A Deep Learning based Approach for Student Attendance Recognition

Minh Nguyet Nguyen[1,2], Phuong Lan Phan[2]
[1] Can Tho University, Can Tho, Vietnam
[2] Can Tho University Software Center, Can Tho, Vietnam
Correspondence: Minh Nguyet Nguyen, nmnguyet@cusc.ctu.edu.vn, Phuong Lan Phan, pplan@cit.ctu.edu.vn
Received: 14/07/2024, revised: 20/09/2024, accepted: 11/11/2024
Digital Object Identifier:: 10.32913/mic-ict-research-vn.v2024.n3.1314

Abstract: Traditional attendance recognition methods - such as using cards or signing into lists are susceptible to fraud - often consumes considerable time and effort in managing and processing information. With advancing technology and increasing security requirements, modern attendance recognition methods offer higher accuracy and convenience for organizations. This paper proposes an approach utilizing DeepFace technology and the deep learning architectures (MobileNet, ResNet) to face detection, face recognition and face anti-spoofing. The proposed approach is experimented on the existing dataset within the research community (CelebA-spoof) and on a dataset (named CUSC-Student) built by ourselves; and is evaluated by two metrics - Accuracy and F1-Score. From the experimental results, this approach also identifies the suitable model that achieves the high accuracy while requiring less computing resources. Therefore, it can be applied in practice, specifically for recognizing student attendance at our division.

Keywords: *Anti-Spoofing, face detection, face recognition, deep learning*

Tiêu đề: **Điểm danh sinh viên theo tiếp cận dựa trên mạng học sâu**

Tóm tắt: Các phương pháp nhận diện điểm danh truyền thống - như dùng thẻ hoặc ký vào danh sách - dễ bị gian lận và thường đòi hỏi nhiều thời gian và công sức để quản lý, xử lý dữ liệu. Với sự phát triển của công nghệ và yêu cầu bảo mật ngày càng cao, các phương pháp điểm danh hiện đại mang lại độ chính xác và tiện lợi cao hơn cho các tổ chức. Bài báo này đề xuất một phương pháp sử dụng công nghệ DeepFace kết hợp với các kiến trúc học sâu (MobileNet, ResNet) cho các tác vụ phát hiện khuôn mặt, nhận diện khuôn mặt và chống giả mạo. Phương pháp này được thử nghiệm trên tập dữ liệu sẵn có trong cộng đồng nghiên cứu (CelebA-spoof) và một tập dữ liệu tự xây dựng (được đặt tên là CUSC-Student), được đánh giá bằng hai chỉ số - độ chính xác và F1-Score. Từ kết quả thực nghiệm, phương pháp này xác định được mô hình phù hợp giúp đạt độ chính xác cao, đồng thời yêu cầu ít tài nguyên tính toán hơn. Do đó, phương pháp này có thể ứng dụng vào thực tiễn, đặc biệt trong việc điểm danh sinh viên tại đơn vị của chúng tôi.

Từ khóa: *Anti-Spoofing, face detection, face recognition, deep learning*

I. INTRODUCTION

Automatic and real-time attendance recognition is a necessary activity for any organization, especially for educational institutions, because it helps to manage attendees effectively. Face recognition [1] is an advanced technology with numerous applications in security, surveillance, and user convenience. However, it faces significant challenges, particularly its vulnerability to spoofing attacks. Attackers can use printed photos, videos, or masks to bypass recognition systems, compromising the security and reliability of the technology. With the rise of social media, people, especially the youth, frequently post personal images online. This trend increases the risk of face recognition systems being spoofed. Criminals can exploit high-resolution images, videos, and masks to deceive these systems, posing serious security risks to institutions.

Deep learning [2] is a machine learning method in the field of artificial intelligence that focuses on building and training deep neural networks to automatically learn and understand data. It is called "deep" because it involves multiple layers of computational units known as neurons. These layers are interconnected, passing data from one layer to the next, thus forming a neural network capable of complex learning. Deep learning models, especially convolution neural networks (CNNs) [3], provide superior accuracy, robustness, and real-time performance in face recognition tasks. These models can handle variations in lighting, angles, and expressions, ensuring reliable recognition under diverse conditions.

This paper proposes an approach for attendance recognition by combining face detection, face anti-spoofing and face recognition. We then develop a web application for the proposed approach by leveraging existing technologies and methods. In this application, we embed some models, then retrain and fine-tune them to suit our dataset.

The paper is structured into five distinct sections to provide a comprehensive overview of our research. The introduction section outlines the context of our study, explaining the importance of face recognition and anti-spoofing systems and the need for enhanced security measures in automated systems. Following this, the related work section presents the existing studies that have tackled similar challenges, highlights their methodologies and findings. The third section is dedicated to detailing our proposed approach. The fourth section delves into experiments. This section covers the tools, datasets and metrics used for training and evaluating the performance of approach, and presents the experimental results. Finally, the conclusion section summarizes the key findings of our study as well as suggests the work future.

