CNN-based Online Access Control Recognition Method Using IoT and Microcontroller

Yan Su*, Yin Wu
College of Information Engineering, Zhengzhou University of Technology, Zhengzhou, 451191, China
E-mail: paperiset@163.com, zzjm_rgzn123@outlook.com
*Corresponding author

Keywords: internet of things, microcontroller, access control, face detection, CNN

Received: August 20, 2024

In order to improve the security and practicality of the access control system, an online recognition access control system based on the Internet of Things and microcontroller is designed. Taking STM32F103RCT6 microcontroller as the core control center, RFID technology is used for personnel information recognition, and convolutional neural networks are introduced for facial image processing. Meanwhile, Raspberry Pi 3B+ is used as an auxiliary controller to achieve liveness detection. The experiment was conducted under the Windows 10 operating system using Intel ® Core TM i5-10400F processor and 8GB memory are used for face detection under different lighting conditions to evaluate the robustness of the system. The results showed that the proposed method detected the target for the first time with an average frame rate of 194, which had stronger performance compared with support vector machines and convolutional neural networks. In addition, the accuracy of the system was 98.3%, and the final loss value was 0.012%. The research shows that this online identification access control system can effectively meet the needs of modern households and businesses for fast and accurate identity verification, demonstrating good practical prospects.

Povzetek: Zasnovan je izboljšan sistem prepoznave obraza za spletni dostop na osnovi CNN, IoT in mikrokrmilnikov.

1 Introduction

The Internet of Things (IoT) can utilize related devices for information exchange and transmission [1]. It can effectively improve the efficiency of production and management in smart devices, which has been well applied in the design of access control systems. The Online Recognition Access Control System (ORACS) can utilize IoT to achieve rapid personnel detection, reducing time and management costs. Radio Frequency Identification (RFID) technology is located in the perception layer of IoT [2]. RFID affects a company's operating costs, system reliability, and flexibility. However, it is easily affected by environmental factors such as temperature and humidity. Therefore, designing a suitable RFID system is related to the practical application effect of intelligent devices. The microcontroller is the terminal core controller of Access Control System (ACS), which is the core module for implementing data storage, processing, and resource allocation [3-4]. RFID and microcontroller are the core of IoT-based ACS. However, the existing technology cannot adapt to the constantly growing job demands, so further improvements are necessary. An et al. proposed a new deep framework that included a new hybrid spatial channel attention module to facilitate cross-age facial recognition tasks. Different pooling strategies were also combined when applying these spatial and channel attention mechanisms. This module generated differentiated facial representations while retaining complete information from the original features, further improving recognition accuracy [5]. To achieve brain tumor classification in IoT medical systems, Haq A U et al. used an improved convolutional neural network to classify meningiomas, gliomas, and pituitary type brain tumors. Experimental data showed that the classification accuracy of this model reached 99.89%, which was superior to existing classification techniques. [6]. Fu proposed the PLV algorithm for collecting facial information based on improved PCA. This proposed algorithm had higher stability and accuracy, which effectively recognized facial information [7]. As a result, intelligent algorithms have good application effects in facial recognition. Therefore, this study considers integrating facial recognition into ACS to improve the efficiency of personnel entry and exit. The recognition method based on Convolutional Neural Network (CNN) can effectively handle Facial Expression Recognition (FER). This method has high learning ability, which can integrate well with other technologies. Therefore, CNN and IoT-related technologies are combined to design a comprehensive ORACS. The innovation of this article lies in the combination of CNN-based FER and IoT-related technologies to improve the efficiency of

ORACS usage. The content consists of four chapters. The first part summarizes existing research. The second part introduces the relevant technologies of the IoT and the optimization methods of CNN. The third part is the performance testing and application effect analysis of the method. The final section summarizes the article and provides an outlook.

2 Related works

The development of IoT has promoted the practical application of intelligent devices in daily life. The IoT plays a guiding role in the development of smart homes, transportation, and cities. It can utilize existing devices to achieve communication between things or objects. The related technologies and elements of IoT are widely applied in various devices and systems [8]. RFID belongs to the perception layer of IoT, used for reading and writing related data and target recognition. Altaf et al. demonstrated that RFID could be used for human pose recognition. However, traditional methods could not achieve full body posture position estimation. Therefore, the neural network was used for learning error estimation. The improved method reduced environmental noise and performed signal segmentation, achieving real-time pose estimation [9]. He et al. proposed a new RFID authentication protocol that utilized hash functions and other computational methods for bidirectional authentication of information. This lightweight authentication protocol could ensure real-time communication between communication parties, while reducing computational costs and having high security [10]. The design of the ACS involved data processing structures such as the central processing unit, memory, and timer. A microcontroller belongs to a microcomputer system, which includes the data processing units mentioned above and can be used for intelligent devices. Zhang et al. confirmed that microcontrollers could achieve relevant circuit design in the smart curtains. The intelligent curtain with STC microcontroller as the core achieved intelligent control based on the design of circuits such as time and light [11]. Kosina pointed out that SiLabs chips could reduce the difficulty of digital control in wireless transmission, thereby improving the efficiency of wireless network usage. This chip covered both short and long waves in wireless transmission, simplifying the structure of wireless design and could be used for decoding FM stereo [12]. Biometric detection is information that access control and other authentication systems need to process correctly. The correct processing results were obtained by processing these data through the corresponding control system, achieving identity recognition in the ACS. However, this process is susceptible to malicious attacks, leading to the failure of the ACS. Murillo-Escobar et al. adopted a microcontroller system combined with hyper chaotic encryption to improve the security of the authentication system. The results confirmed that this new system resisted malicious attacks and improved security levels [13].

Intelligent algorithms can improve work efficiency and quality of life. FER is an important core of facial recognition ACS. However, human facial expressions are variable. Therefore, it is necessary to design a method that can recognize micro expressions. Thuseethan et al. designed an expression detection method using deep neural networks. Facial features from different regions were combined to achieve microscopic prediction of facial expressions. The results confirmed that the proposed method accurately recognized micro-expressions in videos, outperforming baseline methods [14]. CNN can learn facial expression features well and achieve accurate classification. Zhou et al. combined clustering algorithm and CNN for FER. The improved method was applied to dynamic pixel clustering and attention mechanism was used to adjust weights. This method effectively enhanced the regional feature expression of expressions, thereby improving the accuracy of FER, which was superior to existing advanced methods [15]. The poor mental state of drivers may lead to accidents. Therefore, their state detection is related to the safety of driving. Chand and Karthikeyan used CNN for driver sentiment analysis and facial recognition. This model correctly handled the facial expressions of drivers and accurately detected their behavior and emotions [16]. Zaman et al. improved the facial feature extraction of drivers using an improved CNN. This model could perform faster transfer learning, improve high-speed facial feature recognition, and demonstrate high accuracy [17]. Liu combined CNN with rough set theory to design a new FER model. This pattern recognized blurry facial expressions and improved the accuracy of facial feature extraction [18].

In the above research, the application of IoT related technologies and intelligent algorithms can effectively handle data processing related to intelligent control systems. However, existing access control technologies still suffer from issues such as susceptibility to dirt and wear on fingerprint recognition, and the possibility of passwords and RFID cards being hacked or stolen. In addition, the operation of these traditional methods is complex, and users often face multi-step processes in the authentication process, which undoubtedly causes inconvenience to users. A comprehensive ORACS is designed based on single-chip microcontrollers, utilizing IoT related technologies and CNN to filling a significant gap in fast, secure, and convenient identity authentication. Especially, by introducing live detection, it is possible to effectively distinguish between real faces and fake images, thereby significantly improving the security of the system. The relevant work summary table is shown in Table 1.

