2023 International Conference on Emerging Smart Computing and Informatics (ESCI)
AISSMS Institute of Information Technology, Pune, India. Mar 1-3, 2023

Real Time Face Recognition Based Attendance System using Multi Task Cascaded Convolutional Neural Network

Vrushaket Chaudhari
Department of Computer Engineering, SCTR'S Pune Institute of Computer Technology,
Pune, Maharashtra, India
vrushu00@gmail.com

Shantanu Jain
Department of Computer Engineering, SCTR'S Pune Institute of Computer Technology,
Pune, Maharashtra, India
shantanujain18@gmail.com

Rushikesh Chaudhari
Department of Computer Engineering, SCTR'S Pune Institute of Computer Technology,
Pune, Maharashtra, India
rushikeshchoudhari2001@gmail.com

Tanvesh Chavan
Department of Computer Engineering, SCTR'S Pune Institute of Computer Technology,
Pune, Maharashtra, India
tanveshchavan2000@gmail.com

Prof. Priyanka Shahane
Department of Computer Engineering, SCTR'S Pune Institute of Computer Technology,
Pune, Maharashtra, India
priyankashahane3696@gmail.com

2023 International Conference on Emerging Smart Computing and Informatics (ESCI) | 978-1-6654-7524-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ESCI56872.2023.10099879

***Abstract*— Facial recognition has been an important research direction in computer vision. There are countless algorithms presented in related disciplines, and the precision that may be achieved is increasing. However, the implementation of facial recognition technology is hard. In this paper, combination of facial recognition and facial recognition algorithms to build a video-based facial recognition system to efficiently and accurately mark participant attendance. Utilizing FaceNet to extract characteristics and use MTCNN to detect the image of the student for recognition. Lastly, the output is analyzed by a Support Vector Machine (SVM) that recognizes the person of interest in the image. Studies reveal that this technique still yields accurate detection results when the dependent variable has no data and the image quality is unreliable. On the self-generated data set used in this article, the accuracy of the procedure may reach 94.85%.**

Keywords*— *Face detection, Face recognition, MTCNN, FaceNet, SVM.

I. INTRODUCTION

Today, attendance systems are very important in companies, schools, governments, and other places where human resource management is required. The presence of fingerprints requires a queue for identification, which takes a lot of time. Injury to the finger can greatly reduce the accuracy of fingerprint recognition, and fingerprints can also be forged by others. Scanning an ID card for attendance does not verify the identity of the cardholder, which also leads to fraudulent attendance behaviour. Since iris detection is used for presence, the detection speed is slow and the time and equipment costs are high. Location-based attendance checks on mobile phones are like scanning an ID card, but cannot verify the user's identity, and the location can be spoofed.

With the continuous development of machine learning and artificial intelligence technology, the methods of face recognition, facial recognition and facial feature recognition have undergone major changes. As an important biological feature, the human face has been widely used in attendance systems. Dynamic facial recognition technology eliminates the need for users to stop and wait for verification. Users simply appear as part of video surveillance and are automatically recognized by the system. Due to its real-time nature and convenience, this technology has become a hot research target for attendance systems [1]. In this work, a presence system based on multiple facial recognition is designed. To prevent users from using their photos for attendance, the system has the ability to detect blink movements. Finally, for ease of use, a user interaction interface for the attendance system is design.

This project describes the different algorithms used for facial recognition system using videos that can identify a person using his facial features The following work is done in this study to address issues with face recognition systems that use video.

1. The datasets necessary for facial recognition based on video are generated. The method described in this article is used to pinpoint certain video characters. Facial recognition training models was unable to use images from open source datasets since the system needs to detect certain characters. The data set, which consists of facial images is taken from cameras without any target subjects, is used to finish training the face recognition model.
2. Take the video key frames out next. This significantly lowers the computational difficulty of video face recognition, significantly shortens the time it takes to recognize a picture, and accelerates video face recognition. Facial detection using MTCNN is applied to extract faces from the given frames.
3. Adopts a facial feature extraction method based on deep learning. Target faces in videos are recognized by his SVM classifier in this document. To avoid classifier performance degradation caused by category growth, a method is proposed to generate a binary SVM classifier by target. Additionally, the method developed in this article may successfully solve the issues of a lack of relevant sample data and surpass video quality, which are frequent in the real-world use of facial recognition algorithms. Even

with the mentioned problems, accuracy could still be steady.

