Hindawi
Security and Communication Networks
Volume 2021, Article ID 7314823, 11 pages
https://doi.org/10.1155/2021/7314823

Research Article

Smart Identity Management System by Face Detection Using Multitasking Convolution Network

Lubna Farhi, Hira Abbasi, and Rija Rehman

Department of Electronic Engineering, Sir Syed University of Engineering and Technology, Karachi, Pakistan

Correspondence should be addressed to Lubna Farhi; lubnafarhi@yahoo.com

Received 25 June 2021; Revised 31 October 2021; Accepted 24 November 2021; Published 21 December 2021

Academic Editor: Ahmad Samer Wazan

Identity management system in most academic and office environments is presently achieved primarily by a manual method where the user has to input their attendance into the system. The manual method sometimes results in human error and makes the process less efficient and time-consuming. The proposed system highlights the implementation and design of a smart face identification-based management system while taking into account both the background luminosity and distance. This system detects and recognizes the person and marks their attendance with the timestamp. In this methodology, the face is initially resized to 3 different sizes of 256, 384, and 512 pixels for multiscale testing. The overall outcome size descriptor is the overall mean for these characteristic vectors, and the deep convolution neural network calculates 22 facial features in 128 distinct embeddings in 22-deep network layers. The pose of the 2D face from −15 to +15° provides identification with 98% accuracy in low computation time. Another feature of the proposed system is that it is able to accurately perform identification with an accuracy of 99.92% from a distance of 5 m under optimal light conditions. The accuracy is also dependent on the light intensity where it varies from 96% to 99% under 100 to 1000 lumen/m^2, respectively. The presented model not only improves accuracy and identity under realistic conditions but also reduces computation time.

1. Introduction

In many public and educational sectors, the management system is mandatory for analyzing the performance of candidates. When there are a lot of individuals in an organization or institute, it becomes significantly more difficult to mark their presence through the manual procedure and it is also time-consuming. The conventional marking method is obsolete, and in such systems, identification is recorded with traditional approaches that include registers and sheets whereas more advanced methods like RFID and biometric encounter the difficulty of time wastage and are significantly more complicated where you have to wait in line to swipe the RFID card or put your thumb on a scanner which can be a quick way of spreading unwanted diseases. It is also prone to manipulation where individuals can mark the presence of others without any oversight if they possess the RFID card. Sorting and calculating the attendance of every enrolled person is not only tiresome but humans can easily make mistakes while conducting repetitive tasks. Therefore, a smart system is required for marking and recording. In order to do that, we will save an authentic and proper record of persons that can also be analyzed later on if needed.

In addition to reducing errors, the proposed system for management is also more feasible than other methods. For example, the biometric system needs more hardware, and its maintenance is also difficult. The automatic system can resolve a crucial issue within the manual one that occurs when a person transfers the information from the sheet into the system. The face identification method has many steps which include capture, extraction, comparison, and matchmaking. An automated and computerized attendance information and management system with enhanced face identification has been proposed. The initial steps include database creation, face identification, feature engineering, and categorization stages followed by the last stage, i.e., postprocessing phase [1]. At first, facial images of every student would be transferred to the system and saved within

the database. Then, the identification of candidates is recorded by using a camera attached within the area at an appropriate location from where the entire region is often viewed or monitored. The camera will constantly take pictures of candidates, identify the countenances in pictures, recognize the identified countenances, and mark their identity. In some methods, the camera is at a fixed position at the point of entry to capture the image of the candidate as they enter that area. Through this technique, we will save more time as compared to the manual management system. Finally, if sorting is needed, then it can also be done easily.

There have been many types of research work on surveillance systems that have been done on various devices; most of them were embedded systems based on GPU as well as FPGA [2]. The most effective and powerful GPUs [3] have been utilized to implement the rules of the monitoring mechanism for standard facial identification and object identification algorithms [4] with an accuracy that was nearly equal to 88 percent [5]. However, in recent years, various deep learning architectures have shown optimistic accuracy. This includes FaceNet [6] which provides nearly 99 percent on a system based on GPU. Face recognition has been achieved through various approaches. Some of them are feature-based, multimodal fusions, holistic appearance, or multispectral-based implemented for face recognition in the infrared spectrum. Early research on infrared imagery for facial recognition was carried out with the introduction of the method based on eigenfaces. They produced a recognition rate of approximately 96% by running a database of 24 subjects with 12 images, each forming a database of 288 thermal images [7]. The base images in this technique represented the variation of posture and face. Different types of improvements in the linear methods, such as eigenfaces, Local Feature Analysis (LFA), Independent Component Analysis (ICA), and Linear Discrimination Analysis (LDA), significantly improved the precision of the images in thermal and visible data-based face recognition. More specifically, the increase in precision was much higher in the thermal spectrum (about 93% to 98%) than in the visible spectrum. However, despite the improvements, harmful variables persisted in the image databases that have the potential for increasing bias and skewing the result. Similarly, holistic appearance approaches take into account the picture of the face. Such a methodology is unique in the way facial images work and helps to treat faces in a different way from other categories discussed in [8]; therefore, it is not responsible for the independent processing of functions. This approach has been used by various researchers in facial recognition using infrared imagery. Some have investigated the potential of infrared imagery for facial recognition by extraction of a significant shape called "elemental forms," and the structure of these elemental forms was similar to fingerprints [9]. A methodology built on a general Gaussian mixture model uses the Bayesian approach to select the parameters from a sample image [10]. However, this research resulted in an approximately 95% facial identification accuracy in the thermal and appearance-based data. Another facial recognition approach was created with a database of 50 people along with 10 images per individual, which offered authentic evidence for facial recognition within the IR spectrum [11]. In this research, the data considered for classification did not include varieties of intrapersonal variables which is due to different emotional states, or exercise, or even the temperature of the air which is a major drawback in this set of research. The classification approach depends on the results based on a combination of neural networks and local averted appearances and extracts the characteristics of thermal images proposed [12]. The approach was carried out at an ambient temperature from 302 K to 285 K and achieved a recognition rate of 95% when the test and training data were entered at an identical ambient temperature. Over this period, the approach achieved the closest rate of 60 percent for recognition when the difference between training and sample data temperature remained at 17 K. The multimodal method approaches transform coded greyscale projections, eigenfaces, and pursuit filters for matching the pictures in the research of [13]. One depends on the level of data and the other on the decision-making lever. In this method, characteristics are built up by inheriting the data from the two modalities and then further classified. While within the decision level, the precision of two-person matching within the ROI and visible spectrum is computed, which makes the model more complex. Another problem in face recognition is time-lapse; i.e., the performance of an algorithm decreases as time passes between training and test data without taking into account the scanning conditions. Similarly, using the effects of atmospheric temperature on facial temperature and improving the image to standardize the facial regions is studied in [14, 15]. It shows that facial recognition errors produced in the visible spectrum and infrared spectra are affected by the passage of time between sampling and the acquisition of the test data. Later on, it had been observed that face recognition performance decreases due to changes in certain tangible factors that have an impact on the appearance of the face and in particular on the thermal data [16].