II. RELATED WORK

Recent advancements in machine learning have spurred significant progress in the integration of face recognition with anti-spoofing techniques. This intersection of technologies has been driven by several seminal studies.

The authors of research [4] provide an extensive survey of deep learning methods used for face anti-spoofing, covering both single- and multi-modal techniques. The authors review a variety of deep learning models and their application in preventing spoofing attacks such as those involving photos, videos, and 3D masks. The survey highlights the integration of these models with face recognition systems and discusses their effectiveness in improving security and reliability. The comprehensive analysis offers insights into the strengths and weaknesses of different approaches, guiding future research in this field.

The authors of research [5] introduces a novel approach for face anti-spoofing that leverages patch-based and depth-based Convolutional Neural Networks. The methodology involves dividing the facial image into smaller patches and using a CNN to extract features from these patches. Additionally, a depth-based CNN is employed to estimate the depth map of the face, which helps in distinguishing between real and spoof faces by analyzing the texture and depth information. The experimental setup includes two CNN architectures: one for patch-based feature extraction and another for depth estimation. The patch-based CNN analyzes local regions of the face, providing detailed texture information, while the depth-based CNN focuses on the overall depth profile of the face. These two models work in tandem to enhance the anti-spoofing performance. The author utilizes publicly available datasets for training and evaluation, including the CASIA Face Anti-Spoofing Dataset and the Replay-Attack Dataset. The proposed method demonstrates significant improvements in detecting spoofing attacks compared to traditional approaches. The combination of patch-based and depth-based CNNs already achieves 89% reduction of Equal Error Rate, with notable performance on challenging datasets.

The authors of research [6] propose a new approach for tackling the face of anti-spoofing challenges. The method focuses on learning polysemantic spoof traces by disentangling facial features across multiple modalities. This approach enhances the detection of spoof attacks by effectively separating genuine facial attributes from spoofed ones using multimodal information. The research contributes to advancing the face recognition security, leveraging innovative machine learning techniques to address complex spoofing threats in practical scenarios.

Besides, related to facial recognition and facial attribute analysis, DeepFace technology [7] is applied in authentication, security systems, social media, etc. DeepFace, developed by Facebook, utilizes deep neural networks to detect and recognize faces with high accuracy. The process begins with image preprocessing, where face detection algorithms (such as Viola-Jones or HOG + SVM) identify the position of the face in the image. The detected face is then aligned to ensure that the features of the face such as eyes, nose, and mouth are in inconsistent positions across all images, reducing variations caused by angles and expressions. The core of DeepFace is its deep neural network architecture, which includes convolution layers to extract essential features like edges, contours, and textures from the face image. These features are then consolidated through fully connected layers, culminating in an embedding layer that converts the features into a fixed-length vector representation (often 4096 dimensions). This vector, or face embedding, effectively represents the face in a feature space. For face recognition, the embedding of a new face is compared with the embedding of known faces in the database using similarity measures like Euclidean distance or cosine similarity. If the distance between the new face embedding and a known face embedding is below a certain threshold, the new face is recognized as matching the known face. Additional verification steps can be performed to ensure high accuracy, especially in security-sensitive applications.