II. LITERATURE SURVEY

RFID and NFC technologies are commonly used in attendance monitoring systems that rely on unique identification. One such system is AMS, proposed by Rjeib et al., which uses RFID to manage attendance and provide information services [2]. Each student's identifying details and schedule are associated with their RFID tag, which is then recorded in a database and displayed on a web application. Another system called Touch In, developed by Ahmad et al., uses NFC for attendance monitoring [3]. Students can mark their attendance by touching their mobile devices or ID cards with NFC tags on the reader unit, which is connected to a server. The ID-based attendance verification system was later enhanced by Jacob et al. to include one-time password (OTP) technology [4]. This system generates a unique password for each student and sends it to their mobile device once they have been verified by the NFC reader. To complete the process, the student must enter the password through a pre-installed application on their phone. Location-based attendance verification systems often rely on wireless communication technologies, such as Bluetooth and Wi-Fi, to regulate attendance based on communication distance limitations.

Cisar et al. developed an attendance management system that utilized Mobiles and Microcontrollers, which involved students uploading login information to the microcontroller through Bluetooth using Mobile apps [5]. However, managing the wireless signal's coverage made it challenging to restrict the region where attendance was being checked. To address this issue, Abdulkareem et al. incorporated a random forest classifier and NFC (Near Field Communication) technology based on Bluetooth Low Energy (BLE) devices to identify pupils who were inside or outside the classroom [6]. Anand et al. validated a student's identification using a mobile device's camera, and the system only allowed students to be checked for attendance inside the classroom by using Wi-Fi indoor positioning technology [7]. The system employed Wi-Fi fingerprinting technology for indoor locating, and by examining features including device proximity, device movement trajectories, and time sampling, the author of the paper was able to increase the accuracy of both the RSSI gathering process and the total placement. The location-based attendance checking system can also leverage several data processing technologies that increase human localization accuracy [8][9].

The biometrics-based attendance monitoring systems often use face, fingerprint, and other biometric technology to identify attendees. An attendance monitoring system based on fingerprint identification was created by Muchtar et al. [10]. Each individual may be recognized on various fingerprint sensors thanks to the usage of Microcontrollers to manage biometerics, which boosts the effectiveness of managing attendance.

A face recognition attendance checking project called FaceTime was developed by Arsenovic et al. [11]. Students should first enter their identity from their ID cards before FaceTime may gather and recognize their faces via the webcam and record their attendance in the classroom, but the system is limited to marking attendance inside the classroom only.

In addition to developing a comparable mobile device application, Yang et al. presented an smart attendance verification project based on voice recognition and live geographical locating [12]. The application activates the mobile’s microphone during attendance checking, while students have to finish the process by reading a passage of text. About 120 students from a first-year computer science course participated in the testing of this program. The length of the attendance check can be capped at five minutes if the application satisfies the necessary accuracy requirements.

Using the LBPH algorithm for face recognition and the haar cascade for face detection, Bharath Tej Chinimilli et al. [13] trained a model. To train the model, the author of the paper has used their personal dataset. This system offers features like photographing students and recording their information for the database, training the photos in the database and on the camera, and beginning to track persons entering the classroom. This system recognizes the faces of students walking into the classroom from the webcam and pre-processes them for further processing.

The following are some downsides of the aforementioned types of attendance verification systems.

1. Cost of the Identity-based attendance such as Fingerprint, Iris, etc is high, and it is difficult to verify students' identities.
2. The attendance project based on location can only provide an estimate of the number of mobile devices present in the attendance checking area, and cannot accurately determine the exact count of participating students.

III. ALGORITHM IMPLEMENTATION

A. *Multitask Convolutional Neural Network*

Using Multitask Convolutional Neural Network (MTCNN), we can recognize faces and facial landmarks like the lips, eyes, and nose. [14]. To create face candidates in the first stage, a proposal network (P-Net), which is a completely complex network. Non-maximal suppression merges similar candidate boxes (NMS). Another CNN dubbed Refine Network (R-Net), which is employed in the second stage, is used to weed out a lot of fraudulent candidates. In the final phase, the facial recognition results are improved, and the network outputs the locations of her five facial recognition points. According to the actual needs of the attendance system, modify some parameters and thresholds to improve the algorithm.

