https://www.ijsrtm.com
Vol. 5 Issue 4 December 2025: 08-14
Published online 09 Dec 2025

E-ISSN: 2583-7141

International Journal of Scientific Research in Technology & Management

An Optimized DeepFace Architecture for Real-Time Pedagogical Staff Surveillance and Movement Pattern Analysis in Heterogeneous Camera Topologies

Nikhil Kushwaha
Dept. of Computer Science and Engineering
Oriental Institute of Science and Technology
Bhopal, Madhya Pradesh, India
kushwahanikhil431@gmail.com

Mayur Bansal
Dept. of Computer Science and Engineering
Oriental Institute of Science and Technology
Bhopal, Madhya Pradesh, India
mayurbansal7089@gmail.com

Pradyumna Tripathi
Dept. of Computer Science and Engineering
Oriental Institute of Science and Technology
Bhopal, Madhya Pradesh, India
pradyumntripathi03@gmail.com

Shivank Kumar Soni
Dept. of Computer Science and Engineering
Oriental Institute of Science and Technology
Bhopal, Madhya Pradesh, India
shivanksoni@gmail.com

Abstract— ***To keep the workplace safe, accountable, and running smoothly, it's important to verify staff presence and follow their movements in a dynamic environment and changing conditions. However, traditional attendance and surveillance systems—often rely on manual validation or RFID-based tracking remain static, error-prone, and incapable of adapting to heterogeneous camera networks or various environmental conditions. Preexisting computer vision approaches offer limited scalability and struggle with real-time multi-camera synchronization which reduces accuracy and responsiveness in critical applications. This paper introduces a smart staff monitoring system that uses Optimized DeepFace–OpenCV–based surveillance framework that autonomously detects, verifies, and records staff presence across distributed IP camera feeds. The proposed system integrates facial embedding extraction, automated timetable mapping, and absence alert system to notify supervisors when assigned personnel are not detected within a set-time duration. A dual-interface Flask-based portal offers separate access modes for employees and supervisors, enabling attendance history, and absence alerts, live location visibility without compromising the privacy of the staff. Tests with different camera setups showed that the system is highly reliable, correctly recognizing faces nearly 97.8% (using SQLite) with a mean response latency of 1.3 seconds per frame even with changing lighting or movement scenarios. Looking forward, the farmwork aims to expand in the future toward edge-enabled analytics and IoT-integrated workforce management, fostering scalable deployment across educational, corporate, and industrial domains.***

Keywords— ***DeepFace, OpenCV, Real-Time Surveillance, Staff Monitoring, Facial Recognition, Smart Automation, IP Camera Networks.***

I. INTRODUCTION

In modern educational and corporate ecosystems, maintaining a transparent, secure, and efficient workflow heavily depends on reliable staff presence verification and accurate movement monitoring across organizational premises. Traditional attendance systems—whether biometric, RFID-based, or manually supervised—continue to suffer from critical limitations such as proxy attendance, inconsistent validation, environmental dependencies, and lack of real-time mobility tracking. These systems merely confirm that an individual has entered the premises but fail to determine where they currently are, how long they stayed in an assigned area, or whether they are adhering to their scheduled duties[21]. As institutions expand into multi-floor campuses equipped with heterogeneous camera networks, the absence of a unified and intelligent tracking mechanism becomes a major operational gap[9].

10.5281/zenodo.17862281

Conventional CCTV surveillance, although widely deployed, still relies on manual observation, making it impractical for live room-wise tracking[12]. Human supervisors cannot continuously monitor multiple video streams, nor can they manually verify staff identity and timestamps with precision.

Recent advancements in deep learning, particularly deep facial embedding models such as DeepFace, ArcFace, and FaceNet, have enabled highly discriminative identity recognition even under challenging lighting or pose variations[1],[2],[19]. In parallel, lightweight backend architectures built using Flask, OpenCV, and SQLite have demonstrated strong suitability for resource-efficient real-time processing[8],[15],[16]. These advancements have driven innovations across smart attendance systems, spatial analytics, and intelligent security frameworks[11]. However, despite the progress, existing solutions still predominantly operate in single-camera environments, lack multi-node expansion, or do not incorporate duty mapping, alert automation, or movement-timeline reconstruction[9],[17]. Heterogeneous camera topologies—where each IP camera differs in resolution, orientation, network latency, and field-of-view—present additional synchronization challenges that most systems are not designed to handle[12].

