

Face Recognition for Attendance System using Multi-Task Cascaded Convolutional Networks

Annjoe J Prasad

Department Of Computer Science and Technology
St. Xavier's Catholic College of
Engineering, Nagercoil, India

Mr. M Ajin

Department Of Computer Science and Technology
Assistant Professor, St. Xavier's Catholic College
of Engineering, Nagercoil, India

Abstract - The widespread adoption of biometric technologies has opened new avenues for automating traditionally manual processes, with facial recognition standing out as a reliable and non-intrusive approach. This work presents an intelligent attendance management framework that employs Multi-Task Cascaded Convolutional Networks (MTCNN) as the foundational algorithm for face detection and alignment. MTCNN's three-stage cascade—comprising the Proposal Network (P-Net), Refinement Network (R-Net), and Output Network (O-Net)—enables progressive, high-precision localization of facial regions. Recognized individuals are subsequently verified through a Convolutional Neural Network (CNN) trained with transfer learning on domain-specific data, substantially improving identification accuracy. The system captures attendance entries in real time and stores timestamped records in a structured digital format, accessible through an intuitive administrative interface. This approach eliminates the inefficiencies associated with conventional roll-call methods while providing a scalable solution for institutional settings.

Keywords - MTCNN; P-Net; R-Net; O-Net; face recognition; attendance automation; deep learning; transfer learning; CNN.

I. INTRODUCTION

Conventional attendance systems in academic and corporate environments continue to rely on manual procedures that are both time-consuming and susceptible to inaccuracies. As institutions grow in scale, the need for an automated, reliable, and non-intrusive solution becomes increasingly pressing. This work addresses that gap by introducing a face recognition-based attendance system built on the Multi-Task Cascaded Convolutional Networks (MTCNN) architecture [1].

MTCNN has established itself as a preferred solution for facial preprocessing due to its ability to detect faces across varied orientations, scales, and lighting conditions with high speed and accuracy [15]. Unlike conventional image-processing pipelines that treat face detection and landmark localization as separate tasks, MTCNN performs both simultaneously through a cascaded multi-task learning approach, making it highly efficient for real-time deployment.

The proposed attendance framework integrates MTCNN for face detection with a CNN-based recognition module trained through transfer learning. Upon detection, the

system extracts discriminative facial embeddings and compares them against a pre-registered database to identify individuals. Once a match is confirmed, the attendance record is automatically updated with the corresponding timestamp.

Beyond the immediate application domain, this framework demonstrates the broader utility of facial recognition technology in settings such as access control, event monitoring, and organizational security—highlighting the transformative potential of deep learning-driven biometric systems [1].

II. LITERATURE REVIEW

A substantial body of research has been devoted to improving the accuracy and robustness of face detection and recognition systems. Early works predominantly focused on static image datasets; however, recent advances have shifted attention toward real-time, unconstrained scenarios.

A. Aggregate Channel Features for Face Detection

Yang et al. proposed Aggregate Channel Features (ACF) for detecting faces captured from multiple viewpoints. Their method demonstrated improved multi-view detection accuracy through robust feature representation. While effective in controlled settings, the approach incurs considerable computational overhead and exhibits sensitivity to hyperparameter selection, limiting its scalability in dynamic environments.

B. Deep CNNs for Multi-View Face Detection

Farfadi, Saberian, and Li explored the use of deep convolutional neural networks for multi-view face detection (ACM ICMR, 2015). The study highlighted the representational strength of deep CNNs in handling facial variations across angles and orientations. However, the method's reliance on large annotated training datasets raises practical concerns regarding deployment in resource-constrained systems.

C. Polygonal Haar-Like Features

Pham et al. introduced an extended Haar-like feature set using fast polygonal integration to enhance object localization accuracy. This technique offered improved discriminative power and reduced false-positive rates compared to standard

Haar features. Nevertheless, the added complexity and sensitivity to training data quality remain key limitations.

D. Face Detection Without Bells and Whistles

Mathias et al. (ECCV 2014) proposed a straightforward face detection framework that deliberately avoided auxiliary enhancements, achieving competitive accuracy through a clean and generalizable design. While praised for its simplicity, the evaluation did not fully account for extreme occlusion or low-resolution scenarios.