Fig. 1. Face Detection by MTCNN

B. FaceNet

This network was developed by Google in the year 2015 mostly for the purpose of facial detection. Facenet transforms a facial image into 128-dimension feature vector. We can represent distance between two vector using Euclidean distance. Complexity of the process is reduced significantly after transforming images into 128-Dimensional vectors and then calculating the distance between those vectors. This FaceNet algorithm uses a function named triplet loss in order to train the respective neural network. A favorable pattern is used to form the triplet. The triplets which are submitted to the network for training purpose are produced by the dataset. While training a particular model a triple loss function is taken into consideration. Less the distance between embedding of the same class better is the outcome. Following is the triplet loss function

$$L = \sum_{i}^{N} [||f(x_i^a) - f(x_i^p)||_2^2 - ||f(x_i^a) - f(x_i^n)||_2^2 + a]_+ \quad (1)$$

C. SVM

For both classification and regression, SVM is a supervised machine learning technique. Support vectors, or points on the edges of each class, are used as the foundational building blocks of SVM to identify the ideal hyper plane for classification. Support vector regression is the name given to the process of using SVM for regression (SVR).SVMs can be either linear or nonlinear in nature. There are two more types of linear SVMs: hard margins and soft margins. Hyper planes can properly linearly separate the data in hard-margin SVMs, but cannot do so in soft-margin SVMs. Positive slack is introduced via soft-margin SVM. This is a fine for having data points outside of the allowed margin. As the distance from the edge grows, the Slack variable's value rises. Use of a nonlinear SVM is made when the data are nonlinear. Using kernel functions from the training set, the original input space is in this case transferred to a high-dimensional feature space. Radial basis (rbf), polynomial (poly), and sigmoid kernels are the most oftenly used. SVM is more resistant to noise and outliers.

IV. SYSTEM ARCHITECTURE AND IMPLEMENTATION

A. System Architecture Design

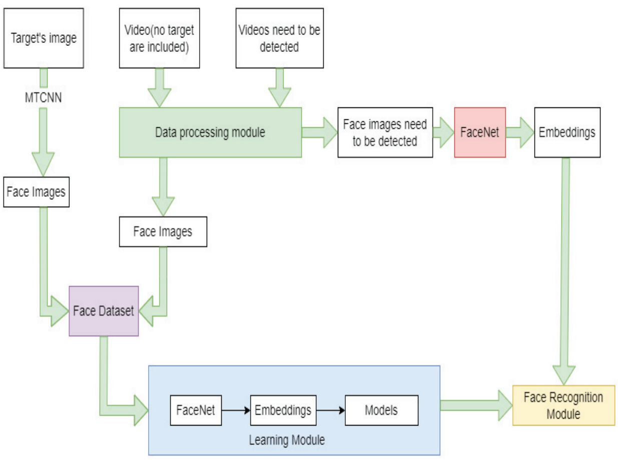

Fig. 2. Architecture Diagram

A system's structure, behavior, and other aspects are all defined by its conceptual model, or system architecture. This is a formal description and representation of the system that is set up to facilitate analysis of face recognition module.

Our project mainly consists of three modules namely, data preprocessing module, the learning module, and the face recognition module. Creating a record involves the following steps:

1. Utilize the crawler tool to sequentially explore each target-related image.
2. Use MTCNN to crop the headshot of the image, remove the extra details, and exclude images of unknown persons.
3. Click the Create button after entering your username in the editable text area.
4. The system uses cameras that are attached to it to take pictures. For further processing, these collected photos are stored in data storage. To increase the precision of face identification, normalize picture rotation based on orientation information.

The face recognition network model in the document can be generated by utilizing the data set obtained from the above-mentioned method as input for the training module, the dynamic part that needs to be identified is processed by the data preparation module, appearing in the video.

Extract the needed face picture. To identify targets in the video, the identification module inputs a previously obtained network model as well as face photos. Hence this a complete low-level overview of the system architecture design.

B. Data Preprocessing Module

The use of the data preprocessing module is to extract the facial images in the video and provide the data to the next facial recognition module. However, there is redundancy in the face data in the video, and the content of adjacent video frames is very similar. To eliminate redundant data, the data preprocessing engine extracts facial images into key frames for facial recognition.