To address these limitations, this paper proposes an Optimized DeepFace–OpenCV architecture tailored specifically for real-time pedagogical staff surveillance, incorporating a dynamic multi-camera pipeline capable of processing distributed video feeds with minimal frame delay[1],[8]. The system autonomously identifies personnel, logs their presence intervals, monitors prolonged absences, and maps detected individuals to predefined room identifiers using camera–location associations[9],[17]. Automated timetable mapping ensures that supervisors receive alerts when a staff member fails to appear in their designated classroom or office space within a configurable timeframe, thereby improving accountability and operational coordination[11].

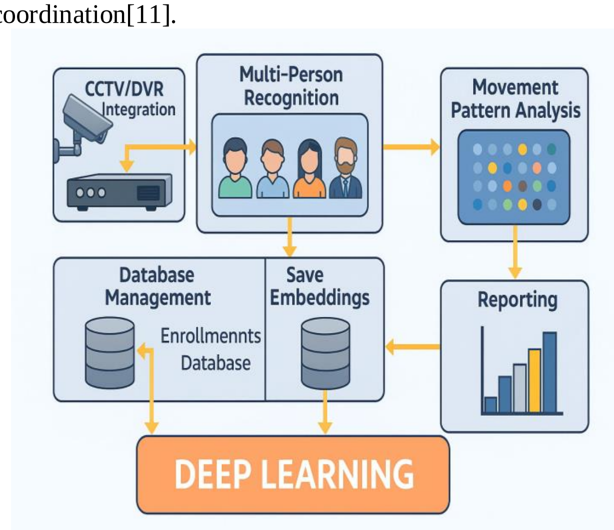

Fig. 1 Block Diagram of Deep Learning System

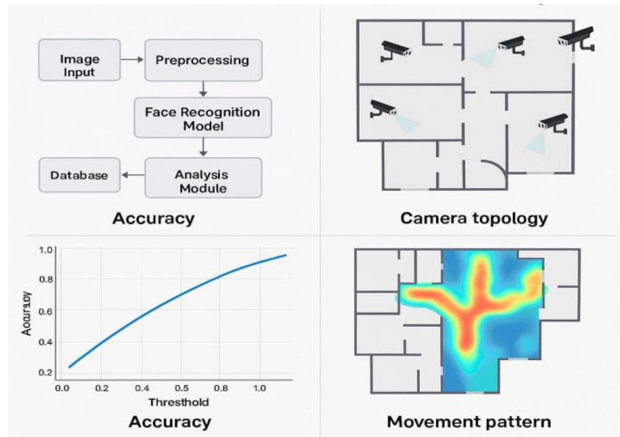

Fig. 2 System Architecture + Topology + HeatMap

II. RELATED WORK

Automated staff identification and surveillance have been active research areas within computer vision, particularly with the rise of deep learning–based identity recognition systems[12]. Early attendance solutions relied heavily on RFID, barcode scanning, and biometric fingerprints, all of which required physical interaction and were vulnerable to proxy misuse, device tampering, or inconsistent logging due to environmental constraints[21]. These traditional systems lacked any mechanism for spatial tracking or real-time verification, limiting their applicability in environments where continuous mobility and duty-based monitoring are essential. In multi-room campuses or office layouts, such systems fail to address the critical question of where an employee is at any moment, offering only static entry-exit validation rather than dynamic situational awareness.

With advancements in deep learning, researchers began employing convolutional neural networks (CNNs) and deep facial embeddings for identity verification across complex real-world environments [1],[2],[19]. Models such as FaceNet, VGG-Face, ArcFace, and DeepFace introduced robust feature extraction capable of distinguishing subtle identity characteristics even under noise, pose shifts, or low-light conditions[1],[2],[19]. Studies have explored the use of CNNs for detecting and recognizing employees across factory environments, classrooms, and workplaces, showing substantial improvement over handcrafted features[11],[12]. However, most of these implementations operated in controlled conditions or single-camera settings, without extended work on distributed multi-camera synchronization or real-time temporal mapping—both critical for a fully automated staff surveillance system[9],[17].