In image classification and object detection, the architectures such as YOLO, GoogLeNet, ResNet, VGG and MobileNet are popular. YOLO [8] is a unified architecture which uses a single convolutional network for classification and bounding box prediction, focusing on real-time object detection. GoogLeNet (or Inception-V1) [9] is a deep convolutional neural network architecture which employs inception modules for multi-scale feature extraction and emphasizes the improved utilization of the computing resources inside the network. ResNet [10] is a deep learning architecture with hundreds or even thousands of layers. It is designed to improve the training of very deep neural networks by incorporating residual blocks and skip connections which enables gradients to flow more easily through the network. VGG [11] is a deep CNN architecture with a simple and uniform design. It uses small convolutional filters throughout the network in capturing fine details in images. MobileNet [12] [13] [14] is a lightweight CNN architecture, focused on efficient deployment on mobile and resource-constrained devices. MobileNetV2 [13] and MobileNetV3 [14] are often considered for developing applications. MobileNetV2 is an improved version of the MobileNet architecture, designed to enhance the performance and efficiency of CNNs on mobile devices. It introduces two key concepts: Inverted Residuals and Linear Bottlenecks. These concepts help preserve more information and reduce computational requirements. MobileNetV2 uses three main layers for each block: 1x1 Convolution with ReLU6 to expand the dimensionality, Depthwise Convolution with ReLU6 to each channel separately, and 1x1 Linear Convolution to compress the dimensionality back. This architecture achieves higher efficiency and performance, making it suitable for mobile and embedded applications. Additionally, it includes hyperparameters to balance speed and accuracy by adjusting the number of channels and input resolution. MobileNetV3 Large ad MobileNetV3 Small are two main variants of MobileNetV3, but they have distinct architectures and use cases. MobileNetV3 Large integrates optimization techniques from MobileNetV2 with new ideas from MnasNet. It employs ReLU6 and h-swish (Hard-Swish) activation functions to enhance performance. MobileNetV3 Large delivers high performance while being efficient and resource-saving. Therefore, it is suitable for various applications, including image classification and object detection. MobileNetV3 Small is optimized for efficiency and speed on low resource devices. It has fewer layers and parameters. MobileNetV3 Small also uses the Hard Swish activation function to improve performance, selects Squeeze-and-Excitation modules to enhance feature recalibration, and uses Network Architecture Search to fine-tune the architecture. The comparison results of RestNet, GooLeNet and VGG models presented in researches [10] [11] [15] [16] have shown that: the test error of VGG to be competitive with that one of GoogLeNet; the test error of ResNet to be the smallest and its accuracy is the largest. The experimental results in [12] shows that MobileNet is more accurate than GoogLeNet while being less computation (the number of multi-adds) and smaller (the number of parameters). According the research in [13], MobileNet achieves competitive accuracy and requires fewer parameters and computations than YOLO and RestNet. The object detection models based on deep learning have been surveyed by [16] shows that for resource constrained environments, the MobileNet is competitive on classification accuracy, latency, and number of parameters. MobileNet has a lightweight architecture, requires less memory and computing resources whereas it achieves competitive accuracy compared to other architectures. It is the balance between performance and accuracy.

Education plays a crucial role in the development of a nation. The number of educational institutions is increasing due to the need for learning, competition in the labor market, technological development, and educational policies. In educational institutions, automatic and real-time attendance recognition is essential because it supports the learner management fast and effectively. However, educational institutions may face equipment limitations due to financial constraints. As a result, the quality and the availability of devices such as computers and cameras can be affected. Devices may not have strong configurations or have limited computing resources. For the attendance recognition problem, if the computer system is not powerful enough, using ResNet or VGG or YOLO can lead to slow and inefficient processing speed because they require more computing resources. Therefore, MobileNet is the best choice in this case.

III. PROPOSED APPROACH

The proposed approach for attendance recognition based is shown in Figure 1 and Algorithm 1. This proposal consists of 3 stages playing the distinct roles.

The first stage "Face Detection" is used to process images, detect faces and recognize faces. The DeepFace technology is utilized to detect and crop the original image, focusing solely on the face and excluding the distracting details like backgrounds, hands and legs. The "Yunet" algorithm is employed to detect faces in the image, adjust their alignment, and crop them based on the coordinates and dimensions of the faces. Subsequently, these processed images are stored for further processing and analysis in anti-spoofing models.

The second stage "Face Anti-Spoofing" is used to distinguish between real face images and fake face images. This stage embeds four models: MobileNetV2, MobileNetV3

Large, MobileNetV3 Small and ResNet18. Depending on the resource constraints of devices, one of four models will be used. The selected model will identify if the face is fake or real. If the face is determined as real, the extracted real face image will be the input of the next component. Therefore, it helps enhance the accuracy and security of the recognition system.

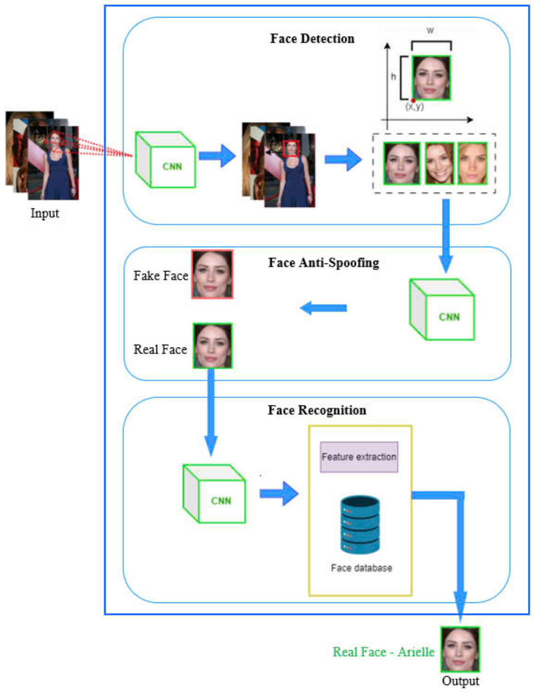

Figure 1. The proposed approach for attendance recognition.

The last stage "Face Recognition" plays a crucial role in converting each face image into a unique feature vector (embedding) by using the ArcFace [17]. This process involves computing embeddings for each image, generating a unique ID for each new face, and adding these embeddings to the FAISS [18] index for efficient management and retrieval. The last stage also utilizes DeepFace to maintain a mapping between the face IDs and usernames in the database; consequently, the process of computing and managing face embeddings becomes efficient, supporting optimal performance for face recognition.