E. Deep Learning Face Attributes

Liu et al. investigated deep learning-based prediction of facial attributes in unconstrained, real-world conditions. Their CNN-based framework classified attributes such as gender, age, and facial hair presence. Despite its practical relevance, the work received limited scrutiny regarding ethical implications and computational efficiency at scale.

III. APPROACH

A. Input Data

The system accepts images or live video frames as primary inputs. Each frame may contain one or more faces to be detected and identified. Prior to feeding frames into the detection pipeline, students are registered by associating their names and enrollment numbers with facial images stored in the system database [1][24]. The quality, diversity, and lighting conditions of input data are critical determinants of overall system performance [21].

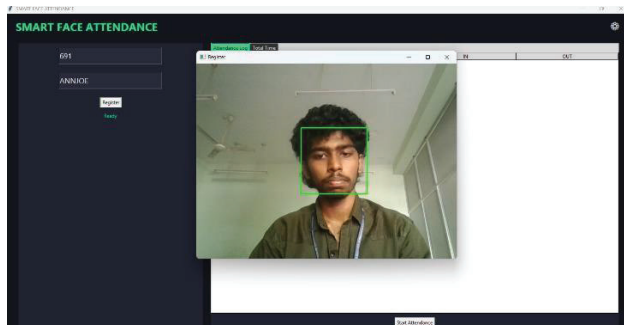


Fig 1. Input Data

B. Face Detection via MTCNN

MTCNN serves as the detection backbone, processing input frames through its three-stage cascade: P-Net, R-Net, and

O-Net. Each stage progressively refines candidate face regions, improving both localization precision and recall [3][22][23]. This cascade design enables the system to efficiently handle faces of varying sizes while suppressing false-positive detections.

C. Bounding Box and Landmark Estimation

For each detected face, the system generates bounding boxes to define spatial extent within the frame. Simultaneously, five facial landmark coordinates—eye centers, nose tip, and mouth corners—are predicted to guide subsequent alignment and recognition steps, ensuring consistent feature extraction regardless of pose variation.

D. Face Recognition

Following detection and alignment, a pre-trained face recognition model (e.g., VGGFace or FaceNet) processes the extracted facial regions to generate embedding vectors [3][4][5]. These embeddings are matched against stored reference embeddings using a distance-based similarity metric. A match exceeding a defined confidence threshold results in successful individual identification [21].

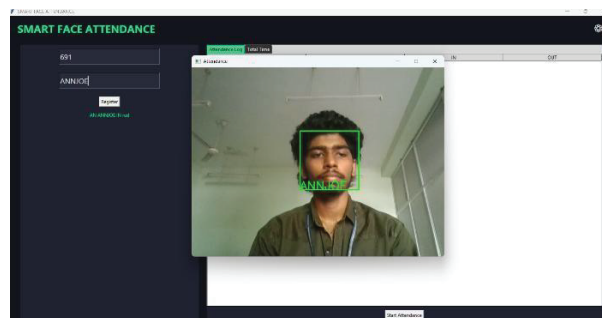


Fig 2. Face Recognition

E. Attendance Recording

Upon successful identification, the system logs the individual's attendance by updating the record with a timestamp and associated metadata [11][16]. This process occurs in real time, ensuring accurate and immediate documentation without manual intervention.

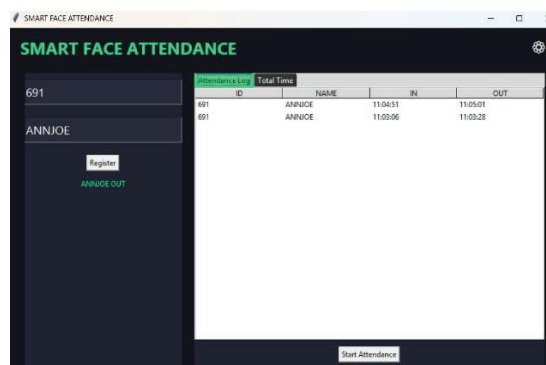


Fig 3. Attendance Marking

F. Database Management

A centralized database stores facial embeddings, student identifiers, and historical attendance records. The database is updated continuously as new registrations are added and new sessions are conducted [6][18].

G. User Interface

An administrative interface enables operators to input images or connect a live camera feed for detection and recognition. The interface also supports viewing attendance summaries, exporting records, and managing registered profiles.