The Convolutional Neural Network (CNN) consists of a combination of convolutional layers, nonlinear layers (e.g., mean, max, or min), and classification layer units. In some cases, this methodology can be used to identify a category of a dog class, a car model, or birds, resulting in these structures having advanced potential outcomes [17]. Nowadays, most researchers are using the *Multitask Cascaded Convolution Neural Networks (MTCNN) algorithm* for facial detection and classification due to its robust nature [18, 19]. Similarly, some existing techniques and their weaknesses are discussed in Table 1.

Facial recognition approaches are hindered by many exigent challenges that include opacity of obstructions between the camera and the subject, the environmental light intensity levels, surrounding atmospheric conditions, the distance between camera and subject, and lastly but not limited to the emotional and physical expression of the subject. Moreover, most appearance-based methods supplement their analysis with complex statistical techniques only to provide specific insight into the analysis instead of a holistic understanding of the outcomes. Present research models fail to incorporate multiple aforementioned factors into their studies and often aim to optimize their models

Table 1: Existing techniques and their limitations.

Author	Title	Model	Weaknesses
Faruk et al. [20]	"Image to Bengali Caption Generation Using Deep CNN and Bidirectional Gated Recurrent Unit"	Deep Face	When the architecture of CNN was made hybrid with multifarious detectors, it limited the variant part of an object in remarkable identification systems. However, this methodology can only be used to identify a few categories of pictures.
Alimuin et al. [21]	"Deep hypersphere embedding for real-time face recognition"	Hybrid combination of MTCNN and FaceNet Framework of deep cascading with Neural Network	Targets only an aspect of constraints in facial features and face recognition with low accuracy.
Jin et al. [22]	"Face recognition based on MTCNN and FaceNet"		Fails to consider surrounding intensity levels and distance between camera and face.
Xiao et al. [23]	"The Improvement of MTCNN ALGORITHM: Face Recognition with Mask"	MTCNN	This model does not consider the environmental light intensity levels, surrounding atmospheric conditions, and the distance between camera and subject.

based on one variable such as improving accuracy based on surrounding light intensity levels or distance between camera and subject, but not both. As a result, their models are only accurate under specific conditions and are not pragmatic as they fail to include the interrelated factors between these variables.

Keeping these shortcomings in consideration, this study provides a novel approach where surrounding light intensity levels, angle of the facial image acquired, and distance between camera and subject are incorporated in the design of the model. This model is then optimized to not only improve accuracy under realistic conditions but also reduce computation time through postprocessing, feature extraction, and Multitasking Cascaded Convolution Neural Networks (MTCNN) algorithm.

This paper is organized into five sections. Section 1 provides an introduction and related work. Section 2 highlights the mathematical modules. Section 3 discusses the implementation of the proposed management system and experimental results in which the performance of the proposed algorithm is evaluated and the results are shown. Section 4 presents the conclusion.

2. Methodology

There are four main modules to this proposed system that are as follows: detecting a face from a real-time stream, extracting countenance, recognizing the face, and providing the countenances.

2.1. Dataset Creation. This first and foremost step to creating a self-based face recognition dataset for an in-house facial recognition system is that physical access to specific individuals is needed to collect sample images of about 126 faces. It would be a typical system for schools, companies, or other organizations where people are physically present themselves on a daily basis. To gather footage of these individuals, we can perform two methods:

(1) Escort them into a special room where a camera is installed. Taking pictures of the person from different angles stores the picture of that person in a labeled directory.

(2) Implement different systems for the different rooms.

In this proposed system, the dataset creation process is going to be implemented at the time of registration of a new student on campus. 10–20 pictures of every student will be taken on-site while creating a brand-new directory for students with reference to their department batch and section and storing the images of the students inside it. At the time of training, we have to settle on a section-based training method where the encoding of every student of the respective section is stored. The dataset contains the three subsequent directories: database directory contains the whole database of the system; i.e., persons, timetables, and student information are also available as attendance records. Next, encodings contain all the encoding files. Similarly, the models have the model file of the system and results; this subdirectory contains all the face recognition testing results such as pictures and videos that contain Labeled Faces, specifically used in testing at the time of training.

2.2. Image Acquisition and Preprocessing. Acquiring images is the first key step in the face recognition method and it is the main phase of any vision system. After the image has been acquired, different strategies for handling are often applied to the image to play out the varied vision assignments required today. In any case, within the event that the image is not recognized, the planned targets might not be achievable even with the guide of any sort of image enhancement. After the image acquisition, the captured frame is passed through the multitasked cascaded convolution neural network which reduces the unwanted features and returns a cropped face. The algorithms applied to normalize the cropped face further are as follows [24, 25]:

$$X_{\text{Normalized}} = \frac{X_{\text{mean}} - X_{\text{min imum}}}{\left(X_{\text{max imum}} - X_{\text{min imum}}\right)}. \quad (1)$$

The mean of the cropped face feature vector is taken which is further subtracted by the minimum of the cropped

face feature vector. The result is then divided by the range of the cropped face feature vector to produce the final result. The resultant normalized face is finally resized to 160×160. Finally, the key points and the bounding box are placed on the original image. Figure 1 represents the flow of the input frame to the detected face.

2.3. Feature Extraction. FaceNet is used as the beginning of the technique for facial recognition [6], identifying, checking, and grouping neural networks for the system. This pretrained FaceNet model contributes as a network that is associated with a layered batch and an extremely deep convolution neural network. The deep convolution neural network is supported by L2 normalization where the integration of the face is the outcome of that standardization. The face embedding is carried out during training with a triplet loss. In case the characteristics are alike, then the loss of triplet will have the lowest distance between good and bad facial points. FaceNet consists of twenty-two deep network layers, whose output is trained over these deep layers directly to achieve a facial feature in 128 distinct embeddings. Once rectified, the completely connected seam will serve as a size descriptor which is converted into a descriptor based on commonality using the embedding module to prepare a distinct feature vector from a given template. The maximum operator has been applied to these features. For specific facial recognition, classification, and verification tasks, the network needs to be refined to anticipate a significant boost. Figure 2 represents the FaceNet architecture.