A web application is developed to implement the proposed approach. As shown in Figure 2, this application uses HTML and CSS to design interfaces; JavaScript to send requests; functions written in Python programming language to handle requests from the user interface; and MySQL to store, select and update student information and attendance records.

The application also integrates the following. Leveraging these advanced technologies ensures security and accuracy in recognizing the attendance.

Algorithm 1: attendance_recognition(frame)

```
size ← (224, 224);
detector ← "yunet";
face_objs ← extractFaces(frame,
  size, detector);
foreach face_obj in face_objs do
    confidence ← getConfi-
     dence(face_obj);
    if confidence ≥ 0.9 then
        facial_area ← getFa-
         cialArea(face_obj);
        model ← selectModel();
        antispoof_flag ← predict-
         Face(frame, facial_area,
         model);
        if antispoof_flag == 1 then
            face ← getFace-
             Frame(facial_area);
            presentor ← "Arcface";
            embeddings ← represent-
             Face(face, presentor);
            indices ← recognize-
             Faces(embeddings);
            recognized_name ← get-
             Name(indices);
            displayInformation(recognized
             _name);
            updateDatabase(recognized
             _name);
        end
        else
            showError();
        end
    end
end
```

- "Yunet" algorithm in DeepFace for image processing and face detection.
- MobileNetV2, MobileNetV3 (Large, Small) and ResNet18 models for face-anti spoofing.
- ArcFace method for face recognition.

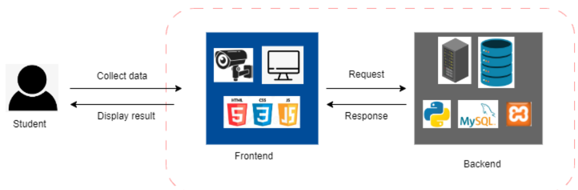

Figure 2. The web application architecture to implement the proposed approach.

The application provides key functions such as user registration and attendance tracking. Initially, students take attendance using a genuine facial image captured by the application. This image undergoes preprocessing using DeepFace technology with the "Yunet" algorithm for image processing and face detection. It is then passed through MobileNetV2, MobileNetV3 (Large and Small) and ResNet18 models for face anti-spoofing to ensure authenticity. Features are obtained from the processed image by the ArcFace method and saved in the embedded database along with the student's personal information. When a student needs to check in, the application captures their image again, preprocesses it, and runs it through the anti-spoofing models. The extracted features are searched in the database. Attendance information is then recorded accordingly. This provides a secure and efficient method for managing student attendance. The results will be visually displayed on the user interface. If the attendance is successful, the screen will show the student's photo, a message confirming that the face is real, and the student's information. Conversely, if no face is detected or if a fake face is identified, the application will present the corresponding alert, as depicted in Figure 3.

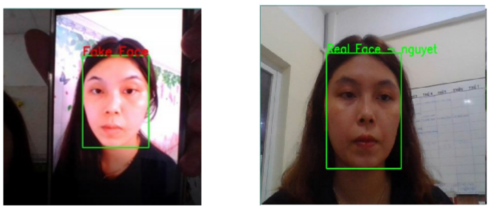

Figure 3. Displaying the result of the user's face detection.

IV. EXPERIMENTS

The performance of proposed approach is evaluated by using K-folds cross validation method, four metrics (Accuracy, Precision, Recall and F1-Score), and two datasets (CelebA-Spoof and CUSC-Student).

In K-folds method, a dataset is partitioned into K folds (subsets) of equal size. The model is trained and validated K times, then gets the average. Each time, (K-1) folds are used for training, and the rest fold is used for validating the model. This method helps reduce bias and variance, leading to a more reliable model. We conduct three experiments.

- Experiment 1 addresses the question "Which model is the most suitable for integration into a student attendance recognition application?" In this experiment, we chose four models for comparisons: MobileNetV2, MobileNetV3 Large, MobileNetV3 Small and ResNet18. The reason for this selection is as follows. As mentioned in the related work, three MobileNet models archieve the competitive accuracy while requiring less memory and computing resources. The RestNet model has the smallest test error and the highest accuracy compared to GooLeNet and VGG.
- Experiment 2 explores the question "How does the face detection stage affect the model's performance?" This experiment compares the performance of the proposed approach both with and without the face detection.
- Experiment 3 examines the question "Which hyperparameter values should we select for integration into a student attendance recognition application?" This experiment changes the values of three crucial hyperparameters (the learning rate, the number of epochs, and the batch size) to identify those that yield optimal performance.

