

# **Application of AI Face Recognition Technology in Swipe Card Attendance Systems for Hospitals**

Te-Kwei Wang<sup>1,\*</sup>, Yu-Hsun Lin<sup>2</sup>, Kai-Ping Li<sup>1</sup>

<sup>1</sup>Department of Electrical Engineering, Ming Chi University of Technology, New Taipei City, Taiwan

<sup>2</sup>Department of Business and Management, Ming Chi University of Technology, New Taipei City, Taiwan

Received 26 November 2021; received in revised form 26 February 2022; accepted 27 February 2022

DOI: <https://doi.org/10.46604/peti.2022.8984>

## **Abstract**

Traditional swipe card attendance systems for hospitals cannot effectively protect employees' personal information and ensure that the employees are swiping their own cards. To solve the problem, the present study proposes a novel hospital swipe card attendance system using an artificial intelligence (AI) face modeling system with an open-source face database. The proposed system employs a multi-task cascaded convolutional network (MTCNN) algorithm and FaceNet to improve the performance of face recognition. The system can compare the face of the one who swipes a card with the faces of cardholders in the database, thereby preventing the one from clocking in on behalf of others. The results show that the application of AI technology in the hospital swipe card attendance system can realize the promise of protecting employees' personal information and verifying employees' identities.

**Keywords:** face recognition, AI face recognition technology, swipe card attendance system, personal information

## **1. Introduction**

Attendance systems have been widely used in the world to record when employees start and stop their work. From basic attendance systems using paper time cards to advanced attendance systems using chips or swipe cards, today's systems can successfully keep track of the attendance of employees. However, most of the systems cannot ensure that the person currently clocking in is the cardholder. Since the identity of the person clocking in still needs to be checked, the cost of manpower is inevitable.

The new-generation fingerprint-based biometric attendance systems can identify who is clocking in by recognizing human biological characteristics. The systems can verify users' identity when users place their fingers or palms on the readers of systems. They are convenient and can prevent one from clocking in on behalf of others. However, the use of fingerprint-based biometric attendance systems may lead to the risk of getting infectious diseases such as COVID-19. To avoid the risk, face recognition-based biometric attendance systems are the alternatives.

The earliest face recognition technology was realized by manually marking facial features (e.g., eyes and mouth) and using a computer for calculation [1-2]. Today's face recognition technology is realized based on artificial intelligence (AI) by collecting a huge number of face photos. Face recognition technology has been widely used and proposed in recent years. The face is important personal information of an individual. If the individual is in a country with personal property protection laws and has not authorized others to record his/her face, then the data of his/her face cannot be collected and aggregated for big data.

---

\* Corresponding author. E-mail address: [yslin@mail.mcut.edu.tw](mailto:yslin@mail.mcut.edu.tw)

Tel.: +886-2-29089899; Fax: +886-2-29084533

In face recognition, face detection and alignment are two fundamental tasks [3]. Mo et al. [3] disclosed that scholars tend to overlook the correlation between face detection and face alignment, and design the hardwired accelerators which produce extra communication and region overhead and cannot achieve energy-efficient acceleration. To remedy this shortcoming, Mo et al. [3] proposed an accelerator based on a multi-task cascaded convolutional network (MTCNN) algorithm to improve acceleration performance by considering face detection and multi-face alignment. MTCNN was proposed by Zhang et al. [4] in 2016. Compared with other similar methods, their proposed face recognition method based on MTCNN has the following several advantages: (1) decreasing multiply-accumulate (MAC) operations and memory access (22.8% and 24.8%); (2) improving the intersection of joint computations and speeding the hardware inference sorting process (16.0%); (3) decreasing memory capacity (38.3%) and achieving similar resolution, same throughput, and high pipeline utilization with only about half multipliers; (4) improving the hardware utilization of the fully connected layer (16.7%).

Recently, Gu et al. [5] have applied MTCNN to improve face recognition performance in a classroom. When using face detection algorithms based on CNN to detect faces in the classroom, Gu et al. [5] have noted that the face frame is not easily identified due to several problems including occlusions, large pose changes, and small-scale faces. They have thus validated their proposed face recognition method using the improved MTCNN algorithm, which can achieve better precision and efficiency than certain leading-edge approaches. To have good precision and efficiency, the hospital swipe card attendance system proposed in this study is therefore also integrated with the MTCNN algorithm.