2.4. Face Detection and Reduction. The detection of facial features from a provided image is a critical task when it comes to facial identification. Without having pictures of variant faces, work will not proceed. An MTCNN is used to identify and bring actual face parts from a given picture in a position to beat multifarious face recognition standards offering real-time performance with high precision studied [26]. In this system, a pretrained model (MTCNN) is used to find the candidate's face in part of the image and interpret it into greater feature facial descriptors [27].

2.4.1. Face Judgement. The initial model resizes the picture to a special degree of size in such order which gradually increases from 12×12 to 256×256, and it is known as a picture pyramid. The subsequent facial portions are presented by the key network, called the proposal network.

The learner target may be a bipartite issue for each sample x_i which uses the cross-entropy loss function:

$$L_i^{\text{det}} = -\left(y_i^{\wedge \text{det}} \log(p_i) + \left(1 - y_i^{\text{det}}\right)(1 - \log(p_i)\right), \tag{2}$$

where the probability of face sample x_i is represented by p_1 which is predicted by the MTCNN. y_i^{det} stands for ground-truth; $y_i^{\text{det}} \in \{0, 1\}$ [26].

2.4.2. Enhancing Image Qualities. R-Net or refine network sharpens limiting boxes. For the applied window, the offset (such as the width, the height, and top-left coordinate) between them and the nearest earthly truth is predicted. The loss function is the square loss function:

$$L_i^{\text{box}} = \left\| y_i^{\wedge \text{box}} - y_i^{\text{box}} \right\|_2^2, \tag{3}$$

where the regressed target $i^{\wedge \text{box}}$, y_i^{box} is the ground-truth 4-dimensional coordinate, including the width, the top-left coordinate, and the height. The R-Net property contains many types of information tagged with relevance, such as expression, blur, invalid, illumination, pose, and occlusion.

2.4.3. Feature Location. O-Net or output network serves as the final network which determines facial landmarks from the given image which is alike to bounding box regression. The loss function is as follows:

$$L_i^{\text{landmark}} = \left\| y_i^{\wedge \text{landmark}} - y_i^{\text{landmark}} \right\|_2^2. \tag{4}$$

Likewise, the regressed feature coordinate from the network is represented as $y_i^{\wedge \text{landmark}}$. The ground-truth contains five coordinates: two corners of the mouth, two eyes, and the nose represented by y_i^{landmark}.

As the dataset for training is different for disparate tasks over the course of training, during training, the loss of another task's training should be zero. Thus, the combination loss function should be as follows:

$$\min \sum_{i=1}^{n} \sum_{j \in \{\text{det}, \text{box}, l\}} \alpha_i^j \beta_i^j L_i^j, \tag{5}$$

where the amount of the training samples is represented by n and the significance of each task is α_j. In P-Net and R-Net, $\alpha_{\text{det}} = 1$, $\alpha_{\text{box}} = 0.5$, and $\alpha_{\text{landmark}} = 0.5$, while in O-Net, for gaining high accuracy face coordinates, the parameters are $\alpha_{\text{det}} = 1$, $\alpha_{\text{box}} = 0.5$, and $\alpha_{\text{landmark}} = 0.5$. β_j^i is the sample type indicator.

These networks can do facial recognition, bounding box regression, and facial landmark tracking which is why they are known as multitasking networks. These networks are cascaded as a result where various stages are taken into account with additional processing. NMS is applied in MTCNN and is used to refine the boundaries of the applicant by refine network and output network prior to delivering output. This facial detection method has numerous advantages over various poses, visual variations, and the lighting conditions of the face. Figure 3 shows the output result of the picture passed from MTCNN.

2.5. Face Recognition. An identification approach is utilized to make sure the candidate faces the task of classification with a Support Vector Machine (SVM) categorization-related problem since it was highlighted. Matching or classification tasks of a refinement problem are solved by a Support Vector Machine [28]. It increases the boundary between the classes within given input-target entries. The classifier is the result of a specific level of robustness at overfeeding. The range represents the effectiveness of class separation. SVM finds the optimal separation of closest points in the training set. This

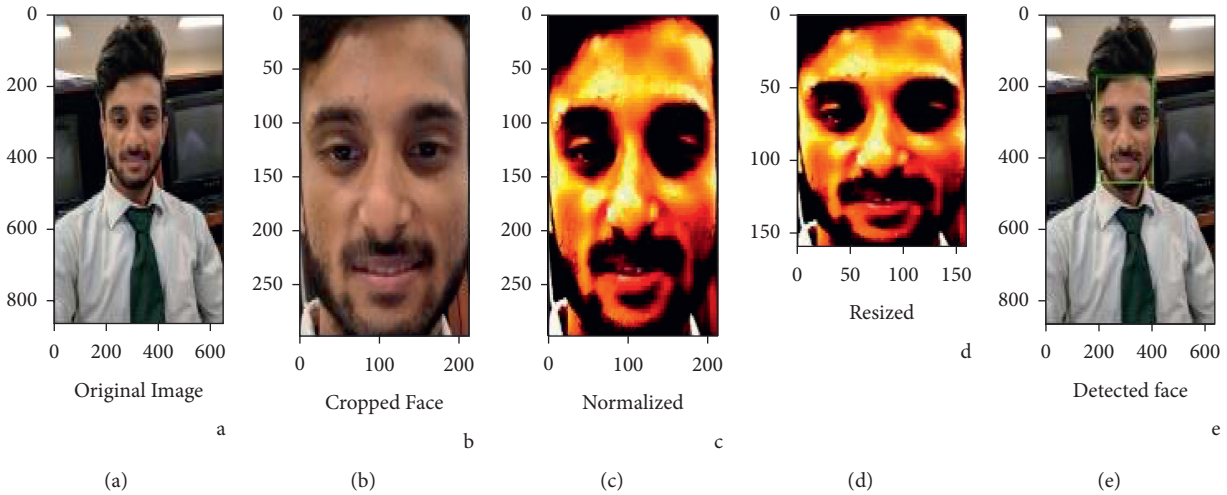

Figure 1: (a) Original face image. (b) MTCNN algorithm minimized the unwanted features and returns cropped face. (c) Normalized cropped face. (d) Resized image to 160 × 160. (e) Face recognized with Bounded Box.

