Sensors and Materials, Vol. 38, No. 8 (2026) 4415–4438
MYU Tokyo

S & M 4574

A Blurred Face Recognition System for Enhanced Surveillance: Integrating Automated Detection and Identification with Existing Infrastructure

Yu-Shiuan Tsai,[1] Quan-Bin Zhang,[2] Po-Yang Chi,[2] Shang-Ze Lin,[2]
Chien-Hsing Chou,[3] De-Yuan Huang,[4] and Yi-Zeng Hsieh[2*]

[1]Department of Computer Science and Engineering, National Taiwan Ocean University
No. 2, Beining Rd., Jhongjheng District, Keelung City 202301, Taiwan
[2]Department of Electrical Engineering, National Taiwan University of Science and Technology
No. 43, Keelung Rd., Sec. 4, Daan Dist., Taipei City 106335, Taiwan
[3]Department of Mathematics and Information Education, National Taipei University of Education
No. 134, Sec. 2, Heping E. Rd., Daan District, Taipei City 106, Taiwan
[4]Department of Computer Science and Engineering, Tatung University, Taipei City, Taiwan

(Received September 12, 2024; accepted October 9, 2025)

Keywords: facial recognition, blurred face detection, surveillance systems, security, privacy, artificial intelligence, image processing, facial detection algorithm, automated monitoring

In this study, we present a facial recognition system that integrates blurred face recognition with existing surveillance systems, enhancing security without the need for upgrading older cameras. The system automatically records key data such as timestamps of suspect appearances, significantly reducing the time required to review footage and track down suspects. The contributions of this research are as follows. (1) The system provides enhanced security by utilizing blurred face recognition without requiring costly hardware upgrades. (2) It enables the automated tracking and time-recording of suspects, reducing manual effort in surveillance review. (3) The system's recognition model maintains high accuracy despite challenges in image quality and environmental factors. (4) The study highlights potential privacy concerns and legal considerations related to facial data storage, encouraging careful regulatory compliance. Experimental results show that the system maintains a high recognition rate within a 5.5 m range, with up to 100% accuracy under ideal conditions. However, its performance declines when the distance increases beyond 6 m, and it struggles with mask-wearing and extreme head angles. Future improvements to the face detection algorithm will increase detection rates and further enhance the system's efficiency. By incorporating better visualization tools for suspect identification, the system can enable faster and more accurate crime investigations.

1. Introduction

Facial recognition technology has seen rapid advancements in recent years, driven by the growing demand for convenience in modern life. Its applications have expanded across various

*Corresponding author: e-mail: yzhsieh@mail.ntust.edu.tw
https://doi.org/10.18494/SAM5365

ISSN 0914-4935
https://myukk.org/

domains, including financial services, smart home systems, retail, and public security. In the realm of identity verification, where traditional methods often rely on complex passwords or physical identification documents, facial recognition now offers a more seamless and efficient alternative. The unique biometric characteristics of an individual's face enable streamlined access to both physical and virtual spaces, underscoring the significant role this technology plays in enhancing the convenience of daily life. However, given its widespread use in sensitive applications such as security control and financial privacy, the need for precise and reliable facial recognition systems is paramount. Factors such as facial occlusions, makeup, masks, poor lighting, and low-quality camera angles present challenges to traditional facial recognition methods, highlighting the need for advancements in blurred face recognition technology.

In parallel, the security of educational institutions has become a growing concern, particularly in open campuses where access is not tightly regulated. The lack of strict controls has led to an increase in criminal activities such as theft and unauthorized photography. Current surveillance practices, reliant on the manual monitoring of footage, are not only inefficient but often ineffective owing to the limitations of outdated, low-resolution cameras. As such, identifying suspects in a timely and accurate manner remains a critical issue. Upgrading all surveillance systems across a campus is typically cost-prohibitive and does not address the fundamental inefficiencies of manual video analysis.

To overcome these challenges, in this study, we explore the application of AI in enhancing campus security through automated suspect identification. By leveraging AI, we aim to accurately detect and recognize individuals in real time, reducing the burden of manual surveillance while improving the efficiency of the identification process. The system captures and segments facial images from video streams, which are then compared with a database of known individuals. Identified individuals are either displayed with their corresponding names if they are part of the campus community or marked as unknown and logged for future reference if they are unfamiliar. This approach has the potential to not only streamline suspect identification but also improve overall campus security.

The system's ability to handle blurred facial images—common in surveillance footage—presents an additional challenge. To address this, we build upon two well-established models: AdaFace, known for its quality-adaptive margin in facial recognition, and Multi-task Cascaded Convolutional Network (MTCNN),(1) a robust multitask convolutional network for face detection and alignment. AdaFace's marginal-based loss function demonstrates superior performance in recognizing blurred faces, while MTCNN effectively segments faces from video streams. By combining these models, we aim to enhance the accuracy of facial recognition in low-quality images, a critical need for campus surveillance.

In summary, we propose an AI-driven system that integrates blurred face recognition to address the growing security concerns in educational institutions. By automating the detection and recognition processes and improving the handling of low-resolution images, we aim to enhance both the effectiveness and efficiency of campus security measures, ultimately contributing to a safer learning environment for students and staff alike. The contributions of this study are as follows.

- Integration of Blurred Face Recognition into Existing Surveillance Systems: In this paper, we present a novel facial recognition system that can effectively recognize blurred faces, thus enhancing security without the need for upgrading the existing surveillance camera hardware. This makes it a cost-efficient solution for improving safety in environments such as campuses.

- Automated Suspect Identification and Tracking: The system automates the tracking and recording of suspect appearances, significantly reducing the time and labor traditionally required for manual surveillance. It automatically logs key data such as timestamps, helping streamline the investigation in security incidents.
- High Accuracy in Challenging Environmental Conditions: Despite factors such as low image quality, poor lighting, or occluded facial features (e.g., masks or makeup), the system maintains a high recognition rate within a range of 5.5 m, with up to 100% accuracy under ideal conditions. This robustness is achieved by leveraging the AdaFace and MTCNN models.
- Privacy Considerations and Legal Implications: The research highlights the legal and ethical concerns associated with storing facial recognition data. By addressing privacy issues, it underscores the importance of complying with regulations when implementing such systems, particularly in public or semipublic spaces such as schools.
- Potential for Future System Enhancements: We suggest several areas for future work, such as improving the face detection algorithm to enhance the system's detection rate and implementing better visualization tools for the quicker identification of suspects during footage reviews. This forward-looking approach ensures that the system will continue to evolve in terms of accuracy and usability.

This paper is organized into the following five sections: In Sect. 1 we provide a brief overview of campus security issues in Taiwan, along with the motivation and objectives of this study. In Sect. 2, we introduce the principles of convolutional neural networks (CNNs) in facial recognition, followed by an exploration of the functionality and application methods of the MTCNN and AdaFace models. In Sect. 3, we explain in detail the facial detection and recognition algorithms employed in this study. In Sect. 4, we describe the program's operational flow, including the practical implementation and performance evaluation of the multiface recognition system developed in this study. The performance results of the model are then discussed in relation to varying levels of facial feature occlusion. In Sect. 5, we summarize the research findings, along with future development directions and important considerations for future work.