H. Ethical Considerations

The system complies with applicable privacy regulations. Consent is obtained from all registered individuals prior to facial data collection. Measures are implemented to mitigate potential biases in recognition accuracy across demographic groups [5].

I. Testing and Evaluation

System performance is assessed across multiple datasets and environmental conditions, measuring recognition accuracy, processing throughput, and real-time responsiveness [7]. Insights from evaluation are incorporated into iterative refinements.

IV. SYSTEM ARCHITECTURE

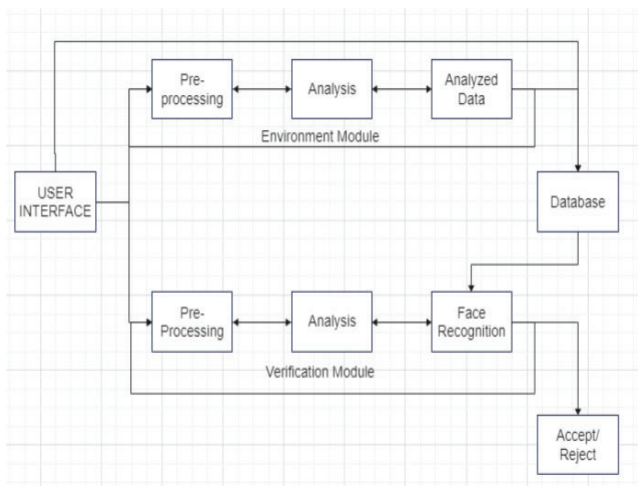


Fig 4. Architecture Diagram

A. User Interface Layer

The front-end interface provides administrators with tools for initiating attendance sessions, reviewing historical records, and managing the registered student database.

B. Data Preprocessing

Raw images are cropped and normalized to ensure uniformity in size, brightness, and orientation. Each image is annotated with a corresponding student identifier prior to model training.

C. Recognition Module

A deep CNN trained using transfer learning on preprocessed institutional data performs feature extraction and identity matching. Architectures such as VGGFace, FaceNet, or OpenFace are suitable depending on performance requirements.

D. Database Layer

The database maintains a mapping between individuals, their facial embeddings, and corresponding attendance histories, enabling efficient lookup and record updates.

E. Analytics Module

An analytics component evaluates system performance over time by tracking recognition confidence scores, processing latency, and attendance patterns. Security and data integrity protocols are enforced at this layer.

F. Accept/Reject Logic

Incoming facial embeddings are validated against the database. If a match is found above the similarity threshold, the individual is recognized and attendance is marked. Faces with no matching record are flagged as unknown.

G. Modules

The system comprises two primary modules. The environment module handles hardware configuration and camera interfacing for image capture [24][2]. The verification module compares detected face embeddings against registered profiles, confirming or denying identity to ensure attendance integrity.

V. METHODOLOGY

MTCNN is a multi-task deep learning framework designed to perform face detection and facial landmark localization jointly within a single unified pipeline [8][17]. Its architecture reflects a deliberate cascading strategy that balances detection thoroughness with computational efficiency.

The cascade begins with the Proposal Network (P-Net), a fully convolutional network that rapidly scans an image pyramid to generate initial candidate face bounding boxes [9]. These candidates are passed to the Refinement Network (R-Net), which discards low-quality proposals and refines surviving candidates, resulting in a significant reduction in false positives [9].

The final stage, the Output Network (O-Net), performs high-precision bounding box regression and predicts five facial landmark coordinates. This stage produces the definitive detection results with the highest level of spatial accuracy.

MTCNN's efficiency derives from the cascaded structure, which focuses computation on genuine face candidates. Its robustness encompasses diverse scenarios including occlusions, variable illumination, and a wide range of face sizes and head orientations [14]. The landmark output enables precise facial alignment prior to recognition, contributing to more consistent embedding generation.

Compared to legacy detection approaches such as R-CNN and Faster R-CNN [12][13], MTCNN achieves competitive accuracy with significantly lower inference time, making it well-suited for real-time applications.