This study aims to develop a face recognition-based hospital swipe card attendance system that can record the attendance and verify the identity of employees at the same time. With the proposed system, when an employee swipes a card, the employees' card number is used as the search condition, and the system automatically retrieves related data (i.e., the facial feature, Euclidean distance, and the threshold value) from the human resource (HR) database. The system will then identify the person who has swiped the card. Thus, integrating the function of human identification into the hospital swipe card attendance system can prevent the cases of swiping cards on behalf of others and achieve the purpose of privacy protection. The rest of this article contains the following sections: literature review of face recognition and MTCNN, proposed system architecture, system implementation, simulation process and result of the proposed system, and conclusions.

## **2. Literature Review**

### *2.1. Face recognition*

In the 1960s, face recognition was proposed by Woody Bledsoe, Helen Chan Wolf, and Charles Bisson and was realized by manually marking facial features (e.g., eyes and mouth) and using a computer for calculation [1-2]. Due to the technical limitations at that time, no major breakthroughs have been made. In the 1970s, Goldstein, Harmon, and Lesk increased the number of the marked features to 21 items, including hair color, lip thickness, etc. Although the accuracy was therefore improved, face recognition was still realized by manually marking facial features, which was very time-consuming and labor-intensive.

In 1988, Sirovich and Kirby introduced linear algebra to improve face recognition performance. In 1991, Turk and Pentland realized how to detect faces in pictures to perform the work of Sirovich and Kirby. Their work is one of the earliest automatic face recognition. Later in the 1990s, the Defense Advanced Research Projects Agency (DARPA) and the National Institute of Standards and Technology (NIST) launched into the facial recognition technology (FERET) program, supplying 2413 face data of 856 people to encourage the use of commercial facial recognition. About ten years later, NIST began to conduct face recognition vendor tests (FRVT). Based on FERET, NIST provided the United States government with an evaluation of available commercial face recognition systems and technologies to determine if there were better face recognition technologies.

In 2006, the Face Recognition Grand Challenge (FRGC) started to promote and develop the existing face recognition system in the United States. Later, in 2010, Facebook began to introduce the face recognition function. Google proposed FaceNet in 2015 and obtained 99.63% accuracy using the database “labeled faces in the wild (LFW)”. In 2016, Zhang et al. [4] proposed MTCNN for face detection and image pre-processing.

2.2. Multi-task cascaded convolutional networks (MTCNN)

MTCNN was proposed by Zhang et al. [4] in 2016. It can be applied to face detection. With MTCNN, face recognition can achieve a good accuracy level and fast execution speed [5]. MTCNN uses the cascaded convolutional network structure to complete face detection and face alignment at the same time, thereby outputting the location of eyes, nose, and mouth in the face box.

3. Proposed System Architecture

Before the proposed face recognition system is developed, MTCNN is used to realize face detection. The flowchart of face recognition based on MTCNN is shown in Fig. 1. First, the image pyramid is used to adjust the size of photos so that the photos can be used as model inputs. The second step is to input the photos to P-Net. P-Net is a fully convolutional network (FCN), and its output is a 32-dimensional feature vector, finding out whether the input possibly contains faces. If so, then the next thing to do is to identify the face candidate frames that match the face area with the original face photo through the operation of non-maximum suppression (NMS). The third step is the operation of R-Net. R-Net is a convolutional neural network (CNN), which uses bilinear interpolation to expand the output of P-Net to  $24 \times 24$  and check whether the output of P-Net contains faces. If there is a face, R-Net finds the candidate frame and uses NMS to filter it, using the same way as the operation of P-Net. The fourth step is the operation of O-Net, which is a CNN. O-Net uses bilinear interpolation to expand the output of R-Net to  $48 \times 48$  as the input for the final face detection and the key point extraction.

FaceNet is used to realize the proposed AI face recognition system [6]. It is a face recognition system proposed by Google in 2015. Since FaceNet uses an algorithm with the feature of easy calculation and available application, it has been widely applied to face verification and recognition. The architecture of the FaceNet model is shown in Fig. 2. FaceNet uses a CNN to map the face photos to the Euclidean space [7]. It employs triplet loss to compare two faces’ similarities in terms of the distance in the Euclidean space. That is, the more similar the two faces in the photos, the smaller the distance. On the contrary, the more dissimilar the two faces in the photos, the greater the distance.

As shown in Fig. 3, the principle of triplet loss is that the operation of feature training is firstly conducted to select the wrong face that is most similar to the original sample (the smallest difference) and the correct face that is most dissimilar to the original sample (the largest difference) at the same time. Next, iterative training is performed until the error is minimized, thereby improving the definition of facial features in the model to achieve the best recognition results.