2. Related Works

2.1 CNNs for image recognition

With the continuous advancement of technology, the computational power and capacity of computers have significantly improved, allowing more tasks related to computation and data interpretation to be managed by machines. As AI continues to evolve, numerous machine learning algorithms have been developed, enabling AI to take a central role in automated data interpretation and analysis.

In the 1950s and 1960s, neurophysiologists Hubel and Wiesel[2] made a key discovery of two types of nerve cell, which they termed simple cells and complex cells.[3] Simple cells transmit localized trigger information to complex cells, which are responsible for large-scale neural perception. This discovery inspired the later development of CNNs. Convolution, a mathematical operation involving matrix multiplication, enables the extraction of features such as image sharpening and edge detection, which are crucial for image analysis. CNNs have demonstrated exceptional performance in tasks related to both image and speech analyses.

2.2 CNN computational architecture

CNNs operate by transforming input images into tensors, which are then processed through layers of convolution and pooling to extract key features. After several convolutional layers, the extracted information is passed into a fully connected network, which is responsible for performing tasks such as classification.

2.3 Margin-based loss function in facial recognition

A loss function plays a critical role in measuring the difference between AI predictions and actual outcomes, which helps evaluate the effectiveness of model training. The purpose of minimizing the loss function is to improve the accuracy of the AI model.[4,5] Margin-based loss functions, specifically designed for tasks such as facial recognition, have been shown to enhance model performance by focusing on the most challenging samples.

2.4 Cross-entropy loss

Cross-entropy loss is one of the loss functions commonly used in machine learning classification problems. Entropy measures the uncertainty of information, a concept introduced in Refs. 4 and 5. Cross-entropy $H(p, q)$, as shown in Eqs. (1) and (2), is a measure of the difference between the actual probability distribution $p(x_i)$ and the predicted probability distribution $q(x_i)$. Higher $H(p, q)$ indicates greater prediction uncertainty. Minimizing cross-entropy improves the model performance.

$$H = -\sum_{i} p(x)\log p(x) \tag{1}$$

$$H(p,q) = -\sum_{i=1}^{n} p(x_i)\log\left(q(x_i)\right) \tag{2}$$

2.5 Triplet loss

Triplet loss, as introduced in Ref. 6, is used for classification tasks such as facial recognition by determining the similarity between different images. The training process involves three types of sample: Anchor, Positive, and Negative. The goal is to minimize the distance between the Anchor and Positive samples, while maximizing the distance between the Anchor and Negative samples. A margin is enforced to ensure sufficient separation between the Positive and Negative samples.(6)

$$\mathcal{L} = \sum_{i=1}^{N}\left[\left\|f_i^a - f_i^p\right\|_2^2 - \left\|f_i^a - f_i^n\right\|_2^2 + \alpha\right] \tag{3}$$

2.6 Margin loss

Instead of classifying data into predefined categories, it evaluates the similarity or distance between two distinct inputs, making it ideal for verification tasks in Ref. 7. The margin-based loss function, a type of ranking loss, was introduced in Ref. 8. Unlike triplet loss, margin loss emphasizes ensuring that positive samples maintain a specific distance and correct ranking, thereby enhancing system robustness.

2.7 Contractive loss, as shown in Eq. (4), was proposed in Ref. 9.

This loss function involves two independent variables. During training, input data pairs are fed into a Siamese Model.(7–9) If the two images belong to the same class, the distance between their corresponding vectors is minimized; conversely, if they belong to different classes, the distance is maximized.

$$\mathcal{L}\left(W, Y, \vec{X}_1, \vec{X}_2\right) = \left(1 - Y\right)\frac{1}{2}\left(Dw\right)^2 + \frac{Y}{2}\left\{\max\left(0, m - D_w\right)\right\}^2 \tag{4}$$

2.8 Challenges in recognizing low-resolution (LR) images

Facial recognition is a widely researched field within AI, with numerous studies focusing on deep learning applications in this domain. While recognizing clear, high-resolution (HR) facial images is now well established, surveillance environments often produce blurry images with occluded features caused by, for example, clothing. Recent research has shifted towards AI models capable of effectively recognizing blurred faces and the development of LR image datasets for testing.(10)

2.9 Development of LR image datasets

Significant progress has been made in facial recognition using high-resolution datasets. However, several factors—such as lighting conditions, distance, pose, and motion blur—can result in unclear images, making recognition more challenging. Research into low-resolution image recognition has become increasingly essential, and datasets play a critical role in model training. The TinyFace dataset,(11) which consists of 169403 facial images with 5139 labels, provides a valuable resource for advancing LR facial recognition, particularly since the images are captured in uncontrolled environments.

In recent years, wearing masks has become a common practice owing to the COVID-19 global pandemic. Consequently, mask-related datasets are crucial for improving facial recognition systems. The SF-MASK dataset,(12–14) which combines the Real World Masked Face Recognition Dataset (RMFRD), Medical Mask Dataset (MMD), and Face Mask Label Dataset (FMLD), includes approximately 20,000 low-resolution masked facial images, many of which feature top-down views to simulate surveillance camera scenarios.

2.10 Deep learning for facial recognition

The development of facial recognition can be traced back to the 1960s when the first facial recognition system was created by determining facial feature coordinates. Since then, many research groups have aimed to enhance facial recognition models by addressing challenges such as head rotation, lighting, facial expressions, and aging.(15–17)

2.11 SphereFace

SphereFace improves recognition accuracy through the use of a novel loss function, as shown in Eq. (5). A-softmax transforms Euclidean space into angular space, which enhances the angular differences between classes while reducing the angular distance within the same class. This results in improved classification precision.(18–20)

$$L_{ang} = \frac{1}{N}\sum_i -\log\left(\frac{e^{\|x_i\|\Psi(\theta_{yi},i)}}{e^{\|x_i\|\Psi(\theta_{yi},i)} + \sum_{j\neq y_i} e^{\|x_i\|\cos(\theta_j,i)}}\right) \tag{5}$$

Here, $\psi(\theta_{yi},i) = (-1)^k \cos(m\theta_{yi},i) - 2k$.

2.12 CosFace

CosFace introduces the Large Margin Cosine Loss to enhance classification precision, as shown in Eq. (6).(21)

$$L_{lmc} = \frac{1}{N}\sum_i -\log\left(\frac{e^{s(\cos(\theta_{yi},i)-m)}}{e^{s(\cos(\theta_{yi},i)-m)} + \sum_{j\neq y_i} e^{s(\cos(\theta_j,i))}}\right) \tag{6}$$

2.13 ArcFace

Introduced in 2018, ArcFace(22) enhances classification precision by adding the parameter m to the loss function's θ term, as shown in Eqs. (7) and (8). However, if the dataset is very noisy, the model's performance can degrade. To address this, Subcenter ArcFace, which divides positive samples into multiple subsamples, was proposed.[23]

$$\mathcal{L} = -\log \frac{e^{s\cos\left(\theta_{yi}+m\right)}}{e^{s\cos\left(\theta_{yi}+m\right)} + \sum_{j=1, j\neq y_i}^{n} e^{s\cos\theta_j}} \tag{7}$$

$$\mathcal{L} = -\log \frac{e^{s\cos\left(\theta_{yi}+m\right)}}{e^{s\cos\left(\theta_{yi}+m\right)} + \sum_{j=1, j\neq y_i}^{n} e^{s\cos\theta_j}} \tag{8}$$

Here, $\theta_j = \arccos\left(\max_k \left(W_{j_k}^T x_i\right)\right), k \in \{1\ldots K\}$.