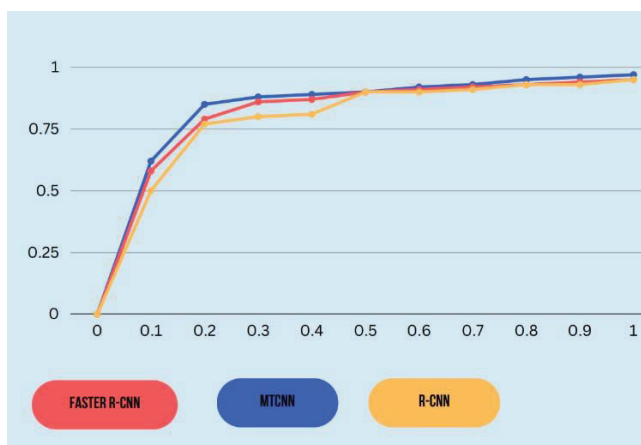


Fig 5. ROC curves of MTCNN, R-CNN and Faster R-CNN

The Accuracy of MTCNN Compared with the previous CNN algorithms are shown in Fig 6



Fig 6. Accuracy Comparisons

VI. RESULT

Experimental evaluation of the proposed framework demonstrated reliable attendance tracking with high recognition accuracy. The system exhibited low rates of false positives and

false negatives under standard indoor lighting conditions, and processed frames at near-real-time speeds.

Attendance records were stored in structured Excel files, organized by date and annotated with session timestamps. Each recognized student's data was archived in a dedicated folder linked to their registration profile. The system's user-friendly interface received positive feedback from administrative personnel during testing, and its digital record-keeping capability substantially reduced workload compared to manual methods.

Performance degradation was observed under extreme lighting variation and heavy occlusion, indicating areas for further development. Ongoing refinement of the recognition threshold and model retraining with expanded data are anticipated to improve robustness in challenging conditions.

	A	B	C	D	E	F
1	ID	NAME	DATE	IN	OUT	TOTAL
2	691	ANNJOE	30-03-26	11:03:06	11:03:28	0:00:22
3	691	ANNJOE	30-03-26	11:04:51	11:05:01	0:00:10
4						
5						
6						
7						
8						
9						

Fig 6. Attendance stored in Excel

The Data get stored in the folder with the name of the student and when the attendance was taken it get compared with the image stored in the file

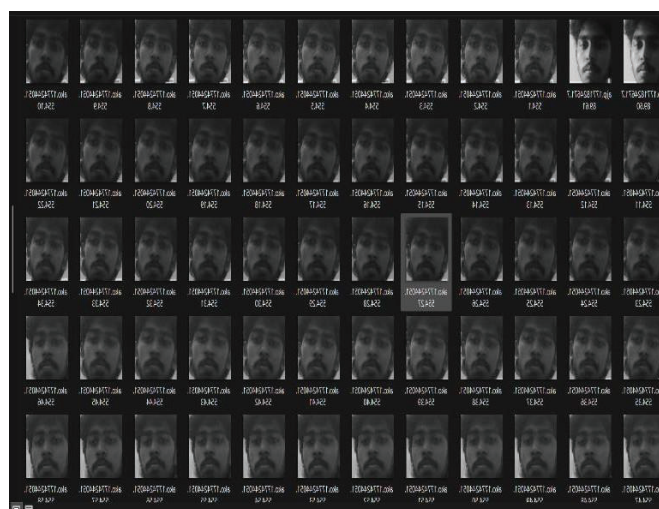


Fig 7. Data (After feature Extracted)

VII. CONCLUSION

This paper introduced a cascaded CNN-based framework for simultaneous face detection and landmark alignment, applied to the task of automated attendance management. Experimental results across standard benchmarks—including FDDB and WIDER FACE for detection, and AFLW for alignment—demonstrated competitive performance relative to existing state-of-the-art methods with real-time processing capability.

The multi-task learning paradigm adopted by MTCNN, which jointly optimizes detection and alignment objectives, proves particularly effective in reducing accumulated error compared to sequential, independent pipelines. This integrated approach strengthens both detection reliability and alignment precision simultaneously.

Looking ahead, the framework can be adapted for edge deployment through model compression and quantization. Further improvements in recognition performance may be achieved by incorporating transformer-based architectures. The system also holds promise for extension into access control, surveillance, and biometric authentication in organizational contexts.