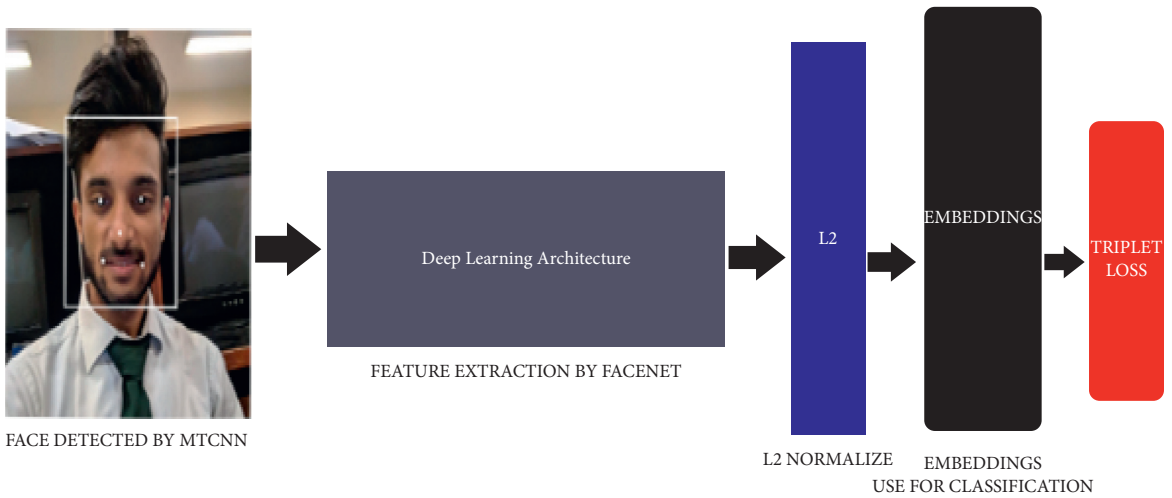

Figure 2: The FaceNet architecture receives the input image after the face has been detected by MTCNN, where the deep learning architecture is used to extract the best features followed by L2 normalization. The result of this architectural processing is facing embeddings that are forwarded to the triplet loss function during training.

separation can be done linearly or nonlinearly both. The proposed methodology compared the test face to other faces using the Support Vector Machine. The result is deemed correct if the distance between the image which we train, and the test of the same person is kept to a minimum. A facial resemblance is measured on the image which we have input and the image of the face formed by estimating a level 2 normalization in the characteristics of the specific points gathered from the network structure.

3. Proposed Management System

This process takes the recognized face which is delivered from face identification utilizing SVM. The name of the face owner will be marked present in our database with the current timestamp by the interfacing of Python with SQLITE3. The number of faces will be obtained through face detection which will be used at the cohort level or individually for the management system. Figure 4 shows the approach of face recognition-based management system as explained in Section 3

The proposed deep net topology is simpler than prior models, and this allows that net to be extended into a deeper network in a straightforward way as shown in Figure 5.

3.1. Experimental Results. In the following, we initially tested the models and instructional data for the specific dissimilarity of the suggested models and instructional information using the Labeled Faces in Wild Dataset to verify performance. Next, we compared the performance of configuration with leading-edge methods on LFW (Labeled Faces in Wild Dataset). Our implementation is predicated in Python related to public libraries of NVIDIA CuDNN to boost the training. All experiments were carried out on NVIDIA GTX 1650 with 4 GB of onboard memory. This is often important

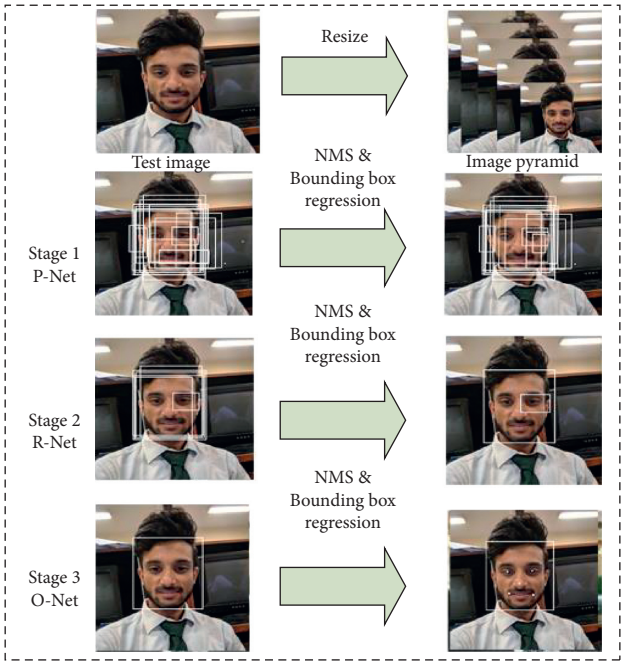

Figure 3: Face detection with facial landmarks.

due to the limited memory footprint and great complexity of very deep networks. For multiscale testing, the face is initially resized to 3 different sizes of 256, 384, and 512 pixels. Accordingly, the cropping method was repeated for each of them. The overall outcome size descriptor is the overall mean for these characteristic vectors. Faces are identified with the help of the methodology explained in "FaceNet: a unified embedding for face recognition and clustering.

Figure 6 represents the vector of 128 numbers which represent the most important features of each tested face. This vector is further converted into 128 distinct embeddings using L2 distance measures between the test faces which will be used to identify the test subject. This means that, for example, once Rajal's image is taken and converted into the facial feature vector, if the vector has a small distance with his distinct embeddings (prestored), then it signifies that his face has been identified. However, if, for example, Hamza's facial feature vector has a larger distance from the distinct embeddings based on the measures, then it means that his face has not been identified and another picture needs to be taken. The figure shows the difference in the embeddings between the 4 detected faces of our dataset which have been used as the input to our classifier model.

The identification accuracy from the distance of 1 m to 5 m from the camera on different scenarios is shown in Table 2.

For identifying, the detected faces are first converted into 128 distinct embeddings. The correlation of these embeddings is taken to represent the distance for recognition in which the Minimum Threshold of the distance is taken as 0 while the maximum is taken as 0.6 (60%) out of 1 (100%). It can be seen, in Figure 6, that for Sufyan, Waqas, and Hamza, the distance is within the "Maximum Threshold" and "Total Distance" bounds, thus indicating that their faces have been detected with an adequate level of certainty. However, for Rajal, the certainty to which his face has been detected is low since it is below the "Maximum Threshold" despite being above the "Minimum Threshold." This can be attributed to various aspects such as a blurry image or a major difference between the facial features of the taken and reference image. If the embedding is within threshold bounds, then it indicates that the image has been identified to a certain degree as shown in Figure 7.