2.14 Challenges in blurred face recognition

Facial recognition has a wide range of applications, but many images suffer from incomplete features because of hardware limitations or human factors. The performance of AI models on low-quality images has been explored in recent research. AdaFace, a facial recognition model published in 2022, adjusts training weights on the basis of image quality. For low-quality but recognizable images, weights are increased to improve learning. Conversely, weights are reduced for low-quality images to prevent degrading the model. Tests on low-quality datasets have shown that AdaFace performs with high accuracy.[24–26]

3. Method

3.1 MTCNN

MTCNN is the facial detection algorithm utilized in this study. Since face detection and alignment are essential preprocessing steps before facial recognition, the algorithm introduced in Ref. 30 has been adopted. MTCNN employs a three-layer CNN to detect face bounding boxes and outputs five facial landmark points. These detected face boundaries form the foundation for facial recognition in this study.[1] The MTCNN architecture is illustrated in Fig. 1.

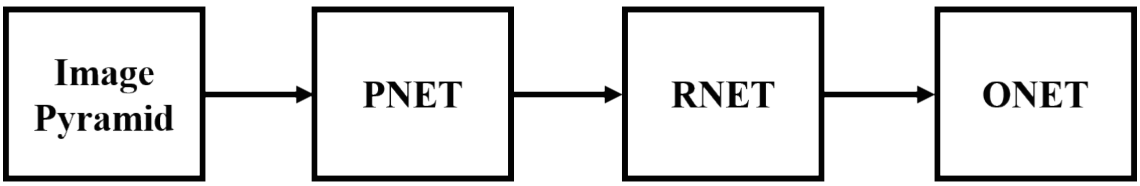


Fig. 1. MTCNN algorithm sequence.

3.2 Image pyramid

The first step in MTCNN involves scaling the input image using a Gaussian pyramid. The image is repeatedly downsampled, and the resized images are fed into subsequent CNNs for feature extraction. This process ensures that faces of various sizes can be detected accurately.

3.3 Proposal network (PNET)

Once the image has been scaled, it is processed by PNET, as depicted in Fig. 2. PNET features a simple architecture and quickly extracts face bounding boxes. However, it is prone to capturing nonfacial features, which lowers its initial accuracy. To improve precision, PNET identifies potential face bounding boxes and applies non-maximum suppression (NMS) to remove redundant boxes.

3.4 Refinement network (RNET)

The bounding boxes generated by PNET are then passed to RNET. As shown in Fig. 3, RNET has a more complex architecture than PNET, allowing it to extract more accurate face boundaries. These boundaries are further refined using NMS to enhance detection precision.

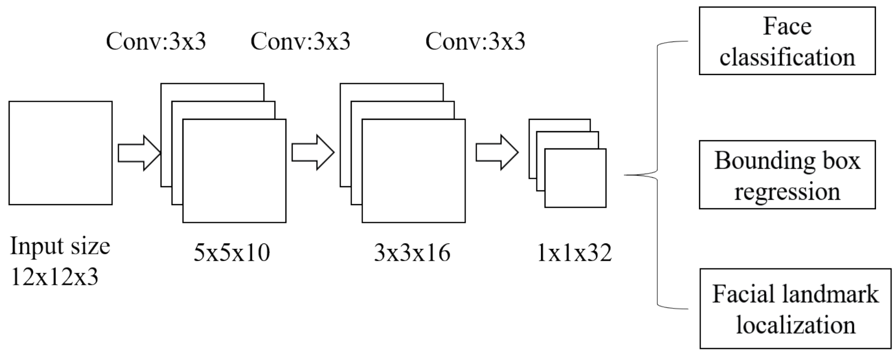


Fig. 2. PNET convolutional network architecture.

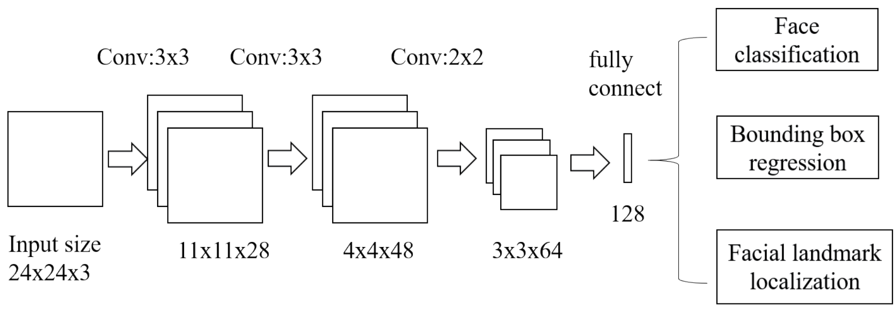


Fig. 3. RNET convolutional network architecture.

3.5 Output network (ONET)

ONET is the final layer in the MTCNN architecture. As shown in Fig. 4, ONET has the most intricate structure among the three networks, and it accurately delineates face boundaries. In addition to producing the face bounding boxes, ONET outputs five facial landmark points to aid in face alignment and recognition.

MTCNN improves the process of extracting face bounding boxes by utilizing a three-layer convolutional network architecture. During training, MTCNN employs multitask learning to simultaneously handle three tasks: face classification, bounding box regression, and facial landmark localization. Each task is optimized using a unique loss function (as summarized in Table 1), enhancing the model's overall performance and accuracy in face detection.

3.6 AdaFace algorithm

Facial recognition is a highly popular research area, with numerous studies focused on developing algorithms capable of delivering stable performance across diverse datasets. While modern facial recognition algorithms perform well with clear images, surveillance images are often blurry, and facial features are obscured by clothing, makeup, or poor angles, making recognition challenging. As such, the development of facial recognition algorithms for low-quality images is crucial.

The AdaFace approach is proposed for modifying training weights on the basis of image quality, as shown in Fig. 5. For high-resolution images with occluded facial features or low-resolution images with clear facial features, training weights are increased. Since these types of image are commonly found in surveillance footage, increasing the training weight improves the

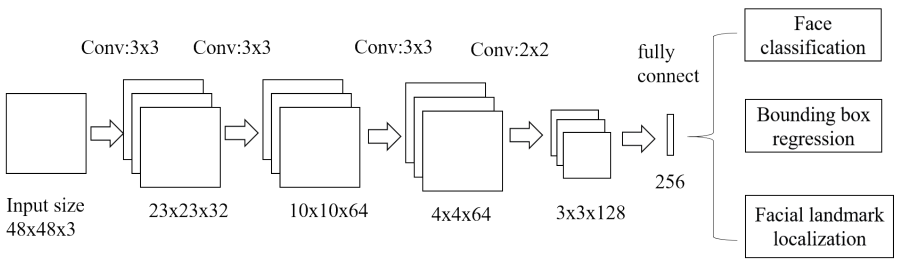


Fig. 4. ONET convolutional network architecture.

Table 1
Loss function overview.