REFERENCES

- [1] N. Zhang, J. Luo, and W. Gao, "Research on face detection technology based on MTCNN," in Proc. 2020 Int. Conf. Computer Network, Electronic and Automation (ICCNEA), 2020, pp. 154–158.
- [2] M. Gu, X. Liu, and J. Feng, "Classroom face detection algorithm based on improved MTCNN," *Signal, Image Video Process.*, vol. 16, no. 5, pp. 1355–1362, 2022.
- [3] E. Kaziakhmedov et al., "Real-world attack on MTCNN face detection system," in Proc. 2019 Int. Multi-Conf. Engineering, Computer and Information Sciences (SIBIRCON), 2019, pp. 0422–0427.
- [4] C. Wu and Y. Zhang, "MTCNN and FACENET based access control system for face detection and recognition," *Autom. Control Comput. Sci.*, vol. 55, pp. 102–112, 2021.
- [5] A. Ghofrani, R. M. Toroghi, and S. Ghanbari, "Realtime face-detection and emotion recognition using MTCNN and MiniShuffleNet V2," in Proc. 2019 5th Conf. Knowledge-Based Engineering and Innovation (KBEI), 2019, pp. 817–821.
- [6] F. Zumstein, *Python for Excel*. Sebastopol, CA: O'Reilly Media, 2021.
- [7] H. Ku and W. Dong, "Face recognition based on MTCNN and convolutional neural network," *Frontiers Signal Process.*, vol. 4, no. 1, pp. 37–42, 2020.
- [8] X. Chen et al., "Eyes localization algorithm based on prior MTCNN face detection," in Proc. 2019 IEEE 8th Joint Int. ITAIC Conf., 2019, pp. 1763–1767.
- [9] Y. Fu, M. Kim, and J. W. Jang, "Research and optimization of face detection algorithm based on MTCNN in a complex environment," *J. Korea Inst. Inf. Commun. Eng.*, vol. 24, no. 1, pp. 50–56, 2020.
- [10] M. Ma and J. Wang, "Multi-view face detection and landmark localization based on MTCNN," in Proc. 2018 Chinese Automation Congress (CAC), 2018, pp. 4200–4205.
- [11] N. Darapaneni et al., "Automatic face detection and recognition for attendance maintenance," in Proc. 2020 IEEE 15th Int. Conf. Industrial and Information Systems (ICIIS), 2020, pp. 236–241.
- [12] M. Rezaei et al., "Assessing the effect of image quality on SSD and Faster R-CNN networks for face detection," in Proc. 2019 27th Iranian Conf. Electrical Engineering (ICEE), 2019, pp. 1589–1594.
- [13] M. Rezaei et al., "Assessing the effect of image quality on SSD and Faster R-CNN networks for face detection," in Proc. 2019 27th Iranian Conf. Electrical Engineering (ICEE), 2019, pp. 1589–1594.
- [14] R. Karmakar, "Facial attendance system using MTCNN and feature mapping," *Int. J. Eng. Applied Sci. Technol.*, vol. 5, pp. 546–550, 2020.
- [15] S. B. Bhaskoro, S. Aminah, and K. Taqi, "Attendance system on moving objects through face recognition using MTCNN and CNN," in Proc. 2021 3rd Int. Symp. Material and Electrical Engineering (ISMEE), 2021, pp. 184–189.
- [16] N. C. Basjaruddin et al., "Attendance system with face recognition, body temperature, and mask detection using MTCNN," *Green Intell. Syst. Appl.*, vol. 2, no. 2, pp. 71–83, 2022.
- [17] M. Azamy, A. B. Ariwibowo, and I. Mardianto, "Face recognition implementation with MTCNN on attendance system prototype at Trisakti University," *Indonesian J. Banking Financ. Technol.*, vol. 1, no. 1, pp. 73–88, 2023.
- [18] M. Varsha and S. Chitra Nair, "Automatic attendance management system using face detection and recognition," in *IoT and Analytics for Sensor Networks: Proc. ICWSNUCA 2021*, Singapore: Springer, 2022, pp. 97–106.
- [19] E. Ramalakshmi, S. Doddapaneni, and S. Gajawada, "Facial recognition attendance system using MTCNN and FACENET," *Grenze Int. J. Eng. Technol. (GIJET)*, vol. 8, no. 1, 2022.
- [20] S. Huang and H. Luo, "Attendance system based on dynamic face recognition," in Proc. 2020 Int. Conf. Communications, Information System and Computer Engineering (CISCE), 2020, pp. 368–371.