Table 3 represents the distance refers to the dissimilarities of the detected image and image stored inside the database.

Table 4 provides the accuracy of detection from various angles. It represents face capturing angle which is one of the causes of face recognition. The results suggest that when pose variation is within 15°.

Figure 8 represents the similarities between the embeddings of the recognized face which were stored in the database and the input face which we acquired through real-time image acquisition.

The result shows that the pose variation at different angles has weight to improve the accuracy. Figure 9 shows that the pose of the face from −15 to +15° provides high accuracy since in that position the features of the face can be detected easily, and face recognition is better as compared to other poses.

In this system, a pretrained model (MTCNN) is used to find the candidate's face in part of the image and interpret it into greater feature facial descriptors. This network works as cascading in this model as shown in Figure 10.

This model is then optimized to not only improve accuracy under realistic conditions but also reduce computation time through postprocessing, feature

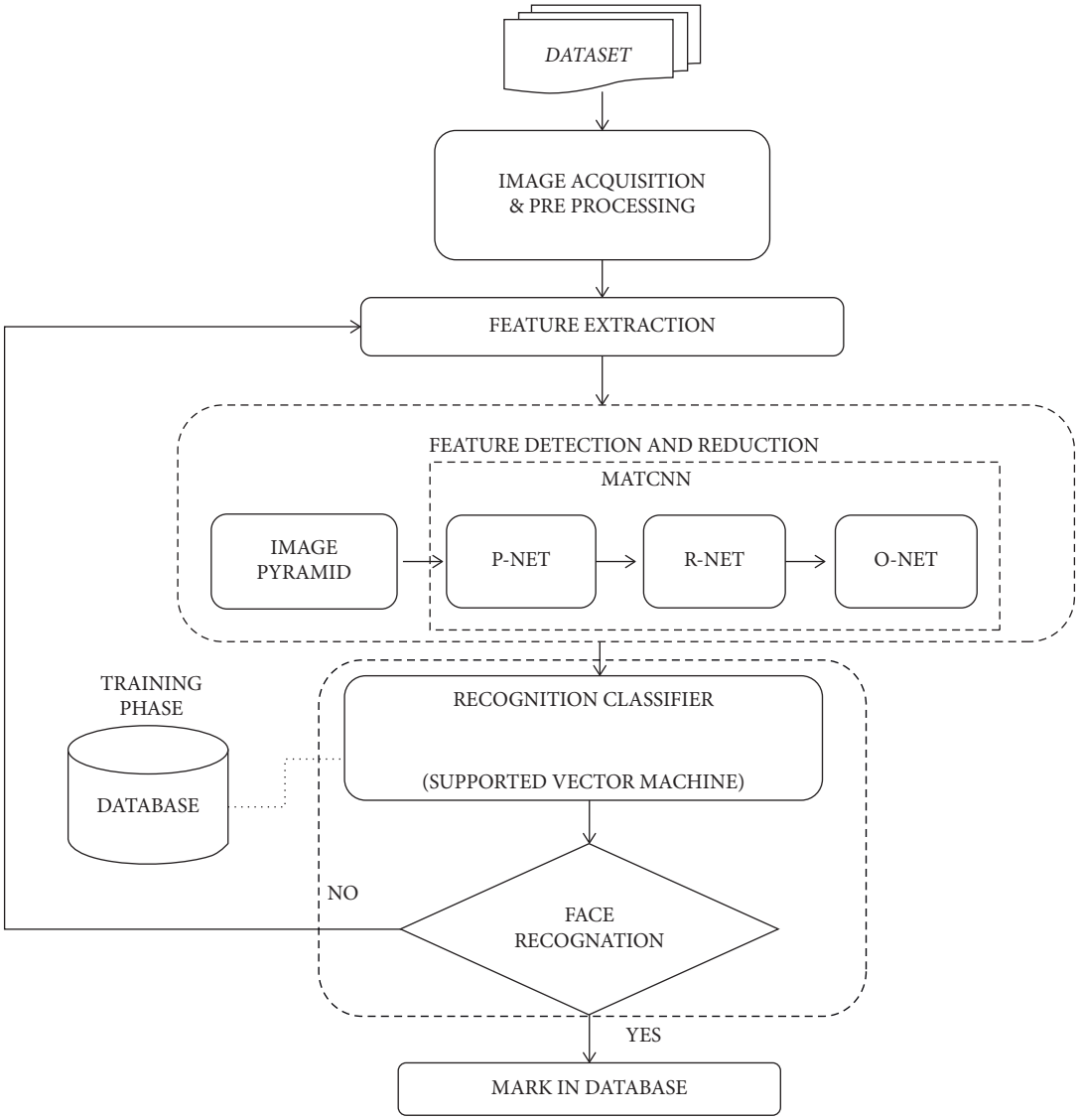

Figure 4: Flow of Smart Identity Management System by face detection.

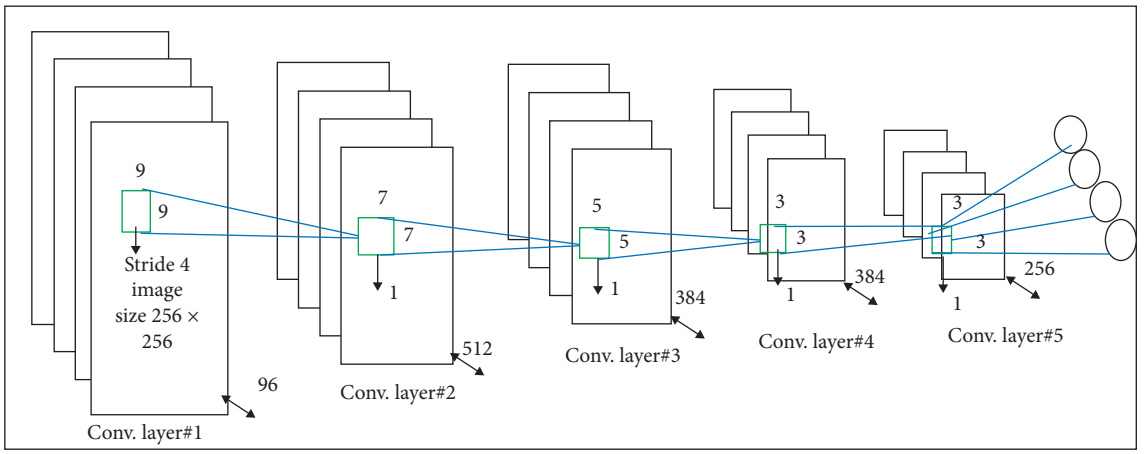

Figure 5: Architecture of deep net: high-order local and global features are learned in an image through multilayer architectures. Kernel size and quantity along with kernel combinations between layers determine the learning capacity of the networks.