Table 1: Summary of related work

Literature	Method	Data set	Key indicators
Reference [9]	RFID	RFID-Pose-Recognition-Dataset	User experience, calibration accuracy
Reference [10]	New RFID Authentication Protocol	RFID_Authentication_Trace	Anti-replay attack capability, anti forgery capability
Reference [11]	STC microcontroller	STC8_Sensor_Data_Logs	Response time, failure rate
Reference [12]	SiLabs	SiLabs_FM_Decoding_Dataset	Decoding quality, interference immunity
Reference [13]	A microcontroller system based on hyper chaotic encryption	AccessControl_Auth_Dataset	Accuracy, efficiency
Reference [14]	Expression detection method using deep neural networks	VGGFace Dataset	Accuracy, processing time
Reference [15]	Combining clustering algorithm with CNN	CelebFaces Attributes Dataset	Accuracy, robustness
Reference [16]	CNN	Cohn-Kanade Dataset	Accuracy, precision, recall, F1-score
Reference [17]	Improved CNN	The Cognitive Activated State Classification	Accuracy, precision, recall
Reference [18]	Combining CNN with rough set theory	OpenFace Dataset	ROC curve and AUC value

3 Online recognition access control system and facial recognition optimization based on the internet of things and microcontroller

IoT can be used for information transmission and exchange in ACS. RFID belongs to its perception layer, which can read relevant personnel information. As a microcomputer system, a microcontroller includes functional modules such as RFID and communication. The first section describes ORACS based on IoT and microcontroller. In the second section, CNN is integrated to optimize this system.

3.1 Online recognition access control system based on the internet of things and microcontroller

ORACS needs to read relevant personnel information and obtain corresponding electronic tags using RFID. The RFID is controlled using a microcontroller to obtain the corresponding encoding information. Then, it is necessary to identify and judge this information to confirm whether the person can pass. Figure 1 shows the operation mode of RFID in ORACS. The picture is hand drawn by the author.

In Figure 1, the computer software is able to issue instructions to RFID and read electronic tags. Then, the system transmits the

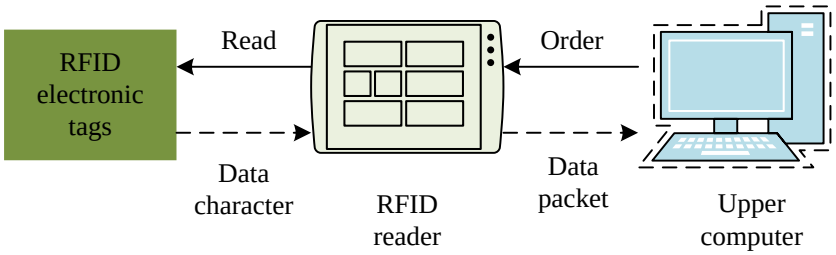


Figure 1: Operation mode of RFID in online identification access control system

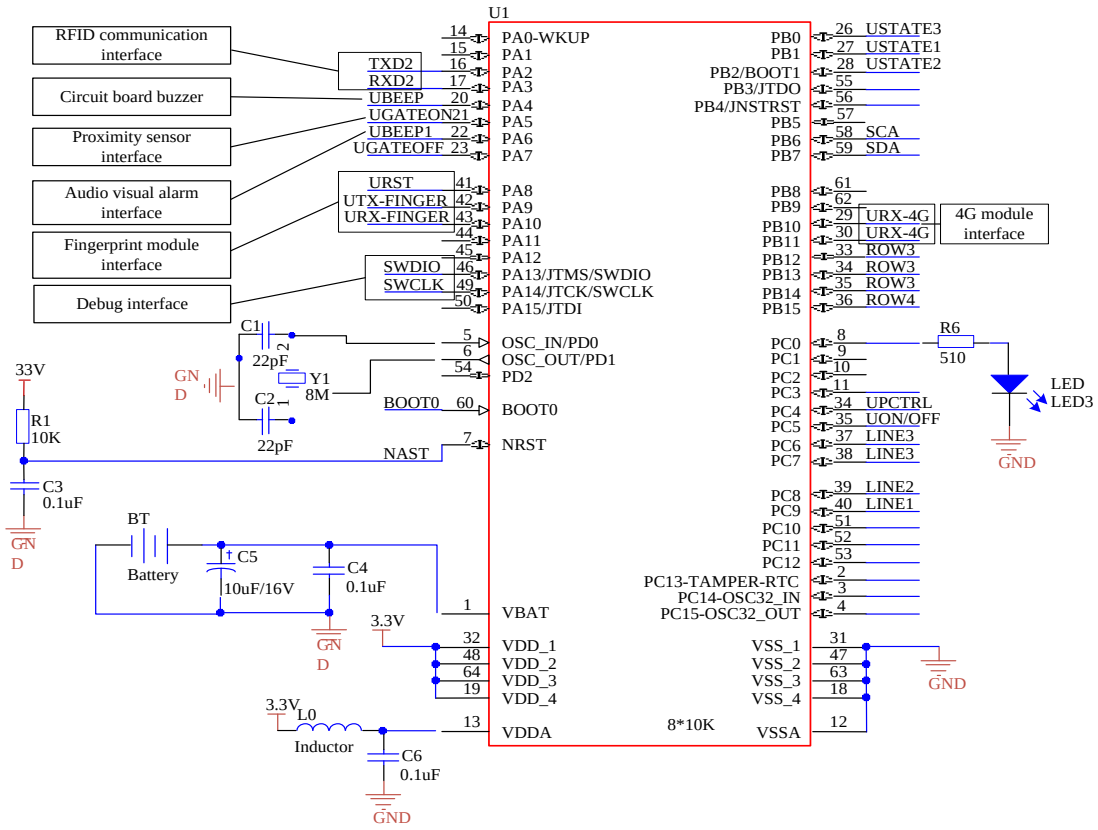


Figure 2: Schematic diagram of a single-chip

obtained information to the RFID reader for tag character interpretation. Next, it sends the interpreted content to the upper computer for the next step of program execution. The RFID technology can identify illegal intrusion information, improving the security and management effectiveness of ACS [19].

The microcontroller in ORACS requires certain technical requirements. These requirements include flexible programming, sufficient I/O interfaces, real-time response, large storage capacity, and the ability to work stably for a long time. The selection of microcontrollers also needs to consider economic costs. Energy efficient and low-cost microcontrollers are a priority consideration in ORACS. The STM32F103RCT6 microcontroller is selected as the core controller after comprehensive consideration. Figure 2 shows the schematic diagram of this

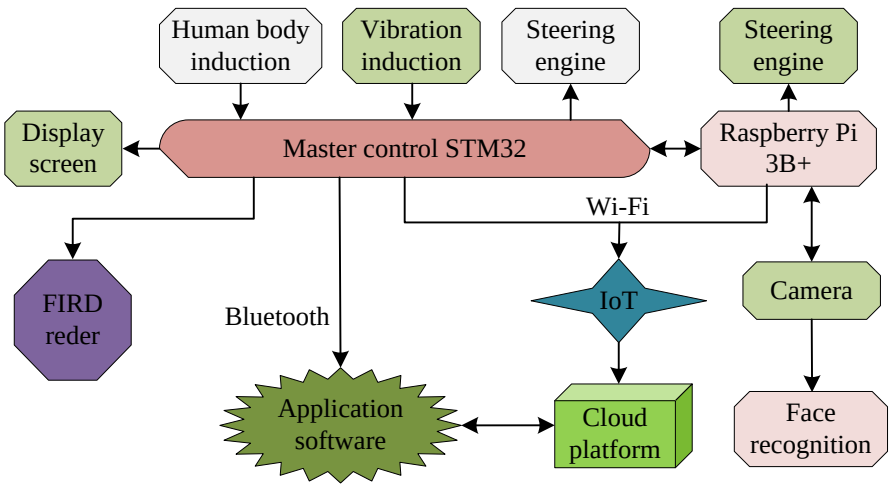


Figure 3: Overall framework of online recognition access control system

microcontroller. The picture is hand drawn by the author.