1. Datasets

a) *CelebA-Spoof dataset*

The CelebA-Spoof dataset [19] is a comprehensive and large-scale resource designed for face anti-spoofing research, featuring 625,537 images from 10,177 subjects. This dataset includes both genuine and spoofed face images, meticulously annotated with 43 detailed attributes related to facial features, illumination, environment, and spoofing techniques. Genuine images are sourced from the CelebA-Spoof dataset, while spoof images are specifically collected and annotated for anti-spoofing purposes. Of the 43 attributes, 40 pertain to genuine images, encompassing a wide range of facial elements and accessories, including skin, nose, eyes, eyebrows, lips, hair, hats, and eyeglasses. The remaining 3 attributes describe spoof images, detailing spoof types, environmental settings, and lighting conditions.

CelebA-Spoof is a critical dataset for training and evaluating algorithms in face anti-spoofing due to its diverse and complex conditions, including various lighting, angles and expressions, which present significant challenges.

However, this proposed approach uses only a subset of the CelebA-Spoof dataset instead of the entire dataset. This decision brings several benefits, such as reducing computational costs and memory usage, while also helping to avoid overfitting. There are 748 folders were randomly selected and each folder contains two sub-datasets: the live folder and the spoof folder. Each folder contains 12,587 real face images and 35,981 fake face images, captured under various conditions, including different colors, face angles, lighting conditions, and backgrounds.

b) *CUSC-Student dataset*

The CUCS-student dataset is collected and processed by ourselves. We use the web application presented in Section

3 to collect the student photos. The system automatically captures photos when students place their faces in front of the camera. Each student's photo is then saved into individual folders corresponding to each student.

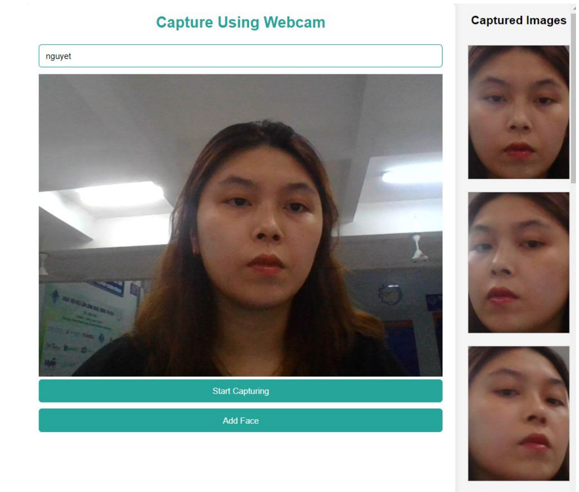

Figure 4. The function "Capture Face" of web application.

The dataset comprises facial images of students participating in short-term or long-term courses at the "Can Tho University Software Center (CUSC)". The dataset contains 50 folders, each corresponding to a student at CUSC. Within each folder, there are two sub-datasets: one is the live set, containing 4136 real face images of the student, and the other is the spoof set, containing 12,408 fake face images of the student. These images are captured under different lighting conditions and various locations within the CUSC campus, ensuring diverse environmental contexts.

During the data pre-processing phase, images are initially resized to a standardized dimension, ensuring uniform resolution and size across all images. This standardization facilitates streamlined model training processes. The resized images are converted into array format for seamless handling in subsequent preprocessing steps. To further enhance model performance, data augmentation techniques such as scaling, rotation, flipping, and adjustment of lighting conditions are applied. These techniques artificially expand the dataset by creating multiple variations of the initial images.

2. Evaluation Metrics

In face recognition and anti-spoofing, evaluating the model performance is crucial. A popular and effective tool used for this purpose is the confusion matrix[20]. We use a 2x2 confusion matrix, as there are only two classes: real face and fake face. This confusion matrix is presented as follows.

Table I
A 2X2 CONFUSION MATRIX.

	Predicted Real	**Predicted Fake**
Actual Real	True Positive (TP)	False Negative (FN)
Actual Fake	False Positive (FP)	True Negative (TN)

Where:

- True Positive (TP): The number of instances correctly identified real faces.
- False Negative (FN): The number of instances correctly identified fake faces.
- False Positive (FP): The number of instances incorrectly identified fake faces as real.
- True Negative (TN): The number of instances incorrectly identified real faces as fake.

From the confusion matrix, four metrics - Accuracy, Precision, Recall, F1-Score - are calculated according to formulas (1)(2)(3)(4), respectively. In which, F1-Score and Accuracy metrics are often used together to evaluate the performance of a model.