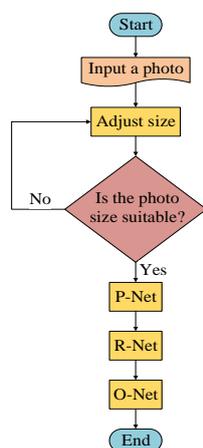


Fig. 1 The operation flowchart of MTCNN



Fig. 2 The framework of the FaceNet model

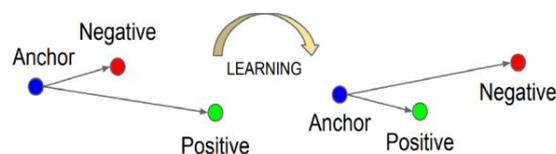


Fig. 3 The operation process of triplet loss

The operation process of triplet loss is to map the face photos to the Euclidean space. The purpose of the operation is to make the dissimilar reference faces  $X_i^a$ (Anchor) and  $X_i^n$ (Negative) far away and make the similar reference faces  $X_i^a$ (Anchor) and  $X_i^p$ (Positive) close to each other in the Euclidean space. The mathematical expression of the relationship can be expressed as:

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2, \forall (x_i^a, x_i^p, x_i^n) \in T \quad (1)$$

The loss function can be obtained by:

$$\sum_i^N \left[ \|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+ \quad (2)$$

The FaceNet architecture is different from other face recognition architectures that require classifiers. FaceNet directly performs spatial mapping, and the mapped dimension is only 128 dimensions. By using the triplet loss, FaceNet does not require a classifier. The accuracy of FaceNet is tested based on the academic data set, achieving an accuracy of  $99.63\% \pm 0.09$  and  $95.12\% \pm 0.39$  by using the LFW database and YouTube Faces database, respectively. The classification test is carried out using the face photo data set of FaceNet, and the results are shown in Fig. 4.



Fig. 4 Results of the classification test using the face photo data set of FaceNet

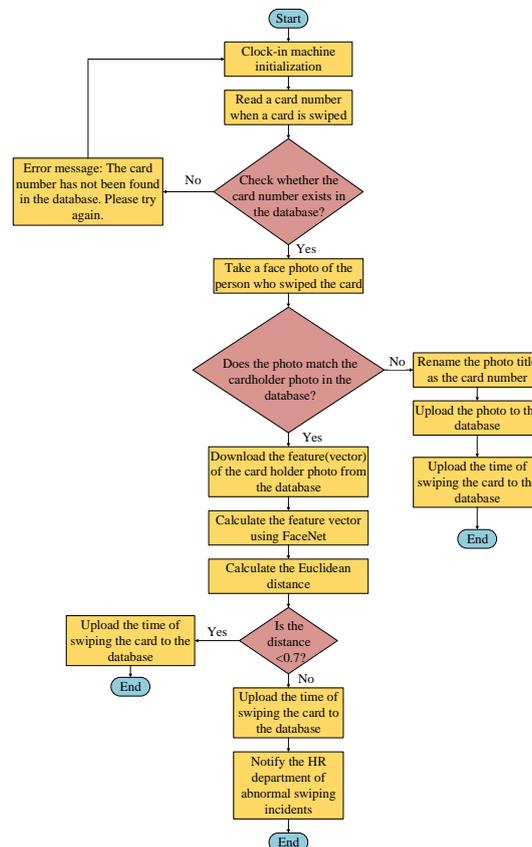


Fig. 5 The proposed swipe card attendance system

Fig. 5 shows the de-capitalization process of the proposed swipe card attendance system [8]. The system is operated based on MTCNN, FaceNet, and face recognition. The system can only receive cardholders' card numbers and personal information from the radio frequency identification (RFID) cards in the existing attendance system. It is possible to prevent the occurrence of clock-in incidents. When one swipes a card on the clock-in machine of the proposed system, the machine reads the information of clock-in, and the cardholder's name and card number recorded in the RFID card are used as the basis to search the database. That is, the proposed system reads the RFID card to obtain the card number and use it as a search condition to query the corresponding cardholder's photo in the database. Because the information of card numbers is available only in searching the database, personal information can be protected. The system can compare the face of the one who swipes the card with the faces of cardholders in the database, thereby preventing the one from clocking in on behalf of others.

#### 4. System Implementation

##### 4.1. Clock-in machine

After the proposed swipe card attendance system is started, the clock-in machine is operated. The operation process of the machine is shown in Fig. 6 [9]. The RFID reader is in the detection state. When an RFID card approaches the RFID reader, the RFID reader starts reading the card's number.