Task	Loss function
Face classification	$\mathcal{L}_i^{det} = -\left(y_i^{det}\log\left(p_i\right) + \left(1 - y_i^{det}\right)\left(1 - \log\left(p_i\right)\right)\right)$
Bounding box regression	$\mathcal{L}_i^{det} = \left\lVert \hat{y}_i^{box} - y_i^{box} \right\rVert_2^2$
Facial landmark localization	$\mathcal{L}_i^{landmark} = \left\lVert \hat{y}_i^{landmark} - y_i^{landmark} \right\rVert_2^2$

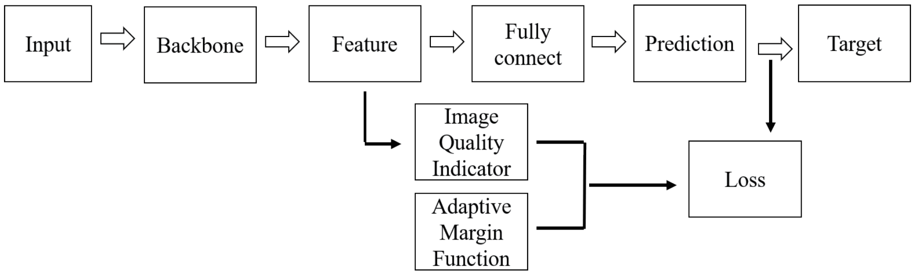


Fig. 5. AdaFace loss function design.

model accuracy on such images. Conversely, for images that are overly blurry or lack sufficient features, the model considers them unrecognizable. If such images are present during training, the model might capture nonfacial features, which can degrade its performance. Therefore, training weights are reduced for these images.

AdaFace introduces a loss function that adjusts the training weights on the basis of image quality. The loss function is designed using the margin-based loss function as shown in Eq. (9) and the adaptive margin function as shown in Eq. (10). Margin-based loss functions are better suited for data classification than cross-entropy loss functions, making them commonly used in facial recognition algorithms. By incorporating the image quality parameter $\hat{z}_i$ into the adaptive margin function, the training weight varies according to the image quality.

$$\mathcal{L} = -\log \frac{\exp\left(f\left(\theta_{yi}, m\right)\right)}{\exp\left(f\left(\theta_{yi}, m\right)\right) + \sum_{j \neq y_i}^{n} \exp\left(s \cos\theta_j\right)}, \tag{9}$$

$$f\left(\theta_j, m\right)_{AdaFace} = \begin{cases} s\left(\cos\left(\theta_j + g_{angle}\right) - g_{add}\right), & j = y_i \\ s\cos\left(\theta_j\right), & j \neq y_i \end{cases} \tag{10}$$

where $g_{angle} = -m \cdot c$, $g_{add} = m \cdot \left\|\widehat{z_i}\right\| + m$, and $\left\|\widehat{z_i}\right\| = \left[\frac{\left\|z_i\right\| - \mu_z}{\sigma_z / h}\right]_{-1}^{1}$.

To address the challenge posed by blurred images, in our method, we adopted the margin function of AdaFace based on the feature norm. In the margin function, the different margin functions can place varying emphasis on the difficulty levels of samples. Additionally, we also established that feature norm can be an effective indicator of low-quality images. These insights are now combined to propose a new loss function tailored for face recognition.

3.7 Image quality indicator of blurred image

The feature norm, $||z_i||$, being a model-dependent value, requires normalization using batch statistics μ and σ_z. Specifically, we define $\left\|\widehat{z_i}\right\|$ as shown in Eq. (11).

$$\left\|\widehat{z_i}\right\| = \left\lfloor \frac{\left\|z_i\right\| - \mu_z}{\sigma_z / h} \right\rceil_{-1}^{1}, \tag{11}$$

where μ_z and σ_z represent the mean and standard deviation of all $\|z_i\|$ within a batch, respectively, and $\lfloor\cdot\rceil$ indicates clipping the value between −1 and 1, with gradient flow being stopped. Since $\frac{\|z_i\| - \mu_z}{\sigma_z / h}$ normalizes the batch distribution of $\left\|\widehat{z_i}\right\|$ to a unit Gaussian, we clip the values between −1 and 1 to enhance handling, knowing that roughly 68% of the unit Gaussian distribution falls within this range. The term h controls concentration, and we set $h = 0.33$ for optimal results.

To further ensure stability, particularly with smaller batch sizes, we utilize exponential moving averages for μ_z and σ_z across multiple steps. Let $\mu(k)$ and $\sigma(k)$ be the k-th step batch statistics for $\|z_i\|$, as shown in Eq. (12); then,

$$\mu_z = \alpha\mu_z^{(k)} + (1-\alpha)\mu_z^{(k-1)}. \tag{12}$$

The momentum coefficient α is set to 0.99, and a similar approach is taken for α_z.

3.8 Adaptive margin function

We adopted the margin function of AdaFace to image quality as follows: (1) emphasizing hard samples when the image quality is high and (2) de-emphasizing hard samples when the image quality is low. This is achieved through two adaptive terms, g_{angle} and g_{add}, corresponding to angular and additive margins, respectively, as shown in Eq. (13).

$$f(\theta_j, m)_{AdaFace} = \begin{cases} s\cos(\theta_j + g_{angle}) - g_{add}, & j = y_i \\ s\cos\theta_j, & j \neq y_i \end{cases} \tag{13}$$

Here, g_{angle} and g_{add} are functions of $\left\|\widehat{z_i}\right\|$. When $\left\|\widehat{z_i}\right\| = -1$, the function behaves as ArcFace; when $\left\|\widehat{z_i}\right\| = 0$, it mimics CosFace; and when $\left\|\widehat{z_i}\right\| = 1$, it becomes a negative angular margin with a shift. Features with high norms receive a higher gradient scale when far from the decision boundary, whereas features with low norms receive a higher gradient scale near the decision boundary. Harder low-norm samples are de-emphasized.

There are several methods to evaluate image quality, and SER-FIQ is a deep learning algorithm used for this purpose. However, using an additional AI model to assess image quality increases computational costs. Therefore, in this paper, we proposed an alternative approach, using feature norms to represent image quality. Thus, feature norms can serve as a proxy for image quality, avoiding the complexity of additional computations. AdaFace uses MS1MV2, MS1MV3, and WebFace4M as training datasets, while the test datasets include low-quality, mixed-quality, and high-quality datasets. The results show that the AdaFace algorithm outperforms other algorithms in recognizing low- and mixed-quality datasets while maintaining stable performance on high-quality datasets. Thus, AdaFace was chosen as the facial recognition algorithm for this study.

4. Experimental Results

We simulated campus surveillance footage by mounting a network camera on a tripod. At the beginning of the program, features are extracted from the images in the folder. The video feed from the camera is processed by MTCNN in the first step, which detects faces and crops them into individual images, returning bounding boxes. These processed images are then fed into the AdaFace facial recognition model, where their similarity to existing images is compared. Finally, the system displays the bounding boxes and name information of the recognized faces on the surveillance screen.

4.1 Equipment introduction

The hardware used in this study include the computer, camera, and tripod. The details of the computer are given in Table 2. We used a Logitech C270 720P HD webcam to simulate campus surveillance cameras (Table 3).

4.2 Model training

The AdaFace model in this system uses the pretrained R50 MS1MV2 model. The training dataset used is CASIA-Webface. The final training result shows that the loss converged to approximately 3.4, as illustrated in Fig. 6.

4.3 Impact of feature occlusion on recognition results

In reality, the video captured by surveillance cameras may be unclear owing to external factors such as lighting, weather, and camera angle. Additionally, people's clothing and accessories often obscure facial features, making facial recognition more challenging. We selected four common factors to test their impact on the system, as described below.