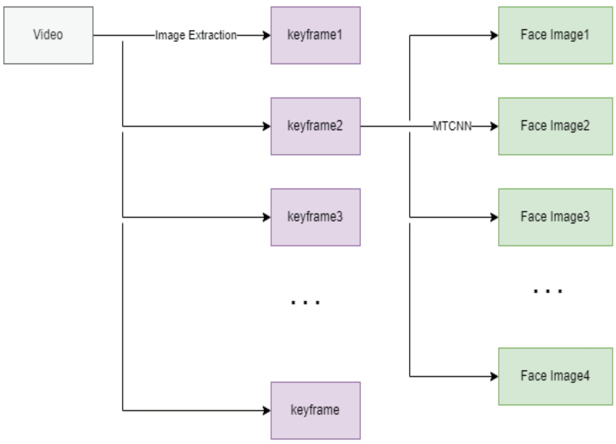

Fig. 3. Data Preprocessing Module

The key frames of the target video and detect the facial images in the key frames by his MTCNN [14]. MTCNN is implemented by a three-layer network structure (P-Net, R-Net, ONet).

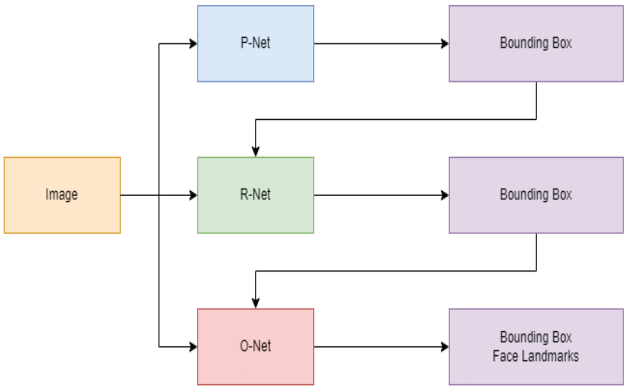

Fig. 4. MTCNN three-level network structure

C. Learning Module

The learning module's ultimate objective is to build a face recognition network model that can be used to recognize specific characters in videos. In this module, this work uses the FaceNet [15] algorithm. Each picture in the dataset used in this study by FaceNet receives an embedding [16] from the system (an embedding is f: indicates a mapping where X -> Y, and the function is injective and structure-preserving). FaceNet is not directly utilized by the learning module to accomplish facial recognition. Reason: The dataset doesn't include enough data. You won't obtain a satisfactory model if you directly train a face recognition neural network using FaceNet.

Data less classification problems are a good fit for the SVM classifier. This is a novel approach to learning that uses brief examples and a sound justification. Because it does not essential rely on probability measurements or the rule of big numbers, it departs from conventional statistical techniques.

It automatically eliminates the conventional derivation-to-estimation method, enabling effective "transfer inference" from training samples to prediction samples, and significantly streamlines typical classification and regression issues. Only a few support vectors are used to determine the SVM's final decision function. The so called "curse of dimensionality" is substantially avoided since computational cost mostly depends on the quantity of support vectors, rather than the dimensionality of the sample space. Only a few support vectors influence the outcome. This not only aids in collecting the most crucial samples but also eliminates a lot of unnecessary samples. SVMs are not just straightforward, but also quite "robust." As a result, the embedding using FaceNet is generated and classify them using SVM.

D. Face Recognition Module

The recognizer in this paper recognizes faces using an SVM classifier. Video-based face recognition systems must handle multiclass classification issues because traditional support vector machine methods only offer binary classification techniques. By creating many binary SVM classifiers, you may achieve multiclass classification. That is, a binary SVM classifier exists for each target. In order to optimize the distance between positive and negative samples in the training set, SVM searches for hyper planes that are ideally separated from one another in the feature space. For binary classification problems, the SVM supervised learning technique is utilized. SVM may be used to handle nonlinear problems once kernel techniques are incorporated. Three different SVM types exist:

1. Support vector machine that is linearly separable: By maximizing the hard interval, a support vector machine that is linearly separable may be produced if the training data is linearly separable.
2. Soft Interval Support Vector Machine [17]: By maximizing the soft interval when the training data are almost linearly separable, a linear support vector machine can also be created.
3. Nonlinear Support Vector Machine: The nonlinear support vector machine may be created using the kernel approach and soft interval maximization if the training data cannot be linearly separated.