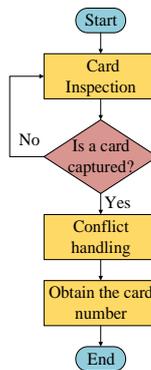


Fig. 6 The operation flowchart of RFID swiping machine

##### 4.2. Face recognition

MTCNN is a popular CNN-based detector for face recognition [5, 10-12]. As shown in Fig. 7, the proposed system uses MTCNN for face detection and face cropping. It resets the size of face photos to  $160 \times 160$  and removes redundant impurities in the photos after whitening. Then, the photos are input to FaceNet to get their feature vectors. The Euclidean distance between the feature vectors of the photos is calculated and face recognition is performed according to the set threshold.

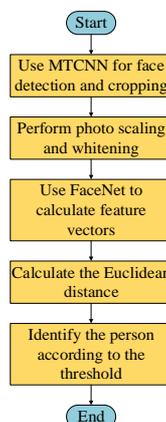


Fig. 7 The operation flowchart of face recognition

## 5. Simulation Process and Result of the Proposed System

### 5.1. The participants involved in the simulation testing of the proposed system

There are two participants involved in the simulation testing of this study. One participant is labeled as NAF09. Another is labeled as AA015. 5 photos for the participant NAF09 are collected in the dataset. 7 photos for the participant AA015 are collected in the dataset. A total of 12 photos are in the data set.

### 5.2. The hardware used in the simulation testing of the proposed system

Table 1 shows the specification of system hardware used in this study. The central processing unit (CPU) for the proposed system is Intel®Core™ i5-10600K @4.1GHz. The graphics processing unit (GPU) is NVIDIA GeForce GTX 1650. The system has a capacity of 16 GB memory. The software for developing the system is Python version 3.8.8 with CUDA (version 11.2) and TensorFlow (version 2.6.0).

Table 1 Specification of system hardware

Central processing unit (CPU)	Intel® Core™ i5-10600K @4.1GHz
Graphics processing unit (GPU)	NVIDIA GeForce GTX 1650
Computer memory	16GB-3000MHz
Python version	3.8.8
CUDA version	11.2
TensorFlow version	2.6.0

### 5.3. The simulation method

Fig. 8(c) shows the simulation process of the proposed system. Step 1 is to load all photos of a participant. Step 2 is to choose one photo as the reference photo and use the rest as the comparison photos. Step 3 is to calculate the Euclidean distance between the reference photo and each comparison photo in sequence and use the Euclidean distance 0.7 as the identification criterion. If the Euclidean distance is less than or equal to 0.7, then the photo is identified as “the photo of the participant”. That is, the identification result is “correct”. Otherwise, the photo is identified as “not the photo of the participant”. That is, the identification result is “wrong”.

Step 4 is to count the number of correct identification results and divide it by the number of comparison photos to obtain the success rate of the reference photo identification. Step 5 is to choose another photo as the reference photo and repeat steps 3 and 4 until all the photos have been tested as reference photos. Step 6 is to load all photos of the other participant and repeat steps 2 to 5.