Accuracy is the proportion of correct predictions out of all predictions. Accuracy measures how often a model correctly predicts the outcome.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision is the proportion of true positive predictions out of all positive predictions. Precision emphasizes correctness.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall (or Sensitivity or True Positive Rate) is the proportion of true positive predictions out of all actual positives. Recall emphasizes completeness.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-Score is often used to evaluate the model's performance because it considers both metrics Precision and Recall simultaneously.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

3. Experimental Results

a) Experiment 1: Finding the most suitable model for a student attendance recognition application

In the first experiment, the performance of the proposed approach where one of four models (MobileNetV2, MobileNetV3 Large, MobileNetV3 Small and ResNet18) is used by the second stage “Face Anti-Spoofing” is evaluated on two datasets: CelebA-Spoof and CUSC-Student. In Experiment 1, these datasets are already processed by the first stage - i.e. it uses the “Yunet” algorithm for face detection and image preprocessing - before training. The goal of this experiment is to assess how each model performs with different datasets to make a decision on the most suitable model for practical deployment.

The experimental results of MobileNet models are presented in two figures (Figure 5 and Figure 6).

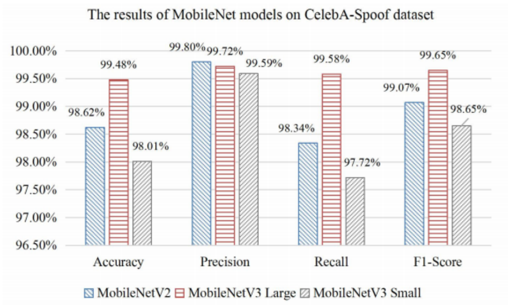

Figure 5. The experimental results of MobileNet models on CelebA-Spoof dataset.

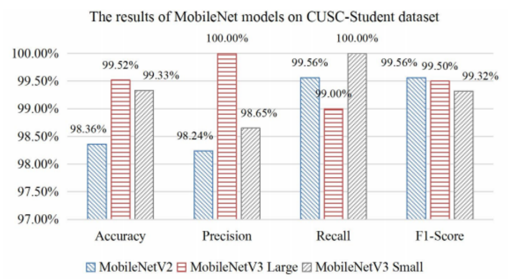

Figure 6. The experimental results of MobileNet models on CUSC-Student dataset.

Among the models trained on the CelebA-Spoof dataset, MobileNetV3 Large shows the best performance, with the highest accuracy at 99.48%, the recall at 99.58%, and the F1-Score at 99.65%. This model effectively balances precision and recall, making it a strong choice for face recognition and anti-spoofing tasks on the CelebA-Spoof dataset. MobileNetV2 also shows good performance, reaching an accuracy of 98.62% and a F1-Score of 99.07%. Although slightly less effective than MobileNetV3 Large, it remains a reliable option. MobileNetV3 Small, while delivering good overall performance, lagged behind the other two models, particularly in recall (97.72%) and F1-Score (98.65%). Despite this, it is still a viable choice, though it may miss more real faces compared to MobileNetV2 and MobileNetV3 Large. Similarly, on the CUSC-Student dataset, MobileNetV3 Large stands out, achieving the highest accuracy at 99.52%, the perfect precision at 100%, and the F1-Score at 99.50%. This model balances the precision and the recall effectively, making it a dependable option for face recognition and anti-spoofing tasks. MobileNetV2 also demonstrated strong performance, with an accuracy of 98.36% and a F1-Score of 99.56%. It is highly reliable and even surpasses MobileNetV3 Large in recall, indicating its strength in correctly identifying real faces. MobileNetV3 Small, while slightly behind the other two models, still delivered strong performance with the accuracy at 99.33%, the precision at 98.65% and the F1-Score at 99.32%.

We continue to compare the performance of MobileNetV3 Large (the superior model of three MobileNet models) and ResNet18 model. The accuracy and F1-Score of these two models on two datasets are presented in Figure 7. We have also observed the total number of floating-point operations per second (FLOPs) and the total number of parameters (Params) of both models as Table II. The results presented in Figure 7 and Table II shows that MobileNetV3 Large is more remarkable than ResNet18.

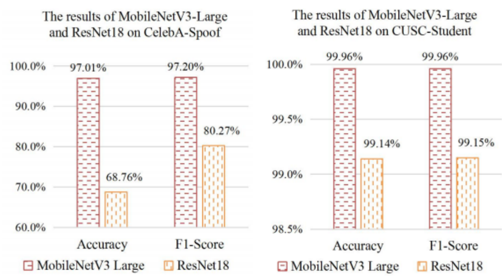

Figure 7. The comparison on the accuracy and the F1-Score of MobileNetV3 Large and ResNet18 models.

Table II
. THE COMPARISON ON THE TOTAL PARAMS AND THE TOTAL FLOPS OF MOBILENETV3 LARGE AND RESNET18 MODELS.

Model	Params	FLOPs
MobileNetV3 Large	5.49M	231.32MMac
ResNet18	11.69M	1.83GMac

The experimental results indicate that all models are effective in face recognition and anti-spoofing. MobileNetV3 Large stands out as the better choice for both the CelebA-Spoof dataset and CUSC-Students dataset because it provides high accuracy and the ability to clearly identify the difference between real and fake facial images while requiring less computing resources. Therefor, MobileNetV3 Large is the most suitable model for integration into the student attendance recognition application.

b) Experiment 2: Examining the impact of the face detection stage on the performance of the proposed approach

The aim of this experiment is to compare the performance of the proposed approach with and without the face detection stage. MobileNetV3 Large model is used in this experiment.