Nonlinear support vector machines are used in this article's identification module. Put the FaceNet embedding he created from the photos the data preprocessing module produced into his SVM model, one for each target. Depending on when and where the target occurs in the data preprocessing module's video output, an SVM binary classifier is passed for each target, and the bounding boxes of the recognized name and person picture are produced along with the matching top-k results. An enclosed box is drawn in the video picture.

V. EXPERIMENTAL RESULTS AND ANALYSIS

As in the below images we can capture multiple face at once from the real-time video which can be further stored into database to mark the attendance. The experiment uses FaceNet and MTCNN to recognize the face and process the face for identification. The experiment produces an accuracy of 94.85% using combination of both the algorithm. Test accuracy of 90.26% was achieved using the given algorithms on a dataset of 476 images out of which 468 were detected accurately. Attendance can be captured using any camera and can be stored in database for further processing.

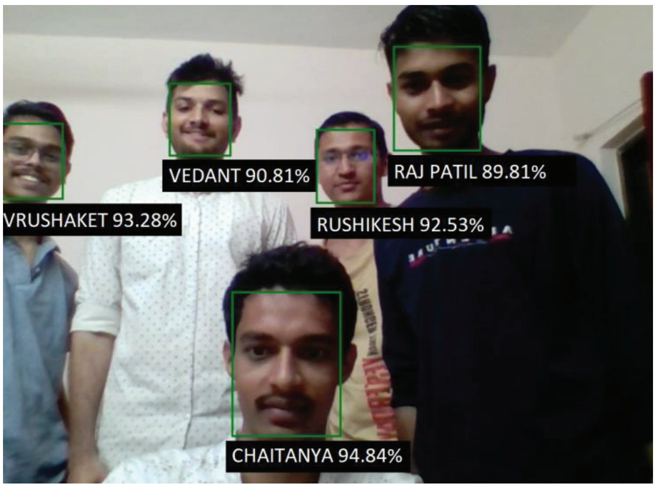

Fig. 5. Image Extracted from video being captured through webcam

Fig. 6. Faces recognized from video being captured through webcam

A. *Training and Testing MTCNN with FaceNet and SVM*

TABLE I. TRAINING AND TESTING DETAILS

Training start time	1668160017
Local Current	Tue Dec 10 2022 11:12:51 GMT+0530 (India Standard Time)
Test images	476
Images identified:	468
Accuracy (Train)	94.85
Training end time	1667160426
Local Current	Tue Dec 10, 2022, 11:22:16 GMT+0530 (India Standard Time)
Training executing time	409
Test Start Time	1668160456
Local Current	Tue Dec 10, 2022, 11:22:16 GMT+0530 (India Standard Time)
Accuracy (Test)	90.26
Test End Time	1668161128

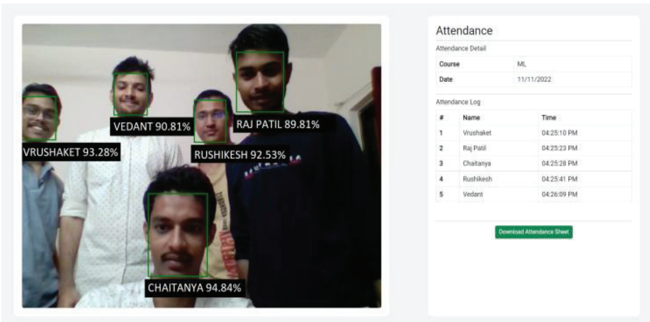

Fig. 7. Attendance System

VI. CONCLUSION

The video face recognition system in this work has a positive impact on attendance systems and many other practical applications. The following improvements have been made to more accurately and quickly identify target characters in videos: By introducing image quality evaluation algorithms, system can filter out obscure video frames with little useful information, eliminating interference while speeding up system operation. A face tracking algorithm is introduced to take full advantage of video timing information and use frame-to-frame differences to determine facial images of the same person for better detection accuracy. In this paper, a dynamic attendance system with face recognition based on MTCNN, FaceNet and SVM has been design. After optimizing the MTCNN face recognition algorithm, the recognition rate reaches 38fps. In this document, the FaceNet neural network to transform face images into 128-dimensional vectors is used. To achieve the face detection function, Euclidean distances between detected faces and faces in the local database are calculated. Based on his three networks above, a presence system with user interaction interface was developed using the Flask Framework. Attendance systems can be used in schools, companies, businesses, and other scenarios where multiple people need to be present at the same time.