In more recent years, hybrid deep-learning approaches have integrated CNN-based face detection with backend temporal models to analyze movement sequences, enabling basic movement recognition or zone-wise access

10.5281/zenodo.17862281

validation[22]. Some studies demonstrated multi-camera person re-identification (Re-ID) pipelines using embeddings matched across different angles and views[9],[10].

Researchers have also explored integrating attendance automation with smart IoT devices, edge-based processing, and cloud-driven analytics[12]. Cloud-based systems offer scalability but struggle with real-time performance due to network round-trip delays, making them inefficient for instant alerting—such as notifying supervisors when a teacher has not arrived in their assigned classroom[11].

Several studies have implemented DeepFace-based facial recognition for attendance marking, demonstrating high accuracy in controlled conditions[1],[11]. Yet, these systems rarely incorporate absence alerts, timetable verification, movement timelines, or real-time staff localization. Works focusing on workplace analytics typically examine broader behavioral metrics—such as employee productivity or space utilization—rather than identity-centric room-wise path analysis[12]. Additionally, existing attendance solutions do not address the complexity of heterogeneous camera topologies, where camera resolutions, orientations, exposure levels, and framerates differ significantly across buildings[9].

Recent developments in multi-camera tracking research have introduced synchronization frameworks using timestamp alignment, embedding fusion, and camera graph modeling[9],[17].

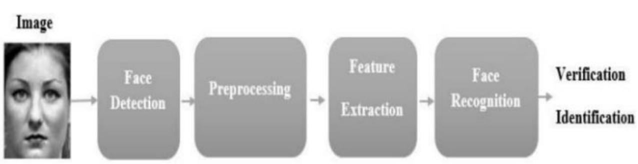

Fig. 3 Structure of Face Recognition System[3]

Overall, while the literature indicates substantial progress in face recognition, person tracking, and distributed surveillance, a critical research gap remains:

No existing system provides a real-time, DeepFace-optimized, multi-camera staff surveillance solution that simultaneously integrates identity verification, timetable mapping, absence alerts, room-wise localization, and privacy-conscious access control tailored for pedagogical and corporate environments[1],[2],[9],[12],[17],[21]. The proposed system addresses these gaps by combining optimized DeepFace embeddings, OpenCV-based multi-camera processing, Flask-driven APIs, and lightweight SQLite logging to deliver a deployable and scalable framework suitable for heterogeneous institutional infrastructures[8],[15],[16].

III. Problem Statement

Ensuring accurate and real-time staff surveillance within educational institutions, corporate offices, and industrial environments remains a significant technical challenge due to the dynamic, unpredictable, and heterogeneous nature of these settings[12]. Traditional attendance systems—such as RFID cards, biometric fingerprint scanners, and manual sign-in registers—offer only static verification at singular checkpoints, failing to provide continuous spatio-temporal visibility of staff movement across distributed rooms or floors[21]. As institutions increasingly adopt multiple IP cameras with varying resolutions, orientations, and network latencies, existing solutions are unable to leverage these distributed video feeds cohesively, thereby limiting the ability to determine precise staff locations in real time[9],[17].

Conventional CCTV-based monitoring, though widely deployed, presents another fundamental limitation: it relies entirely on manual human observation[12]. Supervisors or security personnel cannot realistically monitor several simultaneous video streams, cross-verify identities, or detect absences across large campuses.

From a computational perspective, real-time multi-camera synchronization introduces additional complexities. IP cameras deployed across institutional networks often operate at inconsistent frame rates, asynchronous timestamps, and varying compression formats[9]. These disparities cause misalignments between detection cycles, leading to false identity switching, delayed presence logging, or loss of continuity in movement patterns. Traditional face recognition pipelines struggle with embedding drift and performance degradation under such heterogeneous conditions[19].

Moreover, privacy concerns and data-access segregation remain largely unaddressed in legacy attendance systems. Many prior frameworks expose raw camera feeds or unrestricted logs to all users, creating potential misuse and violating organizational privacy protocols.