Table 2
Laptop specifications.

Component	Specification
CPU	Intel i7-11800H
GPU	RTX-3060
RAM	16 GB

Table 3
Network camera specifications.

Specification	Detail
Maximum resolution	720p/30 fps
Camera megapixels	0.9
Focus type	Fixed focus
Diagonal field of view	55°

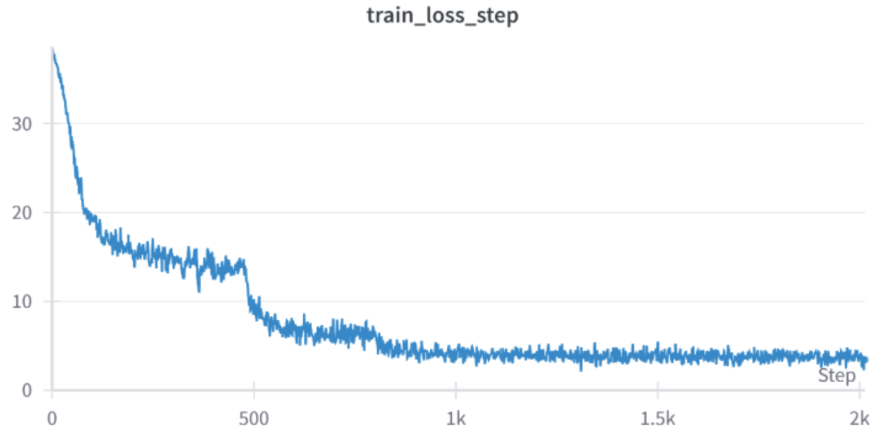


Fig. 6. (Color online) Loss value of training process.

The experiment involved two people standing within the surveillance camera's view, with clear frontal images of both individuals present in the dataset. To test model accuracy, unknown faces were added to the stranger dataset, and detection and recognition rates were recorded. Detection rate refers to the probability of detecting one person (1p) or two persons (2p) in 100 surveillance images while walking within a fixed distance. Once detected, the system cropped the faces from the images, and the recognition success rate was recorded by feeding these cropped images into the model (Tables 4 and 5).

4.4 Distance from camera

Under well-lit conditions, with faces oriented toward the tripod, we observed system performance at various distances. We measured the farthest recognition distance and simulated movement within the recognition range to evaluate the system's response, as shown in Table 6.

The experiment revealed that the system can maintain high recognition accuracy up to 5.5 m. Beyond this distance, however, performance degradation became evident, with a significant increase in the number of misidentifications observed at 6 m and beyond. At 6 m, the system misidentified Chi as a new stranger and stored the photo (Fig. 7). Subsequent recognition attempts produced similar errors, resulting in a final recognition rate of only 36%. While extending the recognition range to 8 m through dataset augmentation improved the system's operational distance, the increased error rate suggests a limitation in scalability. To address this, future enhancements should focus on refining distance-invariant feature extraction methods and integrating domain adaptation strategies to improve recognition robustness under varying image quality conditions.

4.5 Distance from camera (7.5 m)

Since the selected model emphasizes its performance in recognizing blurred faces, we attempted to enhance the system's stability and maximum recognition distance by including blurred faces (captured at 5.5 m) in the dataset, as shown in Tables 7 and 8.

Table 4
(Color online) Internal dataset—face image.

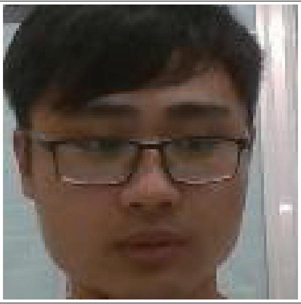
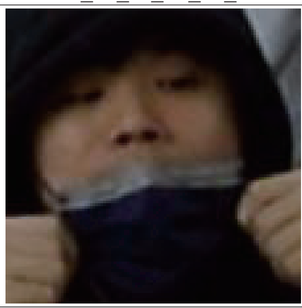
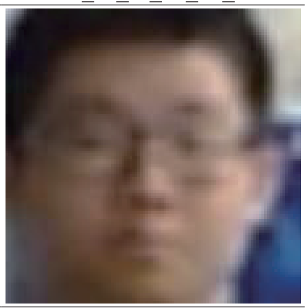
Name	Chi	Chung
Face image		

Table 5
(Color online) Stranger dataset—face photos.

Name	01_09_13_04_53_1	01_10_11_05_38_0
Images		

After incorporating blurred images into the training dataset, the system's maximum recognition distance extended to 8 m. However, despite this improvement, an increase in the number of misidentifications was observed beyond 6 m owing to reduced image resolution and increased noise. At 8.5 m, the system occasionally misidentified Chi as a stranger with occluded facial features (Table 5, left image), leading to a recognition rate of 78%, which is insufficient for reliable surveillance applications. To enhance recognition accuracy at longer distances, future improvements should involve the integration of super-resolution techniques to restore facial details and advanced feature extraction mechanisms to mitigate the loss of discriminatory information in distant images. Additionally, adaptive thresholding based on image quality can be explored to dynamically adjust recognition confidence, thereby improving system scalability in larger areas.

4.6 Lighting

Since surveillance systems typically operate 24/7, lighting conditions may vary throughout the day. Different lighting conditions can affect the brightness of different parts of the captured images, sometimes casting shadows that obscure facial features. We tested the system's performance in darker environments, as shown in Table 9.

Table 6
(Color online) System performance at different distances from camera.

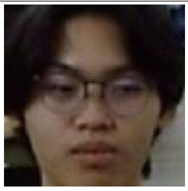
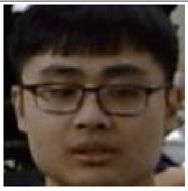
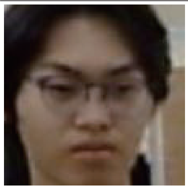
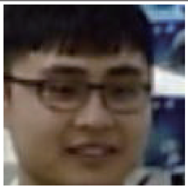
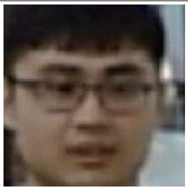
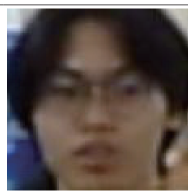
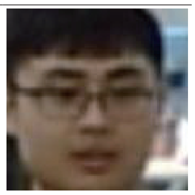
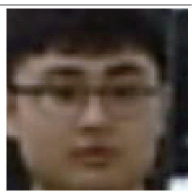
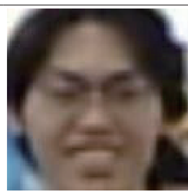
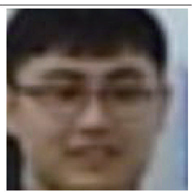
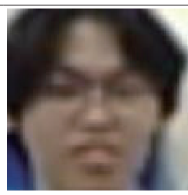
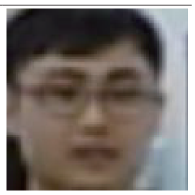
Distance (m)	Captured images	Detection rate (1p, 2p)	Cropped images	
2		53%, 44%	 100%	 100%
3		10%, 86%	 100%	 100%
4		23%, 75%	 100%	 100%
4.5		55%, 44%	 100%	 100%
5		49%, 43%	 100%	 99%
5.5		0%, 95%	 100%	 97%
6		24%, 71%	 36%	 73%
Free movement	—	31%, 61%	100%	99%

Fig. 7. (Color online) System misidentification of Chi.