REFERENCES

[1] Jose, Edwin & Manikandan, Greeshma & T P, Mithun Haridas & M H, Supriya. (2019). Face Recognition based Surveillance System Using FaceNet and MTCNN on Jetson TX2. 10.1109/ICACCS.2019.8728466.

[2] H. D. Rjeib, N. S. Ali, A. Ai Farawn, B. Al-Sadawi, and H. Alsharqi, ‘‘Attendance and information system using RFID and Web-based application for the academic sector,’’ Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 1, pp. 266–274, Jan. 2018

[3] B. I. Ahmad, ‘‘TouchIn: An NFC supported attendance system in a university environment,’’ Int. J. Inf. Educ. Technol., vol. 4, no. 5, p. 448, Jan. 2014.

[4] J. Jacob, K. Jha, P. Kotak, and S. Puthran, ‘‘Mobile attendance using near field communication and one-time password,’’ in Proc. Int. Conf. Green Comput. Internet Things (ICGCIoT), Noida, India, Oct. 2015, pp. 1298–1303.

[5] S. M. Cisar, R. Pinter, V. Vojnic, V. Tumbas, and P. Cisar, ‘‘Android application for tracking students' class attendance,’’ in Proc. IEEE 14th Int. Symp. Intell. Syst. Informat. (SISY), Subotica, Serbia, Aug. 2016, pp. 227–232.

[6] A. Abdulkareem, ‘‘The design, development, and evaluation of a smart attendance system using BLE beacons,’’ M.S. thesis, Dept. Comput. Eng., Kankaya Univ., Etimesgut, Turkey, 2018.

[7] S. Anand, K. Bijlani, S. Suresh, and P. Praphul, ‘‘Attendance monitoring in the class using mobiles & Wi-Fi fingerprinting,’’ in Proc. IEEE 8th Int. Conf. Technol. Educ. (T4E), Mumbai, India, Dec. 2016, pp. 62–67.

[8] Z. Shi, C. Zhou, Y. Gu, and S. He,‘‘A fast and efficient algorithm for coprime array adaptive beamforming,’’ IEEE Trans. Veh. Technol., vol. 67, no. 2, pp. 1099–1112, Feb. 2018.

[9] X. Zhou, W. Liang, K. Wang, H. Wang, L. T. Yang, and Q. Jin, ‘‘Deep learning enhanced human activity recognition for Internet of healthcare things,’’ IEEE Internet Things J., vol. 7, no. 7, pp. 6429–6438, Jul. 2020.

[10] M. A. Muchtar, Seniman, D. Arisandi, and S. Hasanah, ‘‘Attendance fingerprint identification system using Arduino and single board computer,’’ J. Phys., Conf. Ser., vol. 978, Mar. 2018, Art. no. 012060.

[11] M. Arsenovic, S. Sladojevic, A. Anderla, and D. Stefanovic, ‘‘FaceTime— Deep learning based face recognition attendance system,’’ in Proc. IEEE 15th Int. Symp. Intell. Syst. Informat. (SISY), Subotica, Serbia, Sep. 2017, pp. 53–58.

[12] S. Yang, Y. Song, H. Ren, and X. Huang, ‘‘An automatic student attendance tracking system based on voiceprint and location,’’ in Proc. 11th Int. Conf. Comput. Sci. Educ. (ICCSE), Nagoya, Japan, Aug. 2016, pp. 214–219.

[13] B. Tej Chinimilli, A. T., A. Kotturi, V. Reddy Kaipu, and J. Varma Mandapati, "Face Recognition based Attendance System using Haar Cascade and Local Binary Pattern Histogram Algorithm," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 2020, pp. 701-704, doi: 10.1109/ICOEI48184.2020.9143046.

[14] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," in IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.

[15] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682..

[16] LUFT JH. Improvements in epoxy resin embedding methods. J Biophys Biochem Cytol. 1961 Feb;9(2):409-14. doi: 10.1083/jcb.9.2.409. PMID: 13764136; PMCID: PMC2224998.

[17] Shantanu Jain, Vrushaket Chaudhari, Rushikesh Chuadhari, Tanvesh Chavan, Prof. Priyanka Shahane, "A Survey on Face Recognition Techniques in Machine Learning ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8, Issue 6, pp.50-66, November-December-2022. Available at doi : https://doi.org/10.32628/CSEIT228558