The above-mentioned microcontroller is used as the control center for RFID, communication, and other modules to achieve functions such as access control and communication. The STM32 series microcontrollers have a high internal capacity and a built-in memory [20]. In addition, they have higher stability and better information processing capabilities, which are suitable for the needs of ORACS. The selected microcontroller has a processing speed of 72MHz, a Random-Access Memory (RAM) of 48K, and a Read Only Memory (ROM) of 256K. The core of the selected microcontroller is a sequential logic circuit, which needs to be driven by external factors to execute instructions. There are a large number of peripherals in STM32 that require a clock to start. Multiple clock sources are used in this experiment to solve the high-power consumption caused by clock acceleration considering the compatibility of the equipment. Multiple clock sources, including high-speed and low-speed internal and external clock sources, as well as PLL, are used to handle different tasks.

The main core modules of ORACS are discussed in the above content. However, a single identification method is easily affected by illegal intrusion. Therefore, a facial recognition module is added to the ACS to improve the security and management efficiency of the system. Figure 3 shows the framework diagram of ORACS. The picture is hand drawn by the author.

In addition to the microcontroller-based main control module mentioned above, corresponding auxiliary control modules need to be combined in the designed ORACS. The function is to perform facial recognition and other tasks, requiring certain interpersonal interaction. Raspberry Pi 3B+ is selected as the auxiliary controller for ORACS in the experiment taking into account both cost and technical difficulty [21]. The data acquisition module of ACS can detect the human body. The human body thermoelectric sensor can reduce energy loss and wake up the system when personnel approach. Vibration sensors are used to remind users of the presence of outsiders, which can serve as anti-theft devices. The data transmission module mainly utilizes IoT for data processing. Wi-Fi and Bluetooth are used for data transmission, reducing the system cost. The servo control module is used to control the door lock. The MG996R servo is

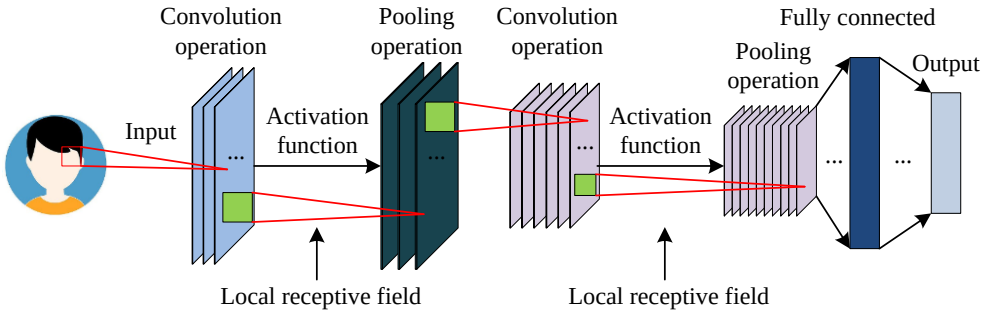


Figure 4: Feature extraction and classification process of CNN

selected in this experiment. The display module facilitates managers to pay attention to the power level of ACS and prompt visitors to enter. The system button circuit can help residents conveniently and quickly open the door lock. The experiment uses a capacitive button, which involves the charging and discharging of the RC circuit in formula (1).

$$V_t = v_0 + (v_1 - v_0) \times [1 - \exp(-t / RC)] \quad (1)$$

In formula (1), V_t is the instantaneous voltage. v_0 represents the initial value of voltage. v_1 refers to the final voltage. t represents time. RC is a time constant. When $v_0 = 0$, the above formula can be simplified, as shown in formula (2).

$$V_t = v_1 \times [1 - \exp(-t / RC)] \quad (2)$$

In formula (2), the capacitance is positively correlated with time, which means that increasing the charging time will increase the capacitance value. Capacitance is used to determine whether the door lock switch has been triggered.

3.2 Optimization of online access control automatic recognition method based on neural network

Traditional ACS mainly adopts methods such as access control, fingerprints, and passwords. However, these methods have issues such as poor flexibility,

difficult management, and low security [22]. Therefore, on the basis of previous designs for the IoT and microcontrollers, this study adds facial recognition functionality to improve the practical application effectiveness of ORACS. In the aforementioned ORACS, the facial recognition module is the focus. Then, a detailed introduction is provided. The training of facial recognition using CNN involves convolution and pooling operations. An activation function is introduced to change the image dimension and extract corresponding image features. Figure 4 shows the feature extraction and classification of CNN. The picture is hand drawn by the author.

In Figure 4, convolution and pooling operations, as well as the activation functions, are used for feature extraction. The fully connected layer reclassifies the feature information obtained through the above operations. Five consecutive convolutional layers are used to extract different levels of facial features. The convolution kernel size of each convolutional layer is 3x3. The number of convolution kernels in the first layer is set to 32, 64 in the second layer, 128 in the third layer, 256 in the fourth layer, and 512 in the fifth layer. Between convolutional layers, 2x2 max pooling layers are used for down-sampling to reduce the dimensionality of the feature map while preserving the main features. The setting of the pooling layer is that the first pooling layer follows closely after the first convolutional layer, and the other pooling layers are applied in the same order to reduce the dimensionality of each feature map. After each convolutional layer, the ReLU activation function is applied to increase the non-linear capability of the network. After passing through convolutional and pooling layers, multiple fully connected layers are connected for final classification, and the number of nodes in the output layer is matched with the facial recognition category. The initial learning rate is set to 0.001, and a learning rate decay strategy is adopted, which decays every 10 epochs. The model training takes a total of 50 epochs. During the training process, the batch size is set to 32 to balance training speed and memory usage. A loss function is used to calculate the training loss value of CNN [23]. Formula (3) is the cross-entropy loss function selected for this study.

$$Loss = -\sum_{j=1}^{n} y_i \ln a_j \tag{3}$$

In formula (3), n represents the number of classifications. y is the true value label. a is a constant. Face detection is divided into two labels: face and non-face. When y refers to a face, $\ln a$ represents the probability of being predicted as a face. The above formula can be simplified, as displayed in formula (4).

$$Loss = -[y \ln a + (1-y)\ln(1-a)] \tag{4}$$

When y represents a face, its value is set to 1. During the CNN training, the closer y is to 1, the smaller the value of the cross-entropy loss function, indicating more accurate results. A center loss function is introduced to calculate the center loss of the Euclidean distance of features to improve the training effect, represented by formula (5).

$$Loss = \frac{1}{2}\sum_{j=1}^{l} (\hat{y}_j^k - y_j^k)^2 \tag{5}$$

In formula (5), y_j^k represents the label of the sample. $\hat{y}_j^k$ refers to the actual network output value. l is the size of the small batch.

The central loss function can be used to fit continuous training samples, thereby obtaining the total loss function value of the entire CNN. The cross-entropy loss function and center loss function are used to jointly supervise the training of CNN. The Adam algorithm is used in this experiment to optimize the above loss function to further improve the robustness of the model, represented by formula (6).

$$\theta_i = \theta_{i-1} - \frac{\eta}{\sqrt{\hat{G}_i + \varepsilon}} \hat{d}_i \tag{6}$$

In formula (6), $\hat{G}_i$ and $\hat{d}_i$ are parameters for error correction. η is the learning rate. θ represents the exponential decay rate. Adam has advantages over other functions such as fast convergence speed and strong learning ability [24]. Adam can improve the oscillation of the loss function and correct slow convergence speed when η decreases.