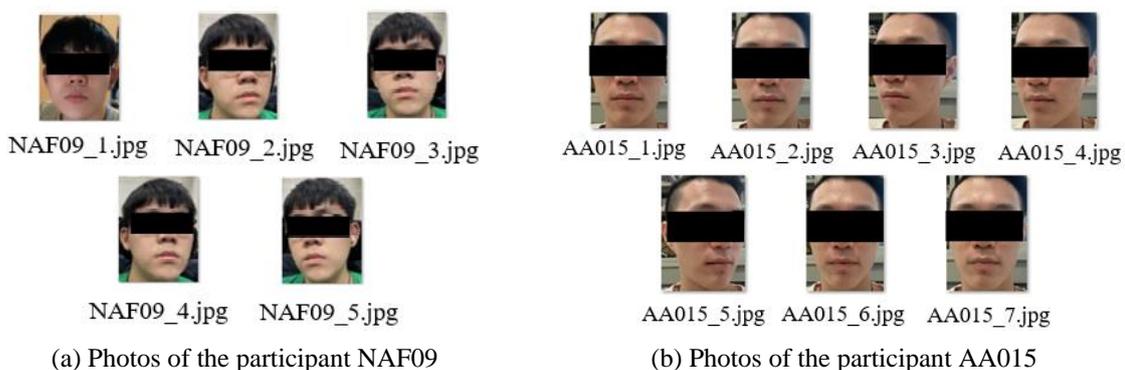
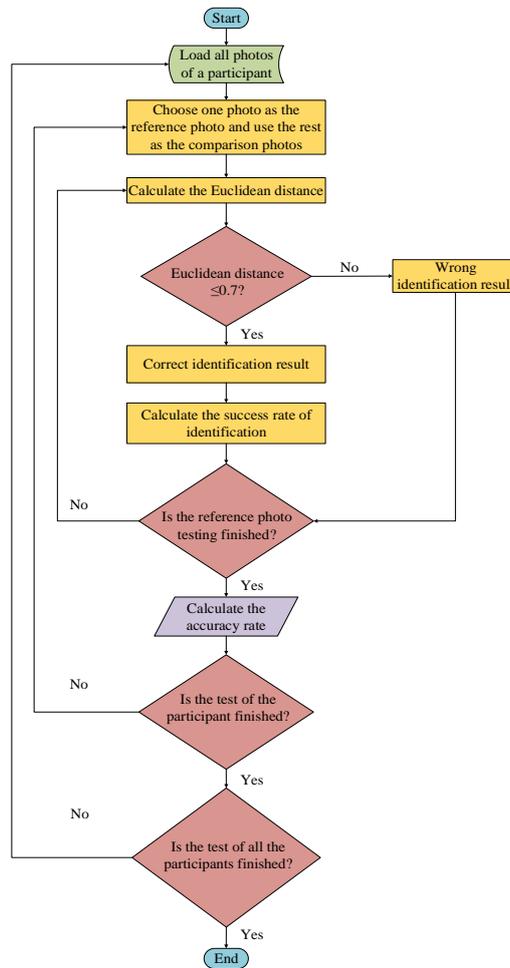
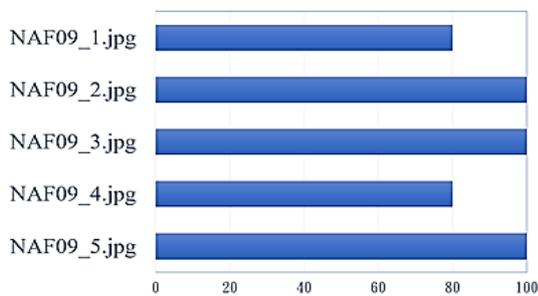


Fig. 8 The simulation of the proposed swipe card attendance system

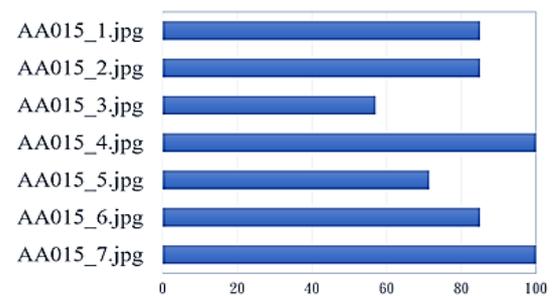


(c) The flowchart of the simulation process

Fig. 8 The simulation of the proposed swipe card attendance system (continued)



(a) Success rate of face recognition using NAF09's photos (minimum: 80%, maximum: 100%)



(b) Success rate of face recognition using AA105's photos (minimum: 57%, maximum: 100%)

Fig. 9 Success rate of face recognition using photos of each participant

As shown in Fig. 9, the success rate of face recognition using AA015's photos is not satisfactory because the minimum is 57%. Compared with the success rate using AA015's photos, the success rate using NAF09's photos is better because the minimum is 80% and the maximum is 100%. Overall, such results may imply that more samples are necessary to enhance the recognition accuracy.

#### 5.4. The simulation result of the proposed system

Each participant has three or more photos with different angles for facial identification in the database. During the simulation, each participant uses the other participant's RFID card to test the function of the proposed system. If the actual result shows "True", it means that the person is swiping the card on behalf of the other. If the result shows "False", the person who swiped the card is the cardholder. The simulation result is shown in Table 2.

Table 2 The simulation result of the proposed system

The face photo of the person who swipes a card	The face photo retrieved from the database	Euclidean distance	Actual result
NAF09_1.jpg	AA015_2.jpg	1.1176	True
NAF09_1.jpg	AA015_3.jpg	1.1005	True
NAF09_1.jpg	AA015_4.jpg	1.2087	True
AA015_1.jpg	NAF09_2.jpg	1.1015	True
AA015_1.jpg	NAF09_3.jpg	1.0909	True
AA015_1.jpg	NAF09_4.jpg	1.1197	True

## 6. Conclusions

The present study develops a swipe card attendance system with a face recognition function for hospitals. The test results verify that the system can identify the event where a card is swiped by a non-cardholder. When this case happens, the system will send an email to notify the HR department, which effectively reduces the labor cost of the HR department for attendance checking.