Table 7
(Color online) New dataset—face photos.

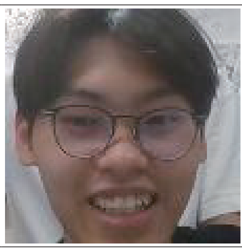
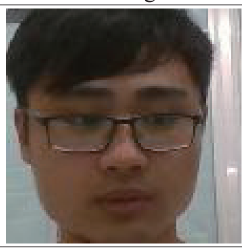
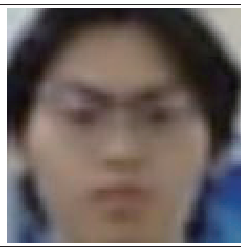
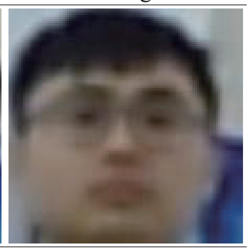
Name	Chi	Chung	Chi2	Chung2
Images				

Table 8
(Color online) System performance with new dataset.

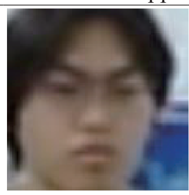
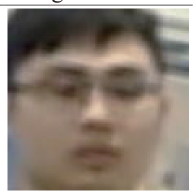
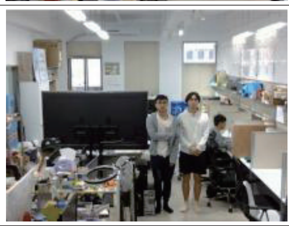
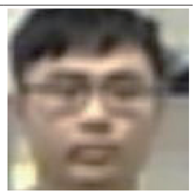
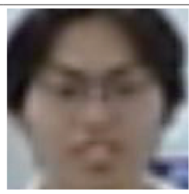
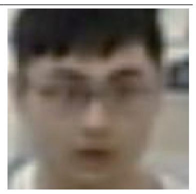
Distance (m)	Captured image	Detection rate (1p, 2p)	Cropped Images	
5.5		43%, 56%	 100%	 100%
6		11%, 88%	 100%	 100%
6.5		29%, 67%	 100%	 100%

Table 8
(Color online) (Continued) System performance with new dataset.

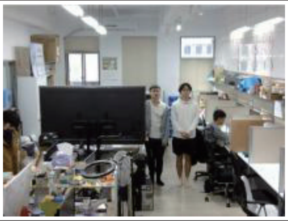
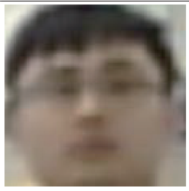
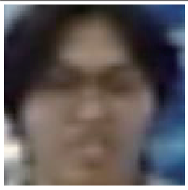
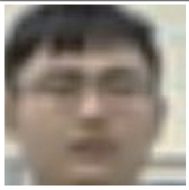
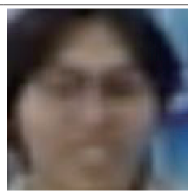
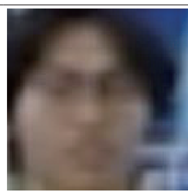
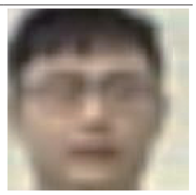
Distance (m)	Captured image	Detection rate (1p, 2p)	Cropped Images	
7		69%, 26%	 100%	 96%
7.5		24%, 73%	 100%	 100%
8		53%, 24%	 100%	 100%
8.5		51%, 41%	 78%	 100%
Free movement	—	36%, 56%	100%	100%

Table 9
(Color online) System performance under different lighting conditions.

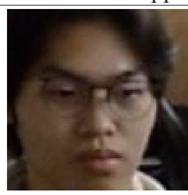
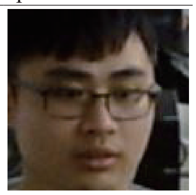
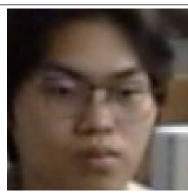
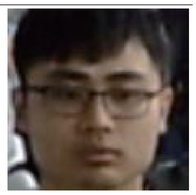
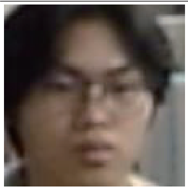
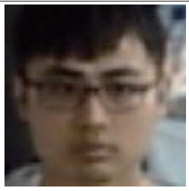
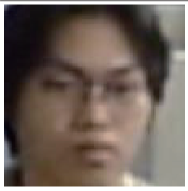
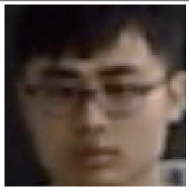
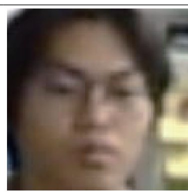
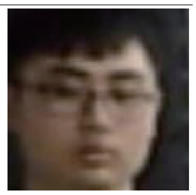
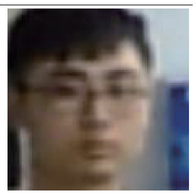
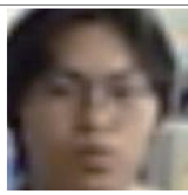
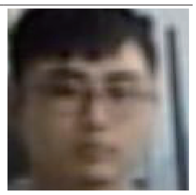
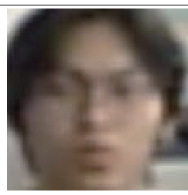
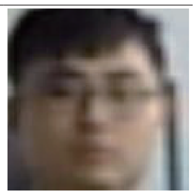
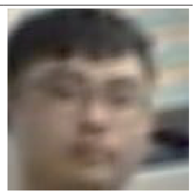
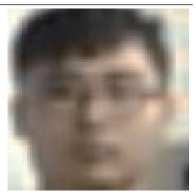
Distance (m)	Captured image	Detection rate (1p, 2p)	Cropped photos	
2		24%, 75%	 100%	 100%
3		35%, 63%	 100%	 100%

Table 9
(Color online) (Continued) System performance under different lighting conditions.

Distance (m)	Captured image	Detection rate (1p, 2p)	Cropped photos	
4		19%, 73%	 100%	 100%
4.5		6%, 92%	 100%	 100%
5		1%, 98%	 100%	 100%
5.5		46%, 53%	 100%	 100%
6		64%, 35%	 100%	 100%
6.5		51%, 47%	 100%	 99%
7		47%, 8%	 100%	 100%
7.5		66%, 11%	 100%	 99%
Free movement		20%, 78%	100%	97%

In darker environments, the system performed well within 6.5 m. However, at greater distances, face detection became unstable. For example, at 7 m, the system failed to detect 45% of the faces. Nevertheless, the recognized faces were still accurately identified, indicating that lighting does not affect the selected facial recognition model.

4.7 Masks

Even after the COVID-19 pandemic subsided, many people continued to wear masks. Additionally, criminals often wear masks to complicate subsequent tracking efforts. Masks block many facial features, making facial recognition models less effective. In this section, we examine the system's performance when masks are worn, as shown in Table 10.