Live detection can improve the accuracy of facial detection in ORACS. This method can achieve detection based on relevant human movements, which is a commonly used eyelid detection method. The curvature of the eyelids needs to be calculated, represented by formula (7).

$$\rho = \frac{\mathrm{V}\alpha}{\mathrm{V}L} \tag{7}$$

In formula (7), $\mathrm{V}\alpha$ represents the angle at which the tangent of the eyelid curve changes. $\mathrm{V}L$ is the arc length of the eyelids. The state of the eyes can be determined based on the curvature calculation. Human live detection can analyze color and texture to determine whether there is facial deception. Figure 5 shows a CNN-based facial detection framework. The picture is hand drawn by the author.

In Figure 5, RGB is a commonly used color space. The HSV color space uses hue (H), saturation (S), and brightness (V) to describe colors. The YCbCr color space is described by brightness (Y), blue concentration offset component (CB), and red concentration offset

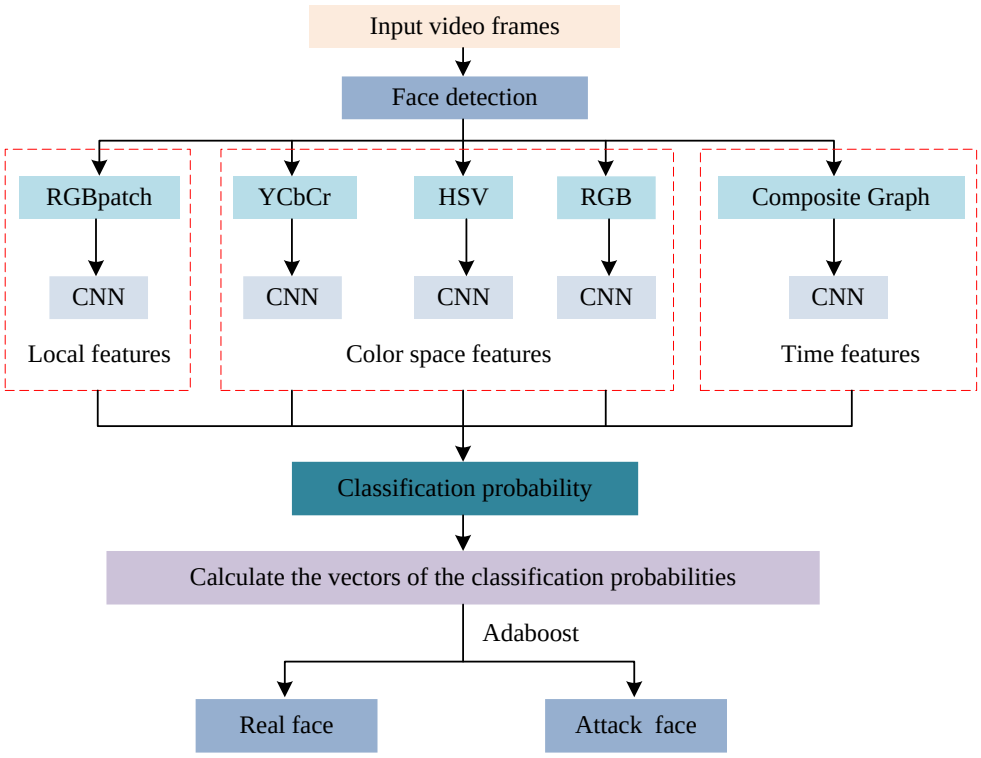


Figure 5: Framework diagram of facial detection

component (CR). RGB, HSV, and YCbCr have different emphasis on color description rules, which can better provide supplementary facial color and texture descriptions for live detection. The advantage of using RGB is that it is easy to understand and process. Especially under good lighting conditions, the RGB space can effectively preserve detailed information in facial images. However, the RGB color space is highly sensitive to changes in lighting, so colors may become distorted under different lighting conditions. The HSV color space is different from RGB in that the brightness component of HSV space is separated from color information, which allows the system to better maintain facial color features under lighting changes. The YCbCr color space is particularly suitable for color expression in video stream data. YCbCr separates brightness and color information for easy processing and analysis. The independence of brightness signals enables the system to better resist the influence of uneven lighting when detecting faces. The research adopts a feature extraction strategy that combines RGB, HSV, and YCbCr color spaces, forming a multi-level feature description. RGB provides basic color information, while HSV and YCbCr enhance adaptability to lighting changes. By integrating the features of these color spaces, the system can accurately recognize faces in complex environments and reduce misidentification caused by lighting changes. Facial images are transformed in RGB, HSV, and YCbCr spaces to extract texture information from each color space. Afterwards, the extracted information is trained through CNN. Finally, the classification results are obtained through an SVM classifier. Finally, SVM is used as the classifier to output the classification results, distinguishing the real faces and attacking faces. By combining features extracted from different color channels, texture descriptions originally designed for grayscale images can be applied to color images. The color and texture information of facial images need to be analyzed in five steps: local binary mode, co-occurrence local binary mode, local phase quantization, binarization statistical image features, and scale invariant feature transformation. The dataset used for model training includes two, namely Colorferet and Large-scale CelebFaces Attributes (CelebA) Dataset. The Colorferet dataset includes a universal face library and universal testing standards. This includes over 10000 photos of over 1000 people, each with different expressions, lighting, poses, and ages. The Colorferrot dataset is commonly used for facial recognition. The CelebA dataset is a large-scale facial recognition dataset published by Professor Tang Xiaoou's laboratory at the Chinese University of Hong Kong. The CelebA dataset is a large-scale facial attribute dataset with over 200000 celebrity images and over 40 types of facial attributes. Each image has 40 annotation attributes. The images in the dataset contain significant pose changes and background confusion. The CelebA dataset is mainly used for facial attribute recognition. The faces recorded in the entry control system belong to real faces. A face that has not been entered into ACS is considered an attacking

face. In addition, the attacking face also includes images generated by artificial intelligence. Formula (8) is the judgment formula for Adaboost.

$$h_i(x)=\begin{cases}1 & p_i f_i < p_i \vartheta_i \\ 0 & others\end{cases} \tag{8}$$

In formula (8), p_i is the offset. ϑ_i refers to the threshold. f_i is the characteristic value. When training Adaboost, the f_i of all samples is calculated. Then, a range is obtained as the threshold. ϑ_i needs to minimize the weighted classification error, as shown in formula (9).

$$e=\min\left\{(P^{'}+(N-N^{'})),(N^{'}+(P-P^{'}))\right\} \tag{9}$$

In formula (9), P refers to the total weight of the face set. $P^{'}$ represents the total weight of samples less than f_i. N is the sum of weights for a set of non-human faces. $N^{'}$ represents the total weight of all samples greater than f_i. The sample weights in formula (10) are set to train the weak classifier $h_j(x)$.

$$w_{1,i}(x)=\begin{cases}\frac{1}{2p} & y_i=1 \\ \frac{1}{2n} & y_i=0\end{cases} \tag{10}$$

In formula (10), p means the total positive samples. n means the total negative samples. Formula (11) represents the normalized weight.

$$w_{t,i}(x)=\frac{w_{t,i}}{\sum_j^n w_{t,i}} \tag{11}$$

In formula (11), t is the training coefficient. According to formula (12), $h_j(x)$ is selected to minimize the error.

$$\varepsilon_j=\sum_{i=1}^n w_t\left|h_j(x_i)-y_i\right| \tag{12}$$

Then, the weights are updated using formula (13).