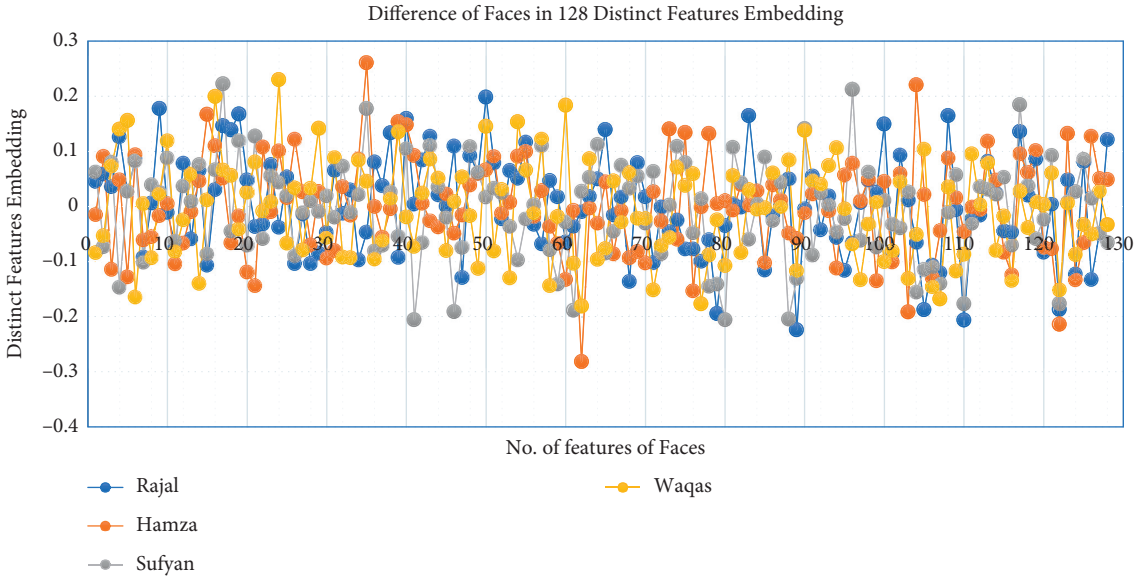

Figure 6: 128 distinct feature embeddings.

Table 2: Face Recognition Accuracy test from distance.

Scenario	Illuminance (lx, lumen/m^2)	Accuracy distance (1 m)	Accuracy distance (3 m)	Accuracy distance (5 m)	Avg. Accuracy (%)
Low light	100	99.630	97.88	92.0	96.47
Normal light	300	**100**	**99.97**	**99.92**	**99.96**
Bright	1000	99.00	99.00	94.00	97.33
Dark	>10	54.00	49.76	44	49.253

Bold values mean reference results at normal light scenario and are used as a comparison for other scenarios.

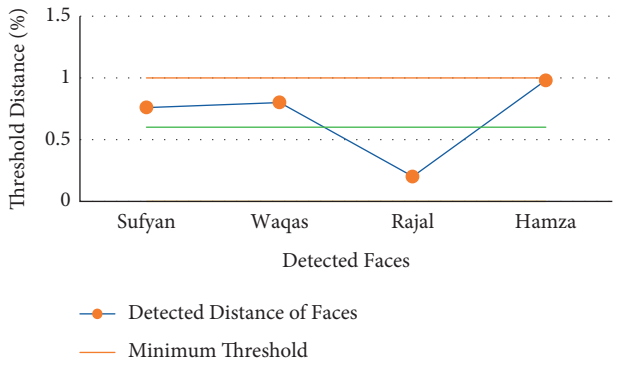

Figure 7: Recognition between threshold distances.

Table 3: Actual Distance from Total Distance for recognition.

Persons	Sufyan	Waqas	Rajal	Hamza
Actual Distance between encoding (m)	0.746564	0.80876	0.225128	0.941572
Total Distance (m)	1	1	1	1
Maximum Threshold	0.6	0.6	0.6	0.6
Minimum Threshold	0	0	0	0

extraction, and Multitasking Cascaded Convolution Neural Networks (MTCNN) algorithm. As a result, the computation time decreased to 0.073–0.40 s for facial recognition and identification. We also utilized a method of having unified face image representation necessary for better recognition of face images. The system provides low-cost memory storage and has data logging features and low maintenance.

Table 4: Performance comparison from various angles.

Pose (degree)	−45	−30	−15	15	30	45
Accuracy (%)	84.5	95.2	**97.1**	**98.8**	95.8	89

Bold values mean good accuracy results at above mentioned angles.

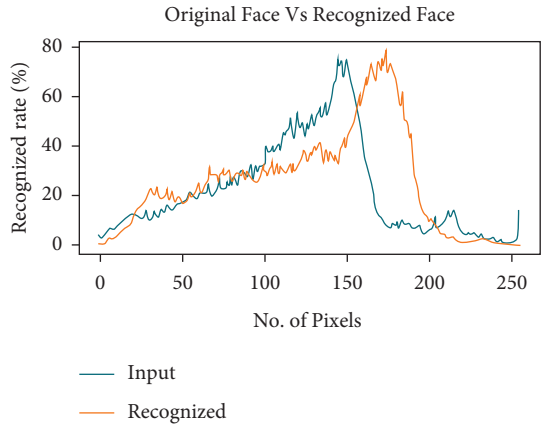

Figure 8: Detected face versus recognized face.

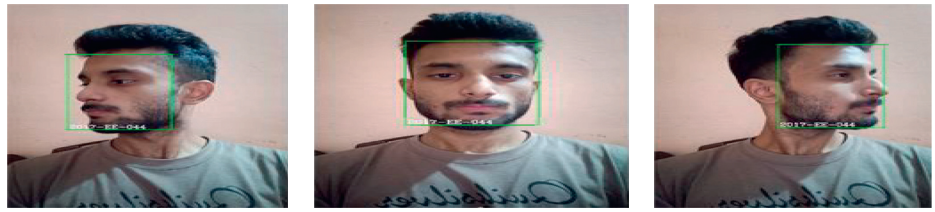

Figure 9: Recognition from varied angles.