The experiment revealed that wearing masks significantly affects system performance, particularly in face detection and recognition accuracy. The current model struggles to extract distinctive facial features from masked individuals, leading to an increased rate of misidentification. To enhance robustness, future improvements to incorporate training datasets that include occluded faces will allow the model to learn feature representations from partial facial information. Additionally, leveraging occlusion-robust recognition techniques, such as self-supervised learning and multi-region attention networks, may enhance system performance under masked conditions. The system sometimes misidentified masked faces as strangers (Table 6, left image), making the system highly unstable. Therefore, the system is not suitable for recognizing masked faces.

4.8 Angle

In normal movement, the faces captured by surveillance cameras are not always fully frontal, and different angles of side faces affect facial recognition accuracy. Extreme side angles can obscure facial features, leading to misrecognition. In this section, we discuss the impact of different side angles at three distances on the system's performance, as shown in Table 11.

At a distance of 5.5 m, the detection rate dropped significantly, with extreme side angles further exacerbating recognition failures. The current model relies heavily on frontal face features, making it less effective when dealing with substantial head rotations. To address this limitation, future work should explore the integration of side-profile recognition techniques and

Table 10
(Color online) System performance with masks.

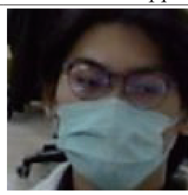
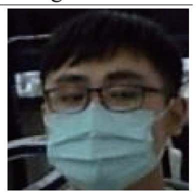
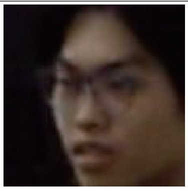
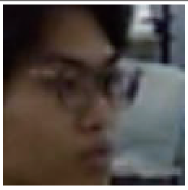
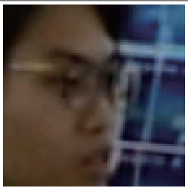
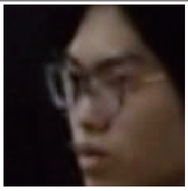
Distance (m)	Captured image	Detection rate (1p, 2p)	Cropped images	
2		51%, 5%	 43%	 100%

Table 11
(Color online) System performance at 3 m with different angles.

Angle	Detection rate	Cropped photo	Angle	Detection rate	Cropped photo
Left 30°	92.6%	 100%	Right 30°	100%	 100%
Left 45°	58.5%	 100%	Right 45°	93.4%	 100%
Left 60°	33.9%	 0%	Right 60°	91.7%	 100%

multi-angle face alignment models, which can improve performance in real-world surveillance scenarios where extreme facial angles are common. In the following experiments, the distance will be increased to determine the model's maximum recognition limit, as shown in Table 12.

At 7.5 m, even small side angles failed to achieve recognition. Although the right side maintained a perfect recognition rate of 100%, the left side achieved less than 50%, resulting in system errors, as shown in Table 13.

4.9 Experimental analysis

The system demonstrated optimal recognition performance within a range of 5.5 m, maintaining nearly perfect detection and recognition accuracy under ideal conditions. However, as the distance increased beyond 6 m, the system began to struggle with accurately identifying faces. At 6 m, the system misidentified individuals and recorded incorrect data, suggesting a critical limitation in the system's range. To further evaluate the effectiveness of our approach, we conducted a comparative analysis with two widely used face recognition methods: ArcFace and SphereFace. The results, summarized in Table 14, indicate that while our system achieves superior performance in recognizing blurred faces at moderate distances, its accuracy declines more significantly than does that of ArcFace in handling extreme facial angles and occlusions. SphereFace, which utilizes hypersphere embedding, exhibits greater robustness against occlusions but underperforms in low-resolution scenarios. This comparison highlights the strengths of our method in handling blurred images while also revealing its limitations in real-world challenges such as extreme occlusions and long-range recognition, as shown in Table 14.

Table 12
(Color online) System performance at 5.5 m with different angles.

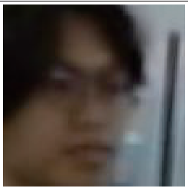
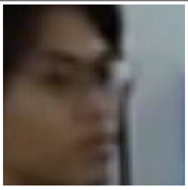
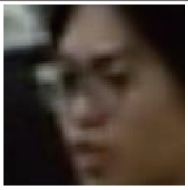
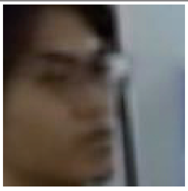
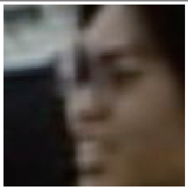
Angle	Detection rate	Cropped photo	Angle	Detection rate	Cropped photo
Left 30°	36.6%	 94%	Right 30°	73%	 100%
Left 45°	16.7%	 94%	Right 45°	46.9%	 100%
Left 60°	Very difficult to detect	 0%	Right 60°	Very difficult to detect	 0%

Table 13
(Color online) System performance at 7.5 m with different angles.

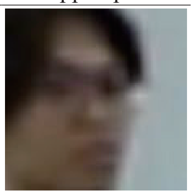
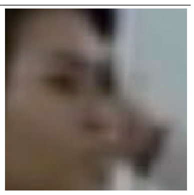
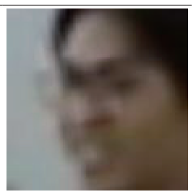
Angle	Detection rate	Cropped photo	Angle	Detection rate	Cropped photo
Left 30°	5%	 39.2%	Right 30°	12%	 100%
Left 45°	Very difficult to detect	 0%	Right 45°	Very difficult to detect	 0%

Table 14
Performance characteristics of proposed and existing face recognition methods.

Method	Recognition accuracy (5.5 m)	Recognition accuracy (8 m)	Occlusion handling	Extreme angle handling	Blurred image performance
Proposed method	97%	78%	Moderate	Poor	**High**
ArcFace	96%	85%	**High**	Moderate	Low
SphereFace	95%	80%	**Very High**	**High**	Low

By adding blurred faces to the training dataset, the maximum recognition distance was extended to 8 m. However, at this extended distance, the number of misidentifications increased, especially when facial features were partially occluded, resulting in lower accuracy.

However, under low-light conditions, the system was able to maintain high recognition accuracy up to 6.5 m, but beyond that, detection became unreliable. Although the system was able to recognize faces correctly when they were detected, the increased difficulty in detecting faces under poor lighting highlights the sensitivity of the face detection model to environmental factors.

We also adapted the mask scenario at low resolution, and the system's performance significantly deteriorated when individuals wore masks, a common scenario in the post-COVID era. Face detection was notably poor, and the system frequently misidentified masked individuals as strangers. This suggests that the model, as implemented, is not well equipped to handle occluded facial features such as when wearing masks, which is a major limitation for real-world applications where mask-wearing is common.

The experiment further showed that the system also struggled with detecting and recognizing faces at extreme angles, particularly when the subject was turned more than 45 degrees from the camera at low resolution. At a shorter distance (3 m), the system performed reasonably well even with moderate side angles (30 degrees), but the accuracy dropped significantly as the angle increased. At 7.5 m, even small side angles led to detection and recognition failures, showing that facial orientation plays a critical role in the system's effectiveness.