The current decision-making method only uses Euclidean distance 0.7 as the criterion, which is slightly rough. In the future, false acceptance rate (FAR) [13], false rejection rate (FRR) [14], and equal error rate (EER) [15] will be added to the experiment to strengthen the decision-making process of face recognition. Also, infrared scanning [16] will be added to increase the features. Moreover, with only two participants, the present study may produce the issue of false positives or false negatives in the identification process, thus decreasing the accuracy of face recognition. Further research should recruit more participants to join the research to improve the rigorousness of the research method.

This study just focuses on one face in one frame. For those cases with several faces in one frame and with low-resolution photos, further research can employ multi-object face recognition using local binary pattern histogram to improve the success rate of face recognition [17]. It should be noted that, face recognition not only can be used in hospital swipe card attendance systems but also can be employed in monitoring the breath function of patients [18]. Future research can integrate face recognition with thermal images to identify the patients and assess their health conditions.

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

- [1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Massachusetts: MIT Press, 2016.
- [2] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*, New York: Springer Science and Business Media, 2005.
- [3] H. Mo, L. Liu, W. Zhu, Q. Li, H. Liu, S. Yin, et al., "A Multi-Task Hardwired Accelerator for Face Detection and Alignment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 4284-4298, November 2020.
- [4] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499-1503, August 2016.
- [5] M. Gu, X. Liu, and J. Feng, "Classroom Face Detection Algorithm Based on Improved MTCNN," *Signal, Image, and Video Processing*, in press.
- [6] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815-823, June 2015.
- [7] J. Hearty, *Advanced Machine Learning with Python*, Birmingham: Packt Publishing, 2016.
- [8] H. Yang and X. Han, "Face Recognition Attendance System Based on Real-Time Video Processing," *IEEE Access*, vol. 8, pp. 159143-159150, July 2020.

- [9] Q. Miao, F. Xiao, H. Huang, L. Sun, and R. Wang, "Smart Attendance System Based on Frequency Distribution Algorithm with Passive RFID Tags," *Tsinghua Science and Technology*, vol. 25, no. 2, pp. 217-226, April 2020.
- [10] N. Zhou, R. Liang, and W. Shi, "A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection," *IEEE Access*, vol. 9, pp. 5573-5584, January 2021.
- [11] B. Yu and D. Tao, "Anchor Cascade for Efficient Face Detection," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2490-2501, May 2019.
- [12] X. Li, Z. Yang, and H. Wu, "Face Detection Based on Receptive Field Enhanced Multi-Task Cascaded Convolutional Neural Networks," *IEEE Access*, vol. 8, pp. 174922-174930, October 2020.
- [13] S. Saito, Y. Tomioka, and H. Kitazawa, "A Theoretical Framework for Estimating False Acceptance Rate of PRNU-Based Camera Identification," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2026-2035, September 2017.
- [14] N. Merhav, "False-Accept/False-Reject Trade-Offs for Ensembles of Biometric Authentication Systems," *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 4997-5006, August 2019.
- [15] D. H. Kaye, "The Error of Equal Error Rates," *Law, Probability, and Risk*, vol. 1, no. 1, pp. 3-8, July 2002.
- [16] R. He, J. Cao, L. Song, Z. Sun, and T. Tan, "Adversarial Cross-Spectral Face Completion for NIR-VIS Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 5, pp. 1025-1037, May 2020.
- [17] R. R. Isnanto, A. F. Rochim, D. Eridani, and G. D. Cahyono, "Multi-Object Face Recognition Using Local Binary Pattern Histogram and Haar Cascade Classifier on Low-Resolution Images," *International Journal of Engineering and Technology Innovation*, vol. 11, no. 1, pp. 45-58, January 2021.
- [18] H. Hendrick, A. Aripriharta, S. K. Chen, M. H. Tung, T. C. Chiang, and G. J. Jong, "Nostril in RGB Imaginary by Using NI Vision LabVIEW," *Proceedings of Engineering and Technology Innovation*, vol. 4, pp. 37-39, October 2016.



Copyright© by the authors. Licensee TAETI, Taiwan. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).