$$w_{t+1,i}=w_{t,i}\beta_t^{1-\varepsilon_i} \tag{13}$$

In formula (13), β refers to the weight coefficient. β_t is represented by formula (14).

$$\beta_t=\frac{\varepsilon_t}{1-\varepsilon_j} \tag{14}$$

When $\varepsilon_t=1$, the classification of sample x_t is incorrect. When $\varepsilon_t=0$, the classification of x_t is correct. Finally, $h_j(x)$ is combined to obtain the strong classifier, as displayed in formula (15).

$$H(x)=\begin{cases}1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2}\sum_{t=1}^T \alpha_t \\ 0 & others\end{cases} \tag{15}$$

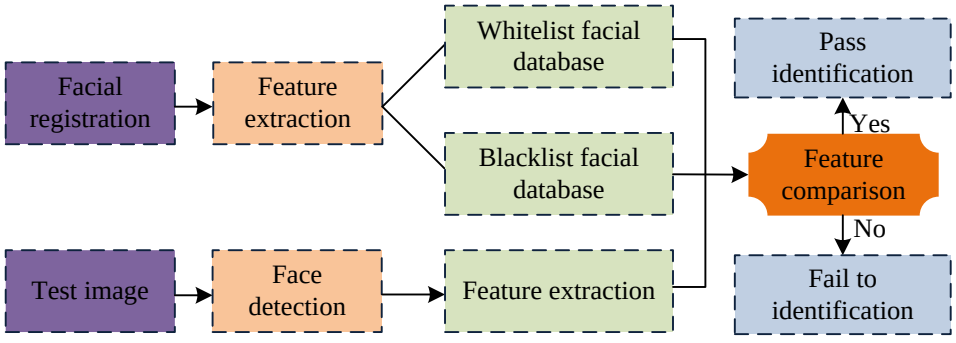


Figure 6: Framework diagram of facial recognition system

In formula (15), α_t refers to the weight value. Based on the above method, the facial recognition framework diagram in Figure 6 is designed and applied to ORACS. The picture is hand drawn by the author. In this system, the first step is to register facial information and then perform facial and live detection. During this process, feature extraction and comparison are required to filter out irrelevant information. Useful feature information are compared and recognition results are obtained.

The face recognition in this study is the foundation of face detection. In facial recognition, it is necessary to detect the input facial information through methods such as FER. Non-input faces belong to attacking faces. Identifying existing facial features can eliminate attacking faces and achieve facial recognition. Therefore, a distinction is made between real faces and attacking faces, which is essentially a facial recognition function. The paper elaborates on facial recognition as a fundamental method and applies it to face detection.

4 Performance test of online access control automatic recognition method

In order to verify the practical application effect of the proposed method, simulation analysis is conducted in the experiment. The testing environment includes Intel ® Core i5-10400F, 6 cores/12 threads, with a base frequency of 2.9GHz and 8GB DDR4 2400MHz RAM. The GPU is NVIDIA GeForce GTX 1660. The operating system is Windows 10 (64 bit). The development environment is Python version 3.7. Deep learning framework: TensorFlow 2. The main datasets are Colorferet dataset and CelebA dataset. This study divides the Colorferet dataset into training set, validation set, and testing set proportionally, accounting for 70%, 15%, and 15%, respectively. The Celebra dataset is divided into 60%, 20%, and 20% of the training set. The experimental indicators include accuracy, average response time, success rate, mean square error and AUC value, precision, recall rate, F1 score, as well as false positive and false negative rates. Among them, accuracy is used to measure the proportion of correctly classified models. The average response time is used to measure the average time required from receiving input to outputting response. The success rate refers to the proportion of correctly identified real faces in the test. The AUC value is used to quantify the model's ability to distinguish faces. Accuracy is used to measure the overall accuracy of the model. Recall is used to evaluate the proportion of true positive samples recognized by the model. F1 score takes into account both accuracy and recall, which is the harmonic mean of the two. All performance metrics reported are based on the test set. The cross-validation method is adopted when calculating AUC and success rate. In Table 2, the average response time of door locks under radio frequency, password, fingerprint, and facial unlocking modes, as well as interference conditions, was tested. Compared with single password, radio frequency,

Table 2: Unlocking situations under different modes

Unlock time comparison			
Unlock modes	Description	Average response time(t/s)	-
Password unlocking	Enter password to end, open the door lock	0.37	-
Fingerprint unlocking	Enter the user's fingerprint to open the door lock	1.22	-
RF unlocking	IC card matching, unlock the door lock	3.65	-
Comparison of unlocking situations in the presence of interference			
Unlock modes	Description	Success times/Experiment times	Variance(s2)
Password unlocking	Other button interference	54/60	0.05
Fingerprint unlocking	Finger stain	42/60	0.72
RF unlocking	Electromagnetic jammer	40/60	1.35
Facial unlocking	Attacking face	56/60	0.02

and fingerprint unlocking methods, the average response time of the door lock in the facial recognition method was the shortest, at 0.24s. This facial recognition method had the highest success rate and was the most stable under different external disturbances. The reasons for the low success rate of the other three methods may be as follows. Fingerprint contamination can cause significant instability in the reading of fingerprint images during unlocking. The success rate of password unlocking is relatively high. However, the operations it requires are relatively cumbersome. RFIC cards are susceptible to electromagnetic interference in RF unlocking, resulting in signal distortion. Overall, facial recognition can achieve both fast response and high stability.

In the designed facial recognition methods, live detection is involved, including eyelid detection. Figure 7 compares the eyelid detection results of different test samples under three lighting conditions. The eyelid detection result was the best, with an average detection efficiency of 85.8% when the detection method was image-based. The average effective detection rate when the detection object was staff ranked second, at 56.7%. The performance of eyelid detection was the worst in different environments, at 45.8% when the detection method was video. This may be because the video contains too many factors and is in a dynamic state, which can affect the effectiveness of eyelid detection. Overall, the average detection efficiency of these three

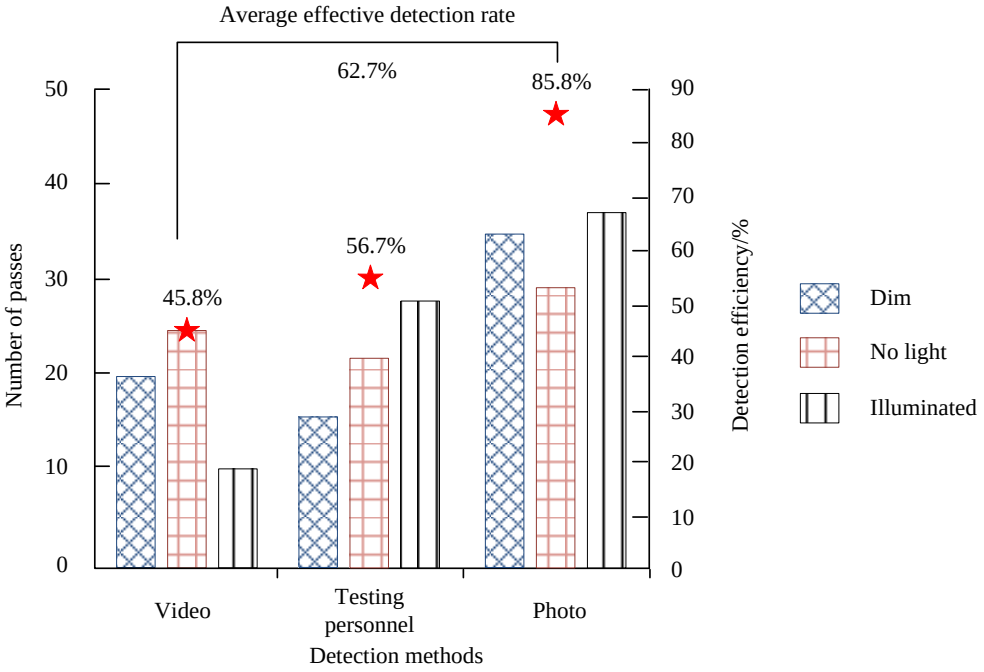


Figure 7: Statistics of correct number of live detections

methods are 62.7%, which can meet the requirements of live detection in facial recognition.