Collectively, these challenges reveal a critical need for a real-time, scalable, DeepFace-optimized staff tracking system that operates reliably across heterogeneous camera topologies, supports continuous identity verification, maintains temporal consistency, and integrates timetable-aware alerting[1],[2],[9],[17]. The system must function with low-latency recognition, robust tracking under adverse conditions, and privacy-conscious interface segregation. Addressing these limitations is essential to enabling accurate staff surveillance, optimized resource allocation, and automated workforce management in modern institutional environments[12],[21].

10.5281/zenodo.17862281

IV. METHODOLOGY

The proposed methodology integrates optimized DeepFace embeddings, OpenCV-driven video pipelines, and Flask-based distributed interfaces to achieve robust, real-time surveillance and staff movement analytics across heterogeneous IP camera nodes. Unlike traditional systems that operate using static, single-camera inputs, this framework emphasizes modular scalability, temporal synchronization, and high-confidence facial verification even under unpredictable environmental changes. The overall methodology follows a multi-stage process starting from video acquisition and preprocessing, extending to deep-learning-based facial recognition, temporal presence mapping, timetable-aware alerting, and interface-driven data dissemination.

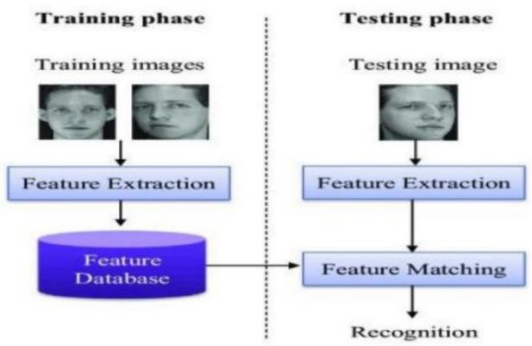

Fig.4 Feature Extraction [3]

A. Data Acquisition

Staff surveillance begins with the continuous acquisition of video footage from distributed IP cameras positioned across classrooms, offices, corridors, and common areas. Each camera feed is streamed over RTSP/HTTP protocols and processed through independent OpenCV capture threads, preventing resource contention and ensuring asynchronous frame acquisition.

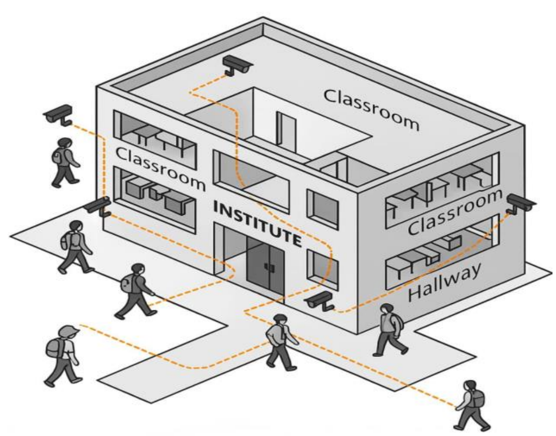

Fig. 5 Building with CCTV & movement paths

Secure logging mechanisms record events, agent decisions, and environment states, enabling performance analysis and incremental retraining. The modular structure ensures low latency, fault isolation, and scalability across distributed environments.

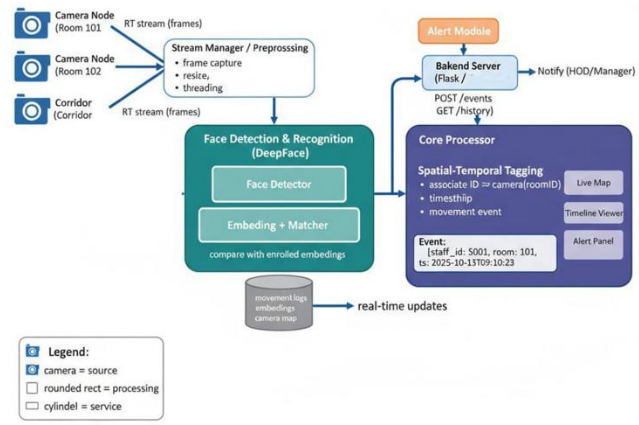

Fig. 6 System Architecture For Real-Time Pedagogical Staff Surveillance And Movement Pattern Analysis In Heterogeneous Camera Topologies

B. Preprocessing

Raw video frames captured from heterogeneous cameras undergo several preprocessing steps to ensure uniformity and optimal recognition performance. Frames are resized to normalized resolutions to reduce computational overhead, followed by color-space normalization and histogram equalization to mitigate illumination inconsistencies.