5. Conclusions

The potential applications of this system are vast and not limited to campus environments. By integrating blurred face recognition with existing surveillance systems, security can be enhanced without the need for costly upgrades to older cameras. Additionally, the system automatically records key information, such as the time when suspects appear, considerably reducing the time traditionally required to review footage and track down suspects. While our proposed approach demonstrates high performance in recognizing blurred images, comparative analysis with ArcFace and SphereFace reveals certain trade-offs. Future research should focus on integrating multi-angle face alignment and occlusion-aware recognition models to further enhance system robustness and scalability. This efficiency can contribute to reducing crime rates. However, several limitations remain. For instance, factors such as wearing masks or significant facial rotation can compromise the system's performance. If an individual deliberately obscures their facial features or wears a realistic mask, the system may fail to accurately capture the suspect. Moreover, since the system not only records the time when a suspect appears but also captures close-up images of the individual, concerns about privacy may arise. Given the growing emphasis on personal privacy, storing such facial data can raise legal concerns, and the proper consideration of relevant regulations is essential when implementing this system.

6. Future Work

This system consists of two main components: it first processes the images from surveillance footage through face detection, and then crops the detected faces and sends them to a facial recognition model. From the results of various experiments, it is clear that the system's limitations are often due to suboptimal detection rates, which hinder its overall effectiveness. However, once images are successfully detected and processed by the facial recognition model, the accuracy remains reasonably high. Therefore, future work should focus on improving the face detection algorithm to increase the detection rate, thereby achieving more efficient and secure recognition. It is crucial to consider that detection and recognition are closely related—if facial images lacking sufficient features are sent to the recognition model, it may result in a high number of false positives. Special care must be taken when designing the algorithm to avoid this problem.

Additionally, although the system currently records precise timestamps of suspect appearances, there is room for improvement in visualization and usability. For instance, when presearching for specific individuals, the system can read the log of recorded timestamps and mark them on a video timeline, allowing for a faster and more accurate identification of suspect appearances during surveillance footage review.

Acknowledgments

This study was partly supported by the National Science and Technology Council (NSTC) of Taiwan under Grant nos. NSTC 114-2221-E-019-015-, 114-2637-8-011 -004 -, 114-2622-E-011 -007 -, 113-2634-F-A49-007-, 115-2637-8-011 -003 -, 115-2622-E-011 -012 -, 115-2222-E-036 -001 -, 114-2622-8-110 -003 -, 114-2637-8-992-009 -, 115-2221-E-019-065 -, and 115-2221-E-152-003 -.

References

1 K. Zhang, Z. Zhang, Z. Li, and Yu Qiao: arXiv:1604.02878.

2 D. H. Hubel and T. N. Wiesel: Brain and visual perception: The story of a 25-year collaboration (Oxford University Press, 2025).

3 Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel: Neural Comput. **1** (1989) 541. https://doi.org/10.1162/neco.1989.1.4.541

4 L.-A. Zeng and Wei-Shi Zheng: IEEE Trans. Image Process. **33** (2024) 1600.

5 K. Jampachaisri, K. Tinochai, S. Sukparungsee, and Y. Areepong: IEEE Access **8** (2020) 51460. https://doi.org/10.1109/ACCESS.2020.2979872

6 S. Hu, X. Wang, and S. Lyu: IEEE Trans. Pattern Analysis Mach. Intell. **45** (2023) 13599. https://doi.org/10.1109/TPAMI.2023.3296062

7 V. A. Krishna, A. Reddy, and D. Nagajyothi: 2021 IEEE Int. Conf. Mobile Networks and Wireless Communications (ICMNWC) (Tumkur, Karnataka, India, 2021) pp. 1–4. https://doi.org/10.1109/ICMNWC52512.2021.9688430

8 Y. Liu, G. Chang, G. Fu, Y. Wei, J. Lan, and J. Liu: 2022 34th Chinese Control and Decision Conf. (CCDC) (Hefei, China, 2022) pp. 721–724. https://doi.org/10.1109/CCDC55256.2022.10034228

9 O. Schlesinger, N. Vigderhouse, D. Eytan, and Y. Moshe: ICASSP 2020 - 2020 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP) (Barcelona, Spain, 2020) pp. 1135–1139. https://doi.org/10.1109/ICASSP40776.2020.9053446

10 Z. Cheng, X. Zhu, and S. Gong: Low-Resolution Face Recognition, C. Jawahar, H. Li, G. Mori, K. Schindler Eds. Computer Vision – ACCV 2018. ACCV 2018. Lecture Notes in Computer Science (Springer, Cham.) **11363** (2019). https://doi.org/10.1007/978-3-030-20893-6_38
11 C. Song, Z. He, Y. Yu, and Z. Zhang: 2021 3rd Int. Conf. Applied Machine Learning (ICAML) (Changsha, China 2021) pp. 76–79. https://doi.org/10.1109/ICAML54311.2021.00024
12 Chengrui Wang, Han Fang, Yaoyao Zhong, and Weihong Deng: 2022 MLFW: Biometric Recognition: 16th Chinese Conf. CCBR 2022 (Beijing, China, November 11–13, 2022) (Springer-Verlag, Berlin, Heidelberg) pp. 180–188. https://doi.org/10.1007/978-3-031-20233-9_18
13 F. Cunico, A. Toaiari, and M. Cristani: A Masked Face Classification Benchmark on Low-Resolution Surveillance Images, J. J. Rousseau and B. Kapralos, Eds. Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 Int. Workshops and Challenges. ICPR 2022. Lecture Notes in Computer Science (Springer, Cham.) **13643** (2023). https://doi.org/10.1007/978-3-031-37660-3_4
14 M. S. Ejaz and M. R. Islam: 2019 Int. Conf. Sustainable Technologies for Industry 4.0 (STI) (Dhaka, Bangladesh 2019) pp. 1–6. https://doi.org/10.1109/STI47673.2019.9068044
15 W. W. Bledsoe: Description of LISP Functions in the Theorem Prover (University of Texas Memo, Department of Mathematics) (January 1970).
16 W. W. Bledsoe: J. ACM **13** (1966) 304.
17 W. W. Bledsoe and H. Chan: A Man-Machine Facial Recognition System-Some Preliminary Results, Technical Report PRI 19A (Panoramic Research, Inc., Palo Alto, California) (1965).
18 W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song: 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI, USA, 2017) pp. 6738–6746. https://doi.org/10.1109/CVPR.2017.713
19 W. Liu, Y. Wen, B. Raj, R. Singh, and A. Weller: IEEE Trans. Pattern Analysis Mach. Intell. **45** (2023) 2458. https://doi.org/10.1109/TPAMI.2022.3159732
20 M. He, J. Zhang, S. Shan, and X. Chen: IEEE Trans. Image Process. **33** (2024) 1588.
21 H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu: 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR) (Salt Lake City, UT, USA, 2018) pp. 5265–5274. https://doi.org/10.1109/CVPR.2018.00552
22 J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou: IEEE Trans. Pattern Analysis Mach. Intell. **44** (2022) 5962. https://doi.org/10.1109/TPAMI.2021.3087709.
23 Y. Li, J. Li, Y. Si, J. Tan, and Q. He: IEEE/ACM Trans. Audio, Speech, Language Process. **32** (2024) 2297.
24 H. Guo, X. Su, C. Wu, B. Du, and L. Zhang: IEEE Trans. Image Process. **33** (2024) 2599.
25 M. He, J. Zhang, S. Shan, and X. Chen: IEEE Trans. Image Process. **33** (2024) 1588.
26 A. George, C. Ecabert, H. O. Shahreza, K. Kotwal, and S Marcel: IEEE Trans. Biometrics, Behavior, Identity Sci. **6** (2024) 158.