The proposed method is compared with Support Vector Machine (SVM), CNN, and CNN Long Short-Term Memory (CNN-LSTM) to verify its superiority [25]. SVM is a common classification and discrimination method. In the machine learning, SVM is a supervised learning model commonly used for pattern recognition, classification, and regression analysis. CNN is a feedforward neural network whose artificial neurons can respond to surrounding units within a certain coverage range. CNN has excellent performance in large-scale image processing. SVM and CNN can serve as basic methods. CNN-LSTM is a new comprehensive algorithm proposed by Adebowale MA et al. in 2023. This method is improved by combining techniques such as phishing detection, making it a relatively advanced algorithm. Therefore, this method is taken as a comparative method to verify the superiority of the proposed method. Live detection can be affected by the lighting environment. Therefore, Table 3 compares the frame rates at which different algorithms first detect targets under different lighting conditions. The average frame rate of the proposed method for the first target detection was 194, CNN-LSTM was 217, CNN was 325, and SVM was 389. The proposed method required the smallest number of frames to detect the target for the first time, with the fastest detection speed among all methods. This is because the proposed method introduces a conversion method for different color texture information, which improves its detection accuracy in different lighting environments. Therefore, the proposed method can quickly lock the target regardless of the

lighting conditions and is less affected by external environments.

To further verify the influence of lighting conditions on facial recognition performance, comparative experiments are conducted under different lighting conditions. The results are shown in Table 4. In Table 4, under normal light conditions, the average recognition accuracy was 90.2%. Under side lighting, the

Table 3: Frames for the first detection of faces using different methods

Video sequence	This paper	CNN-LSTM	CNN	SVM
Illuminated (front)	185	210	315	378
Illuminated (side)	190	217	325	391
No light (front)	179	211	316	379
No light (side)	182	209	313	371
Dim (front)	217	231	346	415
Dim (side)	213	225	337	405
The average frame rate at which a face is first detected	194	217	325	389

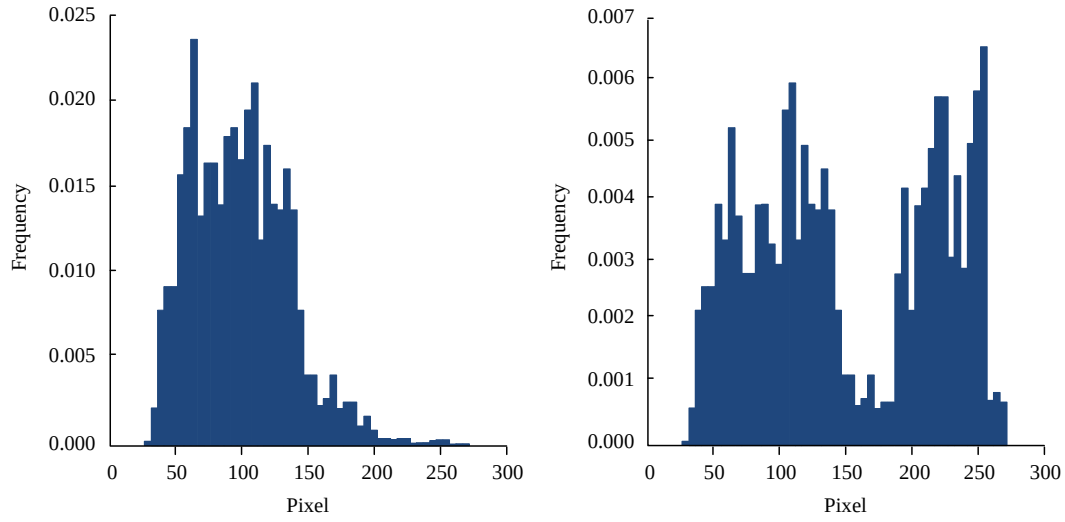
Table 4: Comparison of recognition accuracy under different lighting conditions

Lighting conditions	Accuracy (%)	Number of samples
Normal light	90.2	500
Matt	76.8	500
Side lighting	86.2	500
Weak light	81.6	500

average recognition accuracy was 86.2%. In the absence of light, the average recognition accuracy was 76.8%. Under low light conditions, the average recognition accuracy was 81.6%. From the experimental results, there were significant differences in accuracy and detection rates under different lighting conditions, with a value of p <0.001. Experimental data shows that different lighting conditions have a significant impact on recognition accuracy, with normal light having the highest recognition accuracy and no light having the lowest recognition accuracy. This may be due to facial features becoming blurry in low light conditions, leading to difficulty in recognition. Under side lighting conditions, although there is some improvement compared with no light and low light conditions, facial features are still affected to some extent due to changes in the angle of light. A module for dynamically adjusting lighting is considered for design to enhance robustness in various environments.

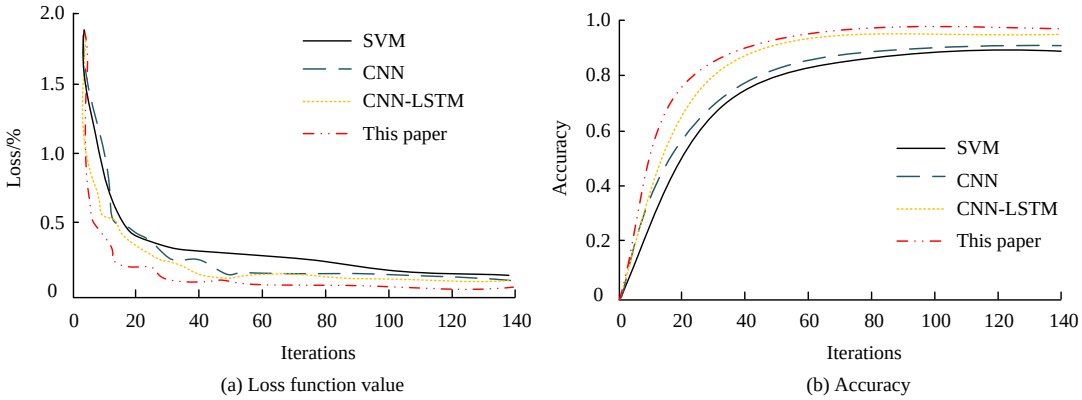
Due to the susceptibility of ORACS to external attacks, Figure 8 tests the recognition performance of the proposed method on both attacking and real faces. Figure 8 (a) shows the histogram of the attacking face, which is the result of taking printed photos as the attacking subject. Figure 8 (b) shows a histogram of a real face. There was a significant difference between the histograms obtained from attacking faces and those of real faces. This is because the printed photos are made of smooth materials, which can easily cause overexposure and result in significant differences from real faces. Therefore, the proposed method can effectively detect certain attack behaviors.

Figure 9 shows the loss functions and accuracy results of each method, which can be



(a) Attack face (b) Real face

Figure 8: Brightness histograms of facial regions under screen attacks and real faces



(a) Loss function value (b) Accuracy

Figure 9: Loss function and accuracy results

used to reflect the training effectiveness of the model. Low loss function values and high accuracy indicate that the model obtains more accurate results during training. In Figure 9 (a), each loss function value significantly decreased, indicating that as training increased, the performance of the model improved. The final loss value of the proposed method was the lowest, at 0.012%. In Figure 9 (b), the accuracy of all methods was improved, and the proposed method had the highest accuracy of 98.3%.