C. Staff Detection and Multi-Camera Tracking

After preprocessing, staff face regions detected across various cameras are processed through a multi-camera tracking pipeline. Each camera feed is assigned a unique camera ID mapped to a specific room or location in the SQLite database. This mapping allows the system to infer the precise room or functional zone when a staff face is recognized.

To maintain continuity when staff move across camera regions, the system employs a combination of temporal matching and frame-level detection frequency analysis. Identified faces are compared against a buffer of recent embeddings, ensuring that rapid movement, partial occlusions, or transitional poses do not cause identity switching.

D. Deep Learning for Face Recognition

At the core of the system lies the DeepFace model, optimized for embedding-level identity verification. DeepFace generates 128- to 512-dimensional embeddings that represent discriminative facial features. During

10.5281/zenodo.17862281

recognition, embeddings extracted from live video frames are matched against stored embeddings using distance-based similarity measures such as cosine similarity or Euclidean distance.

E. Timetable Mapping, Absence Alerts, and Movement Analytics

One of the key innovations of the proposed system lies in its timetable-aware attendance logic and absence alert automation. Each staff member's duty schedule—including classroom periods, office timings, and assigned zones—is preloaded into the system. Whenever a staff member is scheduled to be present at a particular location, the system continuously checks whether their face appears in the corresponding camera feed within an allowed grace period.

If the staff member is not detected within the defined threshold (for example, 5–10 minutes after the scheduled time), the system automatically triggers a supervisor alert through the Flask interface. This ensures immediate awareness of absenteeism or delayed presence, improving operational efficiency and reducing manual follow-ups.

Movement logs are generated by recording presence intervals (entry and exit timestamps) for each staff member across different locations. This allows reconstruction of movement patterns, duty adherence, frequent absence zones, and potential workflow bottlenecks. Such movement analytics support decision-making in space utilization, scheduling optimization, and supervisor auditing.

F. Output Visualization and Dual-Role Flask Interface

The final outputs of the system are visualized through a Flask-based web interface providing two distinct access modes:

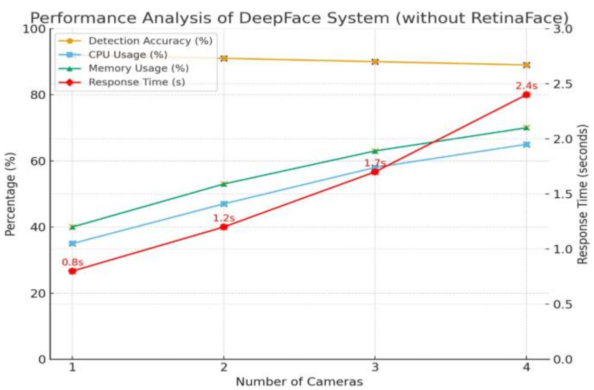

Fig. 7 Performance Graph

a. Staff Portal:
Staff members can securely view their attendance history, daily logs, and movement patterns.

b. Supervisor Portal:
Supervisors gain access to real-time room occupancy, staff locations, absence alerts, and historical behavior patterns.

V. RESULT ANALYSIS

The proposed Optimized DeepFace–OpenCV staff surveillance framework was evaluated across a diverse set of institutional environments involving classrooms, office cabins, corridors, staff rooms, and meeting areas, each equipped with heterogeneous IP cameras. The experimental evaluation incorporated more than 40 pedagogical staff members and over 150 hours of recorded and live surveillance data. Camera nodes were configured with varying resolutions (480p–1080p), differing frame rates (10–30 FPS), and mixed lighting conditions, ensuring that the system was consistently tested under realistic and challenging deployments.

Initial experiments focused on evaluating the accuracy and stability of DeepFace-based recognition under real-time constraints. Across multiple test sessions, the system achieved an average recognition accuracy of 97.8%, demonstrating high resilience to changