mathbb{R}^d$ represents a face image x that is embedded in d-dimensional European space. Among them, the second norm on the left represents the intraclass distance, the second norm on the right represents the distance between classes, and α is a constant. The "+" at bottom right represents that when the value in the brackets is greater than zero, the loss is taken, and when it is less than zero, the loss is zero. M is the base. The optimization process is to use the gradient descent method to make the loss function continuously decrease which means the distance within the class continuously decreasing, and the distance between classes continuously increased.

To draw a conclusion, aiming at the possibility of infection and negligence in human supervision in public places, at the meantime, in order to meet the needs of embedded device platform, this paper

proposes an intelligent access control framework composed of the mask detection algorithm, MTCNN algorithm, face anti-spoofing based on color texture analysis and Face-net identity verification. This framework can meet the requirements of embedded device applications and effectively solve the security problem of the public place access control system under the background of the spreading coronavirus.

3. Experimental setup and results analysis

In this section, we present the experimental evaluation of the proposed scheme for the intelligent access control system. The two data sets used in our experimental evaluation are introduced in detail. Experiments are performed on different anti-spoofing datasets. Details are given in the subsequent sections.

3.1. Benchmark datasets selection

In order to evaluate proposed system objectively, we have chosen widely accepted CASIA-FASD [40] and Replay Attack [41] Datasets for experiments, both of which have well-described evaluation protocols. The CASIA-FASD datasets consist of 600 videos including 50 objects, each of the objects has 12 videos including 3 videos of legal access and 9 videos of illegal access. There are 3 different types of videos in the datasets which is low, normal, high, respectively. Illegal access videos have 3 different types of attack methods including warped photo attacks, cut photo attacks, and video attacks. The CASIA-FASD datasets are shown as Figure 5. The Reply-Attack datasets include 1200 videos which have 50 objects. Each of the objects has 24 videos which include 4 videos of legal access and 20 videos of illegal access, each of the videos has about 10 seconds in length. All videos are classified to 3 different backgrounds and 2 different illuminations. Illegal access videos have 3 different types of attack methods including print attack, digital images attack, videos attack. The datasets have 3 parts of datasets including train datasets, development datasets, text datasets.

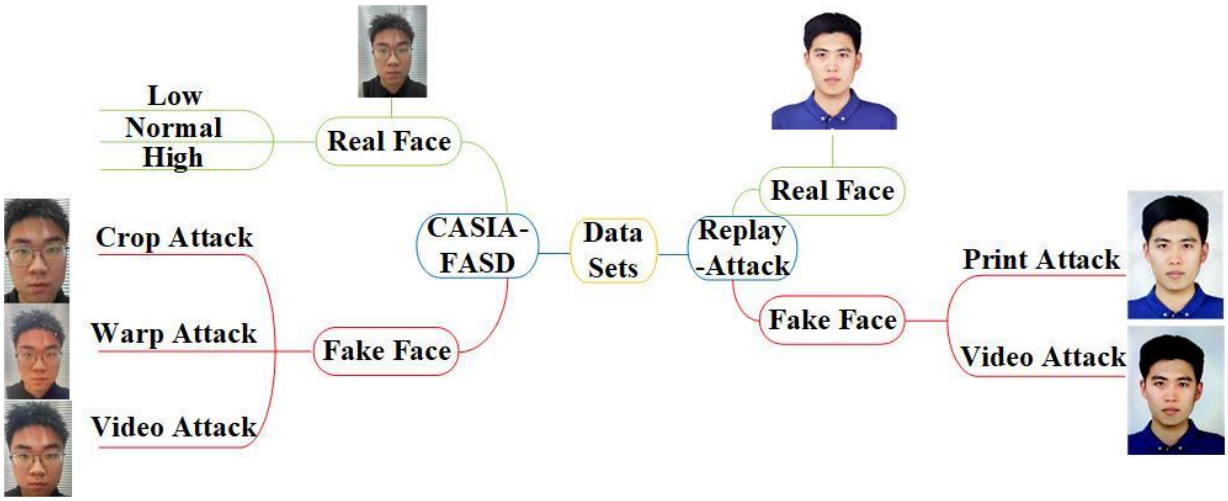

Figure 5. The examples of the real face and fake face in the public CASIA-FASD and Reply-Attack.

3.2. Experimental setup

In order to obtain a face recognition security system with a low calculation amount and a high recognition rate, the live detection is more user-friendly. The experiment verifies the performance of the algorithm by comparing the size of the model parameters, the recognition accuracy, the time required for the system to process a frame of pictures through the entire process, and the CPU occupancy.