Mean Absolute Error (MAE) and Receiver Operating Characteristic (ROC) are selected as validation metrics to further validate the performance of the proposed method. Figure 10 shows the results of MAE and ROC for each method. A large area under the ROC curve and a small MAE value indicates good training effectiveness of the method. In Figure 10 (a), the proposed method had the largest area under the curve, with an

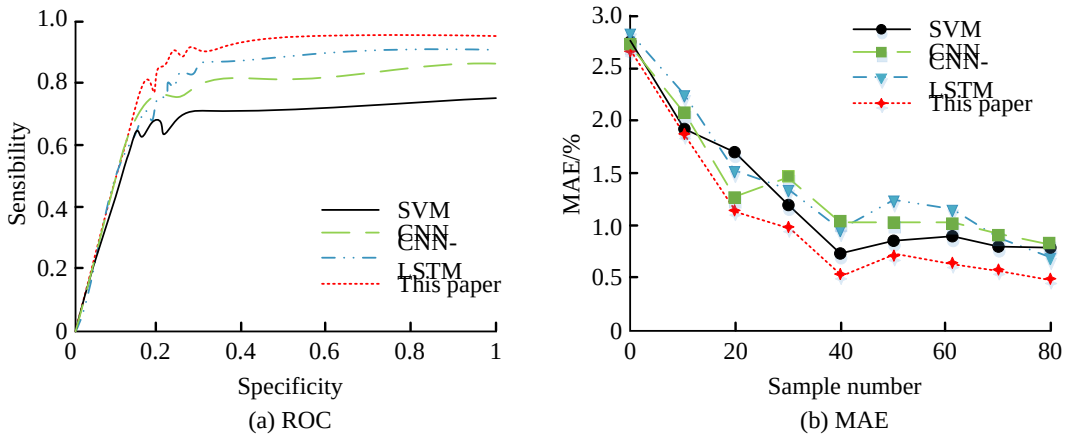


Figure 10: ROC and MAE results of different methods

Table 5: Comparison results with current advanced methods

Model	Average accuracy (%)	Average response time (s)	Number of samples
ORACS	92.4	0.22	500
RetinaFace	89.6	0.31	500
MTCNN	88.9	0.33	500

ROC value of 0.87. In Figure 10 (b), its MAE was the lowest, at 0.5%. The ROC values were higher than other methods, and the MAE was lower than other methods. This once again confirms that the proposed method achieves high performance after continuous learning.

To verify the superiority of the method over other advanced methods, accuracy and detection rate are used as indicators. The current advanced methods RetinaFace and MTCNN are selected for comparative experiments, and the results are shown in Table 5. According to Table 5, ORACS outperformed the comparative methods in terms of accuracy, detection rate, and average response time. Although RetinaFace and MTCNN have also shown good performance, they are not as fast and accurate as the proposed method.

To comprehensively evaluate the performance of the model, accuracy, recall, and F1 score indicators are used for evaluation, and the results are shown in Table 6. According to Table 6, the model accuracy was 95%. The recall rate was 92%, indicating that most real faces were correctly recognized. The F1 score of the model was 0.93, indicating that the model had good balance performance in facial recognition. The false alarm rate was 1% (10/500), which was relatively low in ACS safety critical systems, indicating that the system can accurately recognize faces in most cases. The false negative rate was 8% (40/500), indicating that the system failed to recognize some actual positive samples in certain situations, which may be due to uneven lighting or facial occlusion.

Based on the above performance testing experiments, the ORACS designed in this experiment can effectively recognize facial features during face detection. The proposed method improves its performance through continuous learning. In practical applications, the proposed ACS can meet various needs.

Table 6: Evaluation results of model accuracy, recall rate, F1 value, false alarm rate, and false negative rate indicators

Index	Sample size	Test result
Accuracy	500	95%
Recall	500	92%
F1 value	500	0.93
False alarm rate	500	1%
False negative	500	8%

5 Discussion

This study proposes an online access control automatic recognition system based on the IoT and microcontrollers. Then, the proposed method is compared with state-of-the-art algorithms such as CNN-LSTM and SVM. The superiority of this method is mainly due to the integration of fast image processing technology and optimized CNN algorithm in the system architecture. Comparatively, traditional algorithms such as SVM are usually computationally complex and require more time to process features. Therefore, the proposed ORACS has significant advantages in speed. In terms of success rate, ORACS had the highest success rate in testing, reaching 56/60. This result indicates that in dynamic environments, ORACS can effectively recognize facial features of real people, while SVM and CNN-LSTM perform relatively poorly in handling similar situations. The reason for this difference may be related to the feature extraction and classification mechanisms used. CNN can extract more detailed facial features through its deep structure, enhancing its ability to distinguish between real and fake faces. However, despite the strong performance demonstrated by ORACS, there are still certain limitations. The performance of facial recognition may be affected under different lighting conditions, such as lighting conditions and image quality. Under low light conditions, facial features may become blurry, leading to a decrease in recognition accuracy. In addition, changes in image quality, such as resolution and noise levels, may also have a significant impact on recognition results. The results indicate that the detection efficiency of dynamic videos is lower than that of static images, suggesting that motion blur and background interference may pose more challenges to facial recognition. In summary, although ORACS performs better than existing advanced methods in multiple indicators, the impact of environmental factors still needs to be considered in practical applications.

6 Conclusion

The ORACS needs to meet the growing demand. A fast and accurate online recognition method can improve the practical application effect of ACS. An ORACS was designed based on IoT and microcontroller to improve the convenience and safety of ACS in daily life. CNN was used to optimize it to improve the management effectiveness and security of this system. The average response time of the door lock in the facial recognition method was the shortest, at 0.24s compared with single password, radio frequency, and fingerprint unlocking methods. This facial recognition method had the highest success rate and was the most stable under different external disturbances. When detecting images, videos, and testers, the average detection efficiency of the proposed method was 62.7%, which met the requirements of live detection in facial recognition. Under different lighting conditions, the proposed method achieved an average frame rate of 194 when detecting a target for the first time, which was faster than other methods on detection speed. This method could detect attack behaviors that were mainly based on images. In training, the proposed method had the lowest final loss value, at 0.012%. The accuracy was the highest, which was 98.3%. The area under the ROC curve of this method was the largest, at 0.87. The MAE was the lowest, at 0.5%. Overall, the proposed ORACS can efficiently and accurately perform human detection, which can be applied in practical life. Although the above research achieves certain results, there are still some shortcomings. When developing a more effective ACS, it is necessary to purchase servers with larger bandwidth and more storage. In future research, the experimental content needs to be further improved to enhance the access control recognition system. The test results of on-site sample collection do not meet the ideal test set images. The analysis may be due to significant environmental interference factors in the sample collected during testing. The feature recognition rate is not ideal, especially in low light and hat occlusion situations. Therefore, it is necessary to consider the effects of light, obstructions, and other disturbance factors on facial recognition in ACS in future research. Different skin color, body shape, and other characteristics of different populations may have an impact on the facial recognition function of ACS. Therefore, in future research, the role of body shape, age, skin color, gender, and specific populations in facial recognition should be considered. This can improve the accuracy of the online access control recognition system. This paper aims to improve the effectiveness of online

access control recognition systems in practical life through method improvements.