Table III shows the performance of the proposal without the first stage. The input of MobileNetV3 Large remain CelebA-Spoof and CUSC-Student datasets, but images from these datasets are not preprocessed by the first stage.

Table III
THE PERFORMANCE OF THE PROPOSED APPROACH WITHOUT THE FIRST STAGE.

Dataset	Accuracy	Precision	Recall	F1-Score
CelebA-Spoof	89.27%	89.39%	99.76%	93.57%
CUSC-Student	82.53%	77.04%	67.46%	68.42%

From the result of MobileNetV3 Large in Experiment 1 and the collected data in Table IV, the performance of the proposed approach with and without the use of the first stage is shown in Figure 8. The data in Figure 8 indicates that on both datasets, the proposed approach with the first stage will achieve the higher performance.

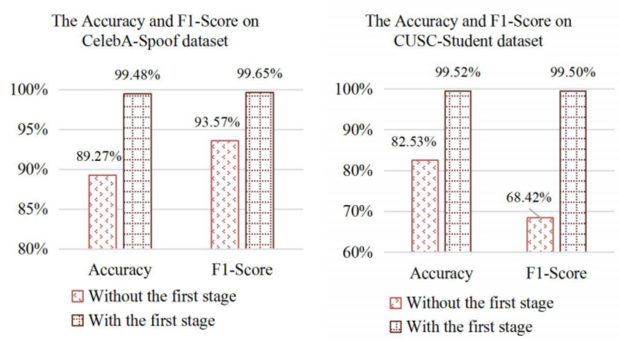

Figure 8. The comparison on the performance of the proposed approach with/without the first stage.

The experimental results indicate that incorporating a face detection phase in the application improves the overall performance of the proposed approach.

c) Experiment 3: Identifying the appropriate hyperparameter values

The MobileNetV3 Large model and its three crucial hyperparameters (the learning rate, the number of epochs, and the batch size) are selected for Experiment 3. The values of these three hyperparameters - a triple set - are changed to evaluate the performance of MobiNetV3-Large on two datasets; thereby identifying the suitable triple set to be used for developing the student attendance recognition application.

The results of Experiment 3 are shown in Table IV where each triple set consists of three values: the learning rate, the batch size, and the number of epochs, respectively.

For the CelebA-Spoof dataset, three different triple sets are tested. The first set (0.005, 128, 20) shows relatively

Table IV
THE PERFORMANCE OF PROPOSED APPROACH WHEN CHANGING THE VALUES OF HYPERPARAMETERS.

Dataset	Triple set	Accuracy	Precision	Recall	F1-Score
CelebA Spoof	(0.005, 128, 20)	68.17%	68.17%	100.00%	81.07%
	(0.002, 64, 30)	78.52%	78.37%	95.29%	83.73%
	(0.0001, 32, 50)	99.48%	99.72%	99.58%	99.65%
CUSC Student	(0.005, 64, 10)	61.57%	56.73%	100.00%	75.73%
	(0.0005, 32, 15)	97.05%	95.56%	100.00%	72.39%
	(0.0001, 32, 20)	95.46%	100.00%	90.60%	95.07%

low performance with an accuracy of 68.17% and a F1-Score of 81.07%. The second set (0.002, 64, 30) performs better, achieving an accuracy of 78.52% and a F1-Score of 83.73%. The third set (0.0001, 32, 50) delivers the best results among 3 sets with an accuracy of 99.48% and a F1-Score of 99.65%. This triple set achieves superior performance regarding accuracy and F1-Score compared to other triple sets.

For the CUSC-Student dataset, the results show that the triple set (0.005, 64, 10) achieves lower performance compared to the other two triple sets. In contrast, the triple set (0.0001, 32, 20) provides the highest performance regarding accuracy and F1-Score, but requires a longer training time. However, the triple set (0.0005, 32, 12) can be considered because it offers a good balance between the model's performance and training time.

The experiment demonstrates that the configuration with a learning rate of 0.0001 and a batch size of 32 consistently performs well across both datasets.

V. CONCLUSION

This paper proposes a face recognition approach combined with anti-spoofing to enhance organizational security. The proposed approach performs well with the MobileNetV3 Large model both on the CelebA-Spoof dataset and CUSC-Student dataset. Real-time applications of this proposed approach can be utilized in various settings, such as schools or universities for student attendance, offices or companies to ensure that only authorized personnel can enter. However, this approach has certain limitations. For example, it may still misidentify when an attacker uses high-quality images, especially as face spoofing and deepfake technologies continue to evolve.

In the future, we will deploy a student attendance system in practice where the networked, high-resolution, motion-tracking cameras will be placed at common locations such as the classroom entrances. We will then conduct experiments to determine the appropriate distances between the camera and the learner, and the suitable lighting conditions to accurately identify the learner. The proposed approach could be improved to enhance face recognition capabilities in low light, uneven lighting, or continuously changing conditions using adaptive lighting algorithms and image enhancement technologies. Additionally, more complex algorithms could be developed to detect and differentiate between real and fake faces created by deepfake technology. This includes utilizing advanced deep learning models to analyze detailed facial features and behaviors.

REFERENCES

[1] Almabdy, S., & Elrefaei, "Deep Convolutional Neural Network-Based Approaches for Face Recognition", *Applied Sciences*, 9(20), 4397, 2019. doi:10.3390/app9204397.

[2] Samira, P., Saad, S., Yilin, Y., Haiman, T., Yudong, T., Maria, P. R., Mei-Ling, S., Shu-Ching, C., & Iyengar, S. S, "A Survey on Deep Learning: Algorithms, Techniques, and Applications", *ACM Computing Surveys*, 51, 1-36, 2018. doi:10.1145/3234150.

[3] Jones, M., & Lee, S, "An Introduction to Convolutional Neural Networks", *Journal of Artificial Intelligence Research*, 37(2), 45-67, 2021.

[4] Li, Y., & Liu, Y, " Deep Learning for Face Anti-Spoofing: A Survey", *IEEE Access*, 8, 105383-105404, 2020. doi:10.1109/ACCESS.2020.2997350.

[5] Xu, X., Li, W., & Zhang, Y," Face Anti-Spoofing Using Patch and Depth-Based CNNs", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1-6, 2018. doi:10.1109/CVPRW.2018.00120.

[6] Li, K., Yang, H., Chen, B., Li, P., Wang, B., & Huang, D ,"Learning Polysemantic Spoof Trace: A Multi-Modal Disentanglement Network for Face Anti-Spoofing", *arXiv preprint arXiv:2212.03943*, 2022.

[7] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. ,"DeepFace: Closing the Gap to Human-Level Performance in Face Verification", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1701-1708, 2014. doi:10.1109/CVPR.2014.220.

[8] Redmon, J., Divvala, S. K., Girshick, R. B., & Farhadi, A," You Only Look Once: Unified, Real-Time Object Detection", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788, 2015

[9] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, " Going deeper with convolutions", *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-9, 2014

[10] He, K., Zhang, X., Ren, S., & Sun, J. ,"Deep Residual Learning for Image Recognition", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778, 2015

[11] Simonyan, K., & Zisserman, A. ,"Very Deep Convolutional Networks for Large-Scale Image Recognition", *CoRR*, 2014.

[12] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. ," MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", *ArXiv*, 2017.

[13] Sandler, M., Howard, A. G., Zhu, M., Zhmoginov, A., & Chen, L.. " MobileNetV2: Inverted Residuals and Linear Bottlenecks", *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4510-4520, 2018.

[14] Howard, A. G., Sandler, M., Chu, G., Chen, L., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. ," Searching for MobileNetV3", *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 1314-1324, 2019.

[15] Ravpreet, K., & Sarbjeet, S. ,"A comprehensive review of object detection with deep learning", *Digital Signal Processing*, 132, 103812, 2023.

[16] Zaidi, S. S., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M. N., & Lee, B. ," A Survey of Modern Deep Learning Based Object Detection Models", *Digital Signal Processing*, 126, 103514, 2021.

[17] Deng, J., et al. ,"ArcFace: Additive Angular Margin Loss for Deep Face Recognition", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4690-4699, 2019. doi:10.1109/CVPR.2019.00482.

[18] Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P. E., & Jégou, H. ," The faiss library", *arXiv preprint arXiv:2401.08281*, 2024.

[19] *CelebA-Spoof Dataset*. Retrieved from https://mmlab.ie.cuhk.edu.hk/projects/CelebA/CelebA_Spoof.html

[20] Salmon, B. P., Kleynhans, W., Schwegmann, C. P., & Olivier, J. C. ," Proper Comparison among Methods Using a Confusion Matrix", In *IGARSS 2015*, 2015.

Minh Nguyet Nguyen received with a B.Eng. degree in Data Communication and Computer Networks from Can Tho University,(CTU), Vietnam, in 2022 and is currently pursuing an M.Eng. degree in Information Technology at Can Tho University. Email: nmnguyet@cusc.ctu.edu.vn

Phuong Lan Phan obtained a Master's degree at Asian Institute of Technology, Thailand (2003). She obtained her Ph.D in Computer Science from the University of Science and Technology - The University of Da Nang, Vietnam (2020). Now, she is a lecturer in the Faculty of Software Engineering. Her research interests include Software Processes, and Recommendation Systems. Email: pplan@cit.ctu.edu.vn