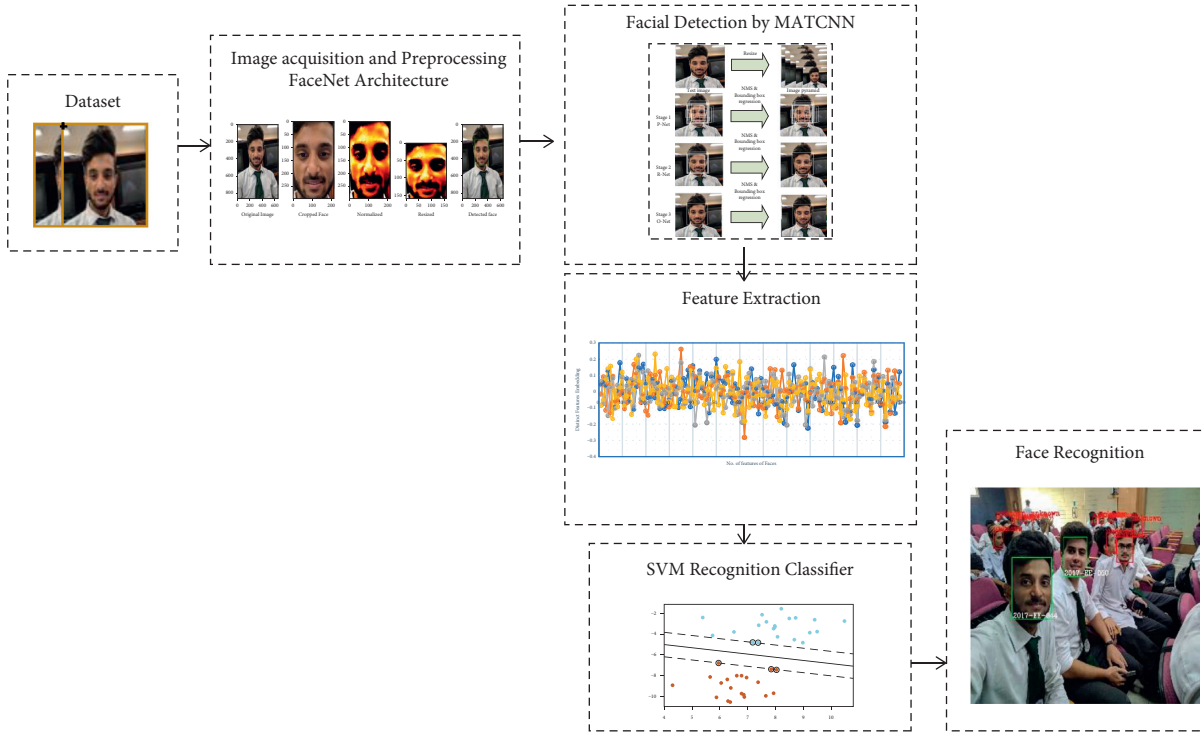

Figure 10: Structure of the deep network of Smart Identity Management System by face detection.

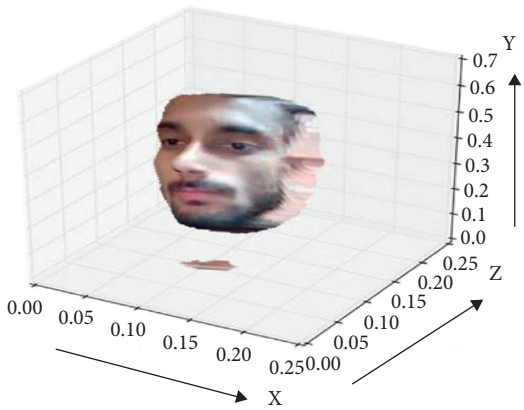

Figure 11: Face plotted in 3-dimension.

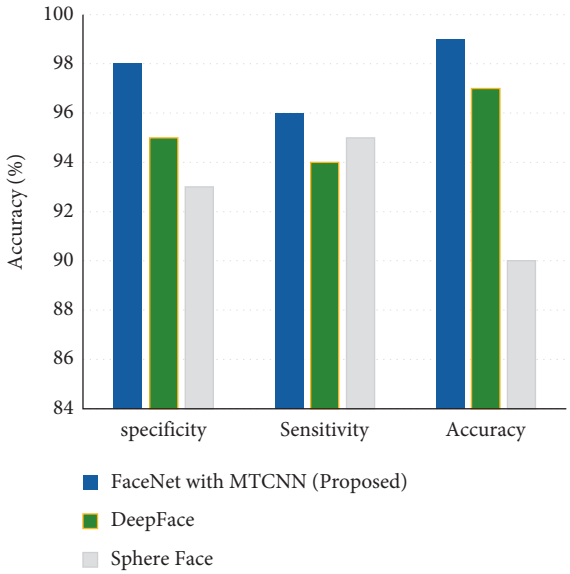

Figure 12: Performance test with other algorithms.

The result of our dataset performed really well in recognition from various angles. Adding additional images which were captured from the various angles makes our system recognition look similar to 3D recognition where a 3D model of a sample face is projected in Figure 11. It shows the actual projected image of the face in the x, y, and z axes.

Varying the algorithm on the system provides minimal variations to the performance of the system. Figure 12 shows the system's performance in terms of its accuracy, sensitivity, and specificity using the Deep Face, Sphere Face, and Proposed FaceNet with MTCNN hybrid Net. The proposed Net adapts to the performance of the system by having a standout result compared with the two other algorithms.

4. Conclusion

This proposed approach has the idea of implementing a smart system that is able to identify the face in real time while taking both luminosity and distance into account. It has been implemented on the features which are successfully able to get results with 97.1% to 98.8% accuracy when the position of the face is in the −15° to +15° range while other positions provide average results. This issue can be improved if we train our database with 3D images taken by a 3D scanner or camera which will boost the performance of recognition and increase the range of which algorithm can identify faces accurately. This identification system is able to recognize a face accurately with 99% to 98% accuracy when the distance of face is 4-5 meters away from the camera under normal light conditions. Additionally, under low light conditions, the average accuracy achieved under low light conditions is 96.47%. The recognition range can be increased by using a high-quality camera that is capable of HD imaging. The approach was also compared with two other industry-standard architectures used for facial detection and it was observed that FaceNet with MTCNN had the highest performance in terms of specificity, sensitivity, and accuracy. In the proposed methodology, postprocessing of images with feature extraction and algorithms reduced the face detection computational time and improved the face identification accuracy. The system was able to accurately identify faces from distances up to 5 m and in both high to low light intensities of the local environment.