7 Authors contribution

In this paper, Online Access Control Automatic Recognition based on the Internet of Things and Microcontroller. Yan Su put forward the research experiment: the proposed online recognition access control system can perform efficient and accurate face detection and live detection. Yin Wu analyzed the data and helped with the constructive discussion. Yan Su and Yin Wu made great contributions to manuscript preparation.

References

[1] M. A. Albreem, A. M. Sheikh, M. J. K. Bashir, and A. A. El-Saleh, "Towards green Internet of Things (IoT) for a sustainable future in Gulf Cooperation Council countries: Current practices, challenges and future prospective," Wireless Networks, vol. 29, no. 2, pp. 539-567, 2022. https://doi.org/10.1007/s11276-022-03133-3

[2] S. Lowes, S. El Tahir, S. Koo, S. Amonkar, A. Leaver, and R. Milligan, "Pre-operative localisation of axillary lymph nodes using radiofrequency identification (RFID) tags: A feasibility assessment in 75 cases," Clinical Radiology: Journal of the Royal College of Radiologists, vol. 78, no. 9, pp. E668-E675, 2023. https://doi.org/10.1016/j.crad.2023.05.017

[3] J. Flanagan, and C. Mcgovernm, "A qualitative study of improving the operations strategy of logistics using radio frequency identification," Journal of Global Operations and Strategic Sourcing, vol. 16, no. 1, pp. 47-68, 2023. https://doi.org/10.1108/JGOSS-04-2021-0030

[4] E. Abbasian, B. Grailoo, and M. Nayeri, "Design of a 10-nm FinFET 11 T Near-Threshold SRAM cell for low-energy Intemet-of-Things applications," Circuits, Systems, and Signal Processing: CSSP, vol. 42, no. 5, pp. 3138-3151, 2023. https://doi.org/10.1007/s00034-022-02251-9

[5] W. An, and G. Wu, "Hybrid spatial-channel attention mechanism for cross-age face recognition," Electronics, vol. 13, no. 7, pp. 1257-1274, 2024. https://doi.org/10.3390/electronics13071257

[6] A. U. Haq, J. P. Li, R. Kumar, Z. Ali, I. Khan, M. I. Uddin, and B. L. Y. Agbley, "MCNN: A multi-level CNN model for the classification of brain tumors in IoT-healthcare system," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 5, pp. 4695-4706, 2023. https://doi.org/10.1007/s12652-022-04373-z

[7] B. Fu, "Face recognition of remote monitoring under the Ipv6 protocol technology of Internet of Things architecture," Journal of Intelligent Systems, vol. 32, no. 1, pp. 20220283, 2023. https://doi.org/10.1515/jisys-2022-0283

[8] D. Gera, S. Balasubramanian, A. Jami, "CERN: Compact facial expression recognition net," Pattern Recognition Letters, vol. 155, pp. 9-18, 2022. https://doi.org/10.1016/j.patrec.2022.01.013

[9] S. Altaf, M. Haroon, S. Ahmad, E. A. Nasr, M. Zaindin, S. Huda, and Z. Rehman, "Radio-frequency-identification-based 3D human pose estimation using knowledge-level technique," Electronics, vol. 12, no. 2, pp. 374-400, 2023. https://doi.org/10.3390/electronics12020374

[10] J. He, C. Peng, Z. Fu, D. Xu, and H. Tang, "Lightweight bidirectional authentication protocol for RFID," Computer Engineering and Applications, vol. 59, no. 18, pp. 268-277, 2023. https://doi.org/10.6633/IJNS.202401 26(1).13

[11] Q. Zhang, J. Zhao, H. Wen, and H. Hao, "Design of intelligent curtain control circuit based on single chip microcomputer," Proceedings of International Conference on Artificial Life and Robotics, vol. 26, no. 1, pp. 676-679, 2021. https://doi.org/10.5954/ICAROB.2021.OS12-7

[12] C. Kosina, "Single-chip silicon labs FM/AM/SW digital radio receiver," Practical Electronics, vol. 51, no. 7, pp. 16-23, 2022.

[13] M. A. Murillo-Escobar, R. M. López-Gutiérrez, C. Cruz-Hernández, E. E. Espinoza-Peralta, and D. Murillo-Escobar, "Secure access microcontroller system based on fingerprint template with hyperchaotic encryption," Integration, vol. 90, pp. 27-39, 2023. https://doi.org/10.1016/j.vlsi.2023.01.002

[14] S. Thuseethan, S. Rajasegarar, and J. Yearwood, "Deep3DCANN: A deep 3DCNN-ANN framework for spontaneous micro-expression recognition," Information Sciences: An International Journal, vol. 630, no. 1, pp. 341-355, 2023. https://doi.org/10.1016/j.ins.2022.11.113

[15] L. Zhou, Y. Wang, B. Lei, and W. B. Yang, "Regional self-attention convolutional neural network for facial expression recognition," International Journal of Pattern Recognition and Artificial Intelligence, vol. 36, no. 8, pp. 1-24, 2022. https://doi.org/10.1142/S0218001422560134

[16] H. V. Chand, and J. Karthikeyan, "CNN based driver drowsiness detection system using emotion analysis," Computers, Materials and Continua (Tech Science Press), vol. 31, no. 2, pp. 717-728, 2022. https://doi.org/10.32604/IASC.2022.020008

[17] K. Zaman, Z. Sun, S. M. Shah, M. Shoaib, L. Pei, and A. Hussain, "Driver emotions recognition based on improved faster R-CNN and neural architectural search network," Symmetry, vol. 14, no. 4, pp. 687-709, 2022. https://doi.org/10.3390/sym14040687

[18] J. Y. Liu, "An improved CNN algorithm with hybrid fuzzy ideas for intelligent decision classification of human face expressions," Soft Computing, vol. 27, no. 9, pp. 5195-5204, 2023. https://doi.org/10.1007/s00500-023-07840-7

[19] S. G. Kim, T. V. Tran, and J. S. Lee, "Iron oxide-immobilized porous carbon nanofiber-based radio frequency identification (RFID) tag sensor for detecting hydrogen sulfide," Journal of Industrial and Engineering Chemistry, vol. 112, no. 1, pp. 423-429, 2022. https://doi.org/10.1016/j.jiec.2022.05.038

[20] M. Gams, and T. Kolenik, "Relations between electronics, artificial intelligence and information society through information society rules," Electronics, vol. 10, no. 4, pp. 514, 2021. https://doi.org/10.3390/electronics10040514

[21] J. A. Bacus, and N. B. Linsangan, "Detection and identification with analysis of carica papaya leaf using android," Journal of Advances in Information Technology, vol. 13, no. 2, pp. 162-166, 2022. https://doi.org/10.12720/jait.13.2.162-166

[22] H. A. Mohammad, and I. M. Husien, "A deep transfer learning framework for robust IoT attack detection: A review," Informatica, vol. 48, no. 12, pp. 55-64, 2024. https://doi.org/10.31449/inf.v48i12.5955

[23] Y. Y. Fanjiang, C. C. Lee, Y. T. Du, and S. J. Horng, "Palm vein recognition based on convolutional neural network," Informatica, vol. 32, no. 4, pp. 687-708, 2021. https://doi.org/10.15388/21-INFOR462

[24] Y. Wang, "A convolutional neural network method based on Adam optimizer with power-exponential learning rate for bearing fault diagnosis," Journal of Vibroengineering, vol. 24, no. 4, pp. 666-678, 2022. https://doi.org/10.21595/jve.2022.22271

[25] M. A. Adebowale, K. T. Lwin, and M. A. Hossain, "Intelligent phishing detection scheme using deep learning algorithms," Journal of Enterprise Information Management, vol. 36, no. 3, pp. 747-766, 2023. https://doi.org/10.1108/JEIM-01-2020-0036