This section focuses on the experiments to verify the anti-spoofing performance of the algorithm, using the CASIA-FASD and Replay-Attack datasets to conduct experiments. The datasets are shown in Table 1.

Table 1. Cases of CASIA-FASD and Replay-Attack.

Dataset	Training Set	Testing Set	Total
CASIA-FASD	7200	10800	18000
Replay-Attack	10800	14400	25200

a) The videos contained in the CASIA-FASD and Replay-Attack datasets are read and saved by OpenCV for 30 frames. After each frame is obtained, the image will be aligned and cropped by MTCNN and normalized to a 224×224 RGB picture.

b) This chapter uses datasets for training and testing separately. In the quantitative analysis, the experiment uses accuracy rate, EER, HTER, and receiver operating characteristic (ROC) curve as the evaluation indicators for evaluating the algorithm. When quantitatively analyzing the real-time performance of the face security system, we verify the performance of the algorithm through the time required for the system to process a frame of pictures and CPU occupancy rate.

c) All experiments have been conducted on a computer with PyCharm® installed in Windows 10. The processor model is Intel Core i5-8250U@ 1.60 GHz, and the memory is 8 GB RAM. The algorithm in this chapter is based on PyTorch 1.5 and TensorFlow 1.12 deep-learning framework, using Python 3.7 language implementation.

3.3. Results analysis

We did experiments on CASIA and Replay-Attack datasets to compare with the EER, HTER, ROC curves of different algorithms, the time required for the system to process a frame of pictures, and the experimental results of the CPU occupancy rate. The advantages and disadvantages of the algorithm are analyzed.

The CASIA dataset has rich attack methods and complex datasets. The Replay-Attack dataset has a total of 25,200 pictures, the attack method is video attack, the control variable is light, and the data set attack method is relatively simple. From Figures 6 and 7, we can see that under the Replay-Attack dataset with a single attack method, the algorithm can perform better classification, while under the complex CASIA data set, the LBP algorithm can still maintain good classification performance.

The lightweight neural network is limited in the size of the dataset. On the one hand, it is subject to fewer model parameters. Even if it is pre-trained under the live detection model, the network convergence process is still slow; on the other hand, under the multi-image classification It can achieve

good accuracy, but live detection pays more attention to the image texture, spectral information, motion information, depth information and other nuances between the real face and the fraudulent face. This is a lightweight neural network for live detection under higher requirements.

It can be seen from Figures 6 and 7, and Table 2 that the size of the model parameters also affects the accuracy of live detection to a certain extent. Figures 8 and 9 verify that the LBP algorithm takes the least time when using SVM for facial feature classification, and the CPU usage is low. Because SVM has a better classification effect for small sample data and has a faster calculation speed.

The overall results of the experiments show that the system fused with the LBP algorithm and the FaceNet algorithm can effectively identify the attacks of forged faces on the face recognition security system while the real-time detection efficiency is high, and the lightweight neural network exists in the real-time live detection feasibility.

Table 2. Performance comparison of each algorithm under the CASIA-FASD and Replay-Attack datasets.

Algorithm	CASIA		Replay-Attack		Parameter
	HTER (%)	EER (%)	HTER (%)	EER (%)	
MobileNetV2	16.7	9.4	12.6	3.6	3.4MB
MobileNetV3	19.3	14.4	10.9	6.5	3.87MB
ShuffleNetV2	21.9	14.9	21.8	6.33	1.4MB
SqueezeNet	28.3	19.1	24.7	13.7	736.45KB
Xception	20.5	9.4	20.5	4.2	20.81MB
LBP	9.7	5.5	10.8	5.6	-

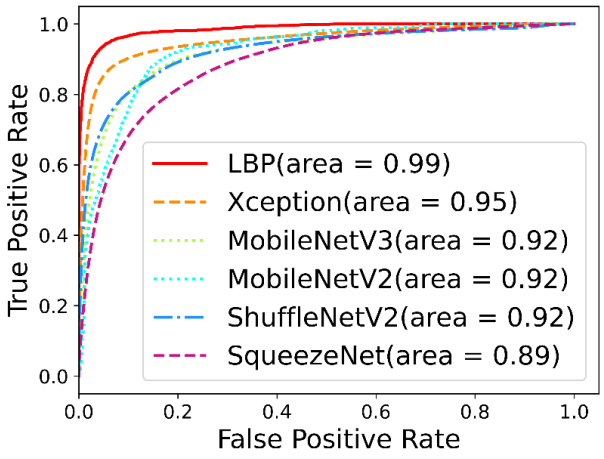

Figure 6. ROC curve of living body detection algorithm under CASIA dataset.

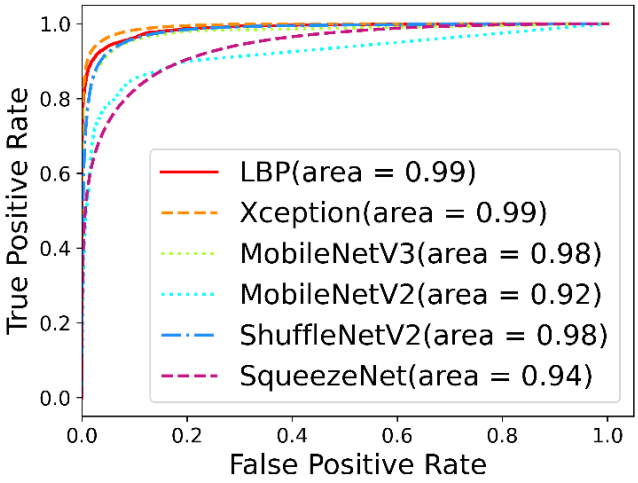

Figure 7. ROC curve of the living body detection algorithm in the Replay-Attack dataset.

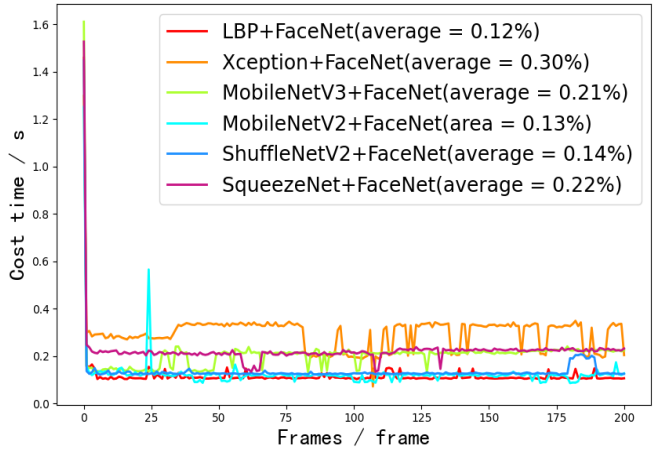

Figure 8. Time spent under different algorithms.

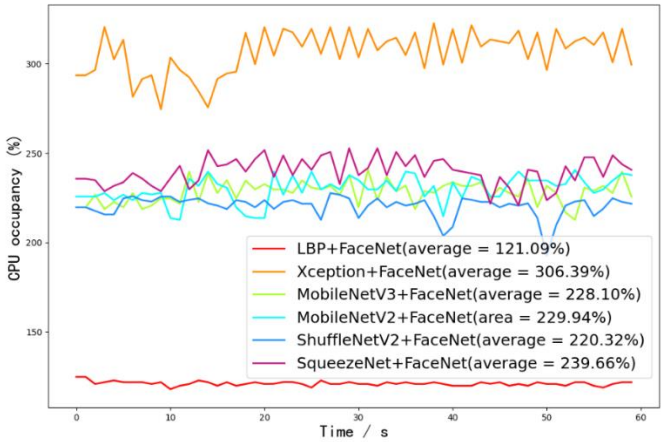

Figure 9. CPU usage of different algorithms.

4. Conclusions

In this paper, we have proposed a novel security classification framework based on face recognition which will possibly contribute to ACS in public condition. This framework is composed of a mask detection algorithm and face authentication algorithm with an anti-spoofing function. According to the epidemic prevention problem, the mask detection algorithm is integrated with a dexterous structure which contains a novel context attention head module to focus on the face and mask features, and a series of lightweight units to remove object cross-class. This configuration enables the mask detection module to meet the operation requirement of embedded devices. Besides, we proposed one novel liveness detection method that analyzes the color and texture of image frames using the uniform LBP method. In order to test the performance of the framework, this paper employs CASIA-FASD and Reply-Attack datasets as a benchmark. The test indicates that the HTER is 9.7%, the EER is 5.5%. The average process time of a single frame is 0.12 seconds. The above results demonstrate that this framework has a high anti-spoofing capability and can be employed on the embedded system to complete the mask detection and face authentication task in real-time, which can effectively support the public entry task under an epidemic background.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (No. 2020YFB1313600).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. B. Qin, D. Li, Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19, *Sensors*, **10** (2020), 5236. https://doi.org/10.3390/s20185236
2. M. S. Ejaz, M. R. Islam, M. Sifatullah, A. Sarker, Implementation of principal component analysis on masked and non-masked face recognition, *2019 1st Int. Conf. Adv. Sci., Eng. Rob. Technol. (ICASERT)*, 2019, 1–5. https://doi.org/10.1109/ICASERT.2019.8934543
3. M. Jiang, X. Fan, H. Yan, Retinamask: A face mask detector, *arXiv*, unpublished work.
4. J. Hosang, R. Benenson, B. Schiele, Learning non-maximum suppression, *2017 IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, 2017, 4507–4515. https://doi.org/10.1109/CVPR.2017.685
5. S. Woo, J. Park, J. Lee, I. Kweon, Cbam: Convolutional block attention module, *Proc. Eur. Conf. Comput. Vision (ECCV)*, 2018, 3–19.
6. Y. Taigman, M. Yang, M. A. Ranzato, L. Wolf, Deepface: Closing the gap to human-level performance in face verification, *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, 2014, 1701–1708.

7. Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, 2014, 1891–1898.
8. D. Nguyen, K. Nguyen, S. Sridharan, D. Dean, C. Fookes, Deep spatio-temporal feature fusion with compact bilinear pooling for multimodal emotion recognition, *Comput. Vis. Image Und.*, **174** (2018), 33–42. https://doi.org/10.1016/j.cviu.2018.06.005
9. J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition, *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)*, 2019, 4690–4699.
10. H. Liu, X. Zhu, Z. Lei, S. Z. Li, Adaptiveface: Adaptive margin and sampling for face recognition, *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)*, 2019, 11947–11956.
11. Y. Jiang, W. Li, M. S. Hossain, M. Chen, A. Alelaiwi, M. Al-Hammadi, A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition, *Inform. Fusion*, **53** (2019), 145–156. https://doi.org/10.1016/j.inffus.2019.06.019
12. Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, et al., Curricularface: Adaptive curriculum learning loss for deep face recognition, *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)*, 2020, 5901–5910.
13. Z. Boulkenafet, J. Komulainen, A. Hadid, Face anti-spoofing based on color texture analysis, *2015 IEEE Int. Conf. Image Proc. (ICIP)*, 2015, 2636–2640. https://doi.org/10.1109/ICIP.2015.7351280
14. Z. Boulkenafet, J. Komulainen, A. Hadid, Face spoofing detection using colour texture analysis, *IEEE T. Inf. Forensics Secur.*, **11** (2016), 1818–1830. https://doi.org/10.1109/TIFS.2016.2555286
15. X. Li, J. Komulainen, G. Zhao, P. C. Yuen, M. Pietikäinen, Generalized face anti-spoofing by detecting pulse from face videos, *2016 23rd Int. Conf. Pattern Recognit. (ICPR)*, 2016, 4244–4249. https://doi.org/10.1109/ICPR.2016.7900300
16. I. Chingovska, N. Erdogmus, A. Anjos, S. Marcel, Face recognition systems under spoofing attacks, In: T. Bourlai, *Face recognition across the imaging spectrum*, Springer, 2016, 165–194. https://doi.org/10.1007/978-3-319-28501-6_8
17. S. Q. Liu, X. Lan, P. C. Yuen, Remote photoplethysmography correspondence feature for 3D mask face presentation attack detection, *Proc. Eur. Conf. Comput. Vision (ECCV)*, 2018, 558–573.
18. I. Manjani, S. Tariyal, M. Vatsa, R. Singh, A. Majumdar, Detecting silicone mask-based presentation attack via deep dictionary learning, *IEEE T. Inf. Forensics Secur.*, 2017, 1713–1723. https://doi.org/10.1109/TIFS.2017.2676720
19. R. Shao, X. Lan, P. C. Yuen, Joint discriminative learning of deep dynamic textures for 3d mask face anti-spoofing, *IEEE T. Inf. Forensics Secur.*, **14** (2018), 923–938. https://doi.org/10.1109/TIFS.2018.2868230
20. J. Määttä, A. Hadid, M. Pietikäinen, Face spoofing detection from single images using micro-texture analysis, *2011 Int. Joint Conf. Biometrics (IJCB)*, 2011, 1–7. https://doi.org/10.1109/IJCB.2011.6117510
21. J. Määttä, A. Hadid, M. Pietikäinen, Face spoofing detection from single images using texture and local shape analysis, *IET Biom.*, **1** (2012), 3–10. https://doi.org/10.1049/iet-bmt.2011.0009

22. Y. Atoum, Y. Liu, A. Jourabloo, X. Liu, Face anti-spoofing using patch and depth-based CNNs, *2017 IEEE International Joint Conference on Biom. (IJCB)*, 2017, 319–328. https://doi.org/10.1109/BTAS.2017.8272713
23. Y. Liu, A. Jourabloo, X. Liu, Learning deep models for face anti-spoofing: Binary or auxiliary supervision, *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, 2018, 389–398.
24. G. Pan, L. Sun, Z. Wu, S. Lao, Eyeblink-based anti-spoofing in face recognition from a generic web camera, *2007 IEEE 11th Int. Conf. Comput. Vision*, 2007, 1–8. https://doi.org/10.1109/ICCV.2007.4409068
25. A. Zadeh, P. P. Liang, N. Mazumder, S. Poria, E. Cambria, L. P. Morency, Memory fusion network for multi-view sequential learning, *Thirty-Second AAAI Conf. Artif. Intell.*, **32** (2018), 5642–5649.
26. T. Baltruaitis, C. Ahuja, L. P. Morency, Multimodal machine learning: A survey and taxonomy, *IEEE T. Pattern Anal. Mach. Intell.*, **41** (2019), 154–163. https://doi.org/10.1109/TPAMI.2018.2798607
27. T. Li, Q. Yang, S. Rong, L. Chen, B. He, Distorted underwater image reconstruction for an autonomous underwater vehicle based on a self-attention generative adversarial network, *Appl. Opt.*, **59** (2020), 10049–10060.
28. T. Li, S. Rong, X. Cao, Y. Liu, L. Chen, B. He, Underwater image enhancement framework and its application on an autonomous underwater vehicle platform, *Opt. Eng.*, **59** (2020), 083102. https://doi.org/10.1117/1.OE.59.8.083102
29. K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, *IEEE Signal Proc. Let.*, **23** (2016), 1499–1503. https://doi.org/10.1109/LSP.2016.2603342
30. J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, et al., Dual attention network for scene segmentation, *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit. (CVPR)*, 2019, 3146–3154.
31. W. Sanghyun, H. Soonmin, I. S. Kweon, Stairnet: Top-down semantic aggregation for accurate one-shot detection, *2018 IEEE Winter Conf. Appl. Comput. Vision (WACV)*, 2018, 1093–1102. https://doi.org/10.1109/WACV.2018.00125
32. T. Ojala, M. Pietikäinen, T. Mäenpää, Gray scale and rotation invariant texture classification with local binary patterns, In: *Computer Vision-ECCV 2000*, Lecture Notes in Computer Science, Springer, **1842** (2000), 404–420. https://doi.org/10.1007/3-540-45054-8_27
33. T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE T. Pattern Anal. Mach. Intell.*, **24** (2002), 971–987. https://doi.org/10.1109/TPAMI.2002.1017623
34. W. S. Noble, What is a support vector machine? *Nat. Biotechnol.*, **24** (2006), 1565–1567. https://doi.org/10.1038/nbt1206-1565
35. F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size, *arXiv*, unpublished work.
36. F. Chollet, Xception: Deep learning with depthwise separable convolutions, *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, 2017, 1251–1258.

37. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, 2018, 4510–4520.
38. A. Howard, M. Sandler, G. Chu, L. C. Chen, B. Chen, M. Tan, et al., Searching for mobilenetv3, *Proc. IEEE/CVF Int. Conf. Comput. Vision (ICCV)*, 2019, 1314–1324.
39. N. Ma, X. Zhang, H. Zheng, J. Sun, Shufflenet v2: Practical guidelines for efficient CNN architecture design, *Proc. Eur. Conf. Comput. Vision (ECCV)*, 2018, 116–131.
40. Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, S. Z. Li, A face antispoofing database with diverse attacks, *2012 5th IAPR Int. Conf. Biom. (ICB)*, 2012, 26–31. https://doi.org/10.1109/ICB.2012.6199754
41. A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, Sebastien Marcel, The replay-mobile face presentation-attack database, *2016 Int. Conf. Biom. Spec. Interest Group (BIOSIG)*, 2016, 1–7. https://doi.org/10.1109/BIOSIG.2016.7736936

AIMS Press