Data Availability

The authors used third-party data and do not have the rights to share it. The third-party data cannot be publicly shared. Researchers must request to gain access to the data in which case the authors will apply to gain access and share it with them.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Chintalapati and M. V. Raghunadh, "Automated attendance management system based on face recognition algorithms," in *Proceedings of the 2013 IEEE International Conference on Computational Intelligence and Computing Research*, pp. 1–5, IEEE, Enathi, India, December 2013.

[2] A. Ezzahout and R. O. H. Thami, "Conception and development of a video surveillance system for detecting, tracking and profile analysis of a person," in *Proceedings of the 2013 3rd International Symposium ISKO-Maghreb*, pp. 1–5, IEEE, Marrakech, Morocco, November 2013.

[3] L. Zhang, J. X. Wang, and K. Zhang, "Design of embedded video monitoring system based on S3C2440," in *Proceedings of the 2013 Fourth International Conference on Digital Manufacturing and Automation*, pp. 461–465, IEEE, Shinan, China, June 2013.

[4] A. Singh, D. Patil, M. Reddy, and S. N. Omkar, "Disguised Face Identification (DFI) with facial key points using spatial fusion convolutional network," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 1648–1655, IEEE, Venice, Italy, October 2017.

[5] Y. Yu, X. Duan, S. Wang, and B. Jiao, "The design and implementation of Bluetooth video surveillance devices," in *Proceedings of the 2010 3rd International Congress on Image*

and Signal Processing, pp. 495–498, IEEE, Yantai, China, October 2010.

[6] E. Jose, M. Greeshma, M. T. Haridas, and M. H. Supriya, "Face recognition-based surveillance system using Face-Net and MTCNN on Jetson TX2," in *Proceedings of the 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pp. 608–613, IEEE, Coimbatore, India, March 2019.

[7] R. G. Cutler, *Face Recognition Using Infrared Images and Eigen Faces*, University of Maryland, College Park, MD, USA, 1996.

[8] I. Intan, "Combining of feature extraction for real-time facial authentication system," in *Proceedings of the 2017 5th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1–6, IEEE, Denpasar, Indonesia, August 2017.

[9] F. J. Prokoski, R. B. Riedel, and J. S. Coffin, "Identification of individuals by means of facial thermography," in *Proceedings of the 1992 International Carnahan Conference on Security Technology: Crime Countermeasures*, pp. 120–125, IEEE, Atlanta, GA, USA, October 1992.

[10] T. Elguebaly and N. Bouguila, "A Bayesian method for infrared face recognition," in *Machine Vision beyond Visible Spectrum*Springer, Berlin, Heidelberg, Germany, 2011.

[11] Z. Lin, Z. Wenrui, S. Li, and Z. Fang, "Infrared face recognition based on compressive sensing and PCA," in *Proceedings of the 2011 IEEE International Conference on Computer Science and Automation Engineering*, pp. 51–54, IEEE, Shanghai, China, June 2011.

[12] D. Polsgrove and C. Woods, "Biometric verification of visible and MWIR images suitable for optical correlators," in *Frontiers in Optics*Optical Society of America, Washington, DC, USA, 2004.

[13] R. S. Ghiass, O. Arandjelović, A. Bendada, and X. Maldague, "Infrared face recognition: a comprehensive review of methodologies and databases," *Pattern Recognition*, vol. 47, no. 9, pp. 2807–2824, 2014.

[14] S. Moon, S. G. Kong, J. H. Yoo, and K. Chung, "Face recognition with multiscale data fusion of visible and thermal images," in *Proceedings of the 2006 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety*, pp. 24–27, IEEE, Alexandria, VA, USA, October 2006.

[15] X. Chen, P. J. Flynn, and K. W. Bowyer, "IR and visible light face recognition," *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 332–358, 2005.

[16] X. Chen, Z. Jing, and G. Xiao, "Fuzzy fusion for face recognition," *International Conference on Fuzzy Systems and Knowledge Discovery*, Springer, Berlin, Heidelberg, Germany.

[17] C. Zhu, Y. Zheng, K. Luu, and M. Savvides, "CMS-RCNN: contextual multi-scale region-based CNN for unconstrained face detection," in *Deep Learning for Biometrics*Springer, Cham, Switzerland, 2017.

[18] E. Sun, "Small-scale image recognition based on cascaded convolutional neural network," in *Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 2737–2741, IEEE, Chongqing, China, 12 March 2021.

[19] X. Li, Z. Du, Y. Huang, and Z. Tan, "A deep translation (GAN) based change detection network for optical and SAR remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 179, pp. 14–34, 2021.

[20] A. M. Faruk, H. A. Faraby, M. Azad, M. Fedous, and M. Morol, "Image to Bengali caption generation using deep CNN and bidirectional gated recurrent unit," 2020, https://arxiv.org/abs/2012.12139.

[21] R. Alimuin, E. Dadios, J. Dayao, and S. Arenas, "Deep hypersphere embedding for real-time face recognition," *Telkomnika*, vol. 18, no. 3, pp. 1671–1677, 2020.

[22] R. Jin, H. Li, J. Pan, W. Ma, and J. Lin, "Face recognition based on MTCNN and Facenet," 2021, https://jasonyanglu.github.io/files/lecture_notes/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0_2020/Project/Face%20Recognition%20Based%20on%20MTCNN%20and%20FaceNet.pdf.

[23] J. Xiao, J. Wang, S. Cao, and Y. Li, "Research on the improvement of MTCNN ALGORITHM: face recognition with mask," in *Proceedings of the 2021 2nd International Conference on Artificial Intelligence and Information Systems*, pp. 1–7, ACM, Chongqing China, May 2021.

[24] J. Heo, S. G. Kong, B. R. Abidi, and M. A. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition," in *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop*, p. 122, IEEE, Washington, DC, USA, June 2004.

[25] S. G. Kong, J. Heo, F. Boughorbel et al., "Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition," *International Journal of Computer Vision*, vol. 71, no. 2, pp. 215–233, 2007.

[26] M. Ma and J. Wang, "Multi-view face detection and landmark localization based on MTCNN," in *Proceedings of the 2018 Chinese Automation Congress (CAC)*, pp. 4200–4205, IEEE, Xi'an, China, November 2018.

[27] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[28] W. Hizem, L. Allano, A. Mellakh, and B. Dorizzi, "Face recognition from synchronised visible and near-infrared images," *IET Signal Processing*, vol. 3, no. 4, pp. 282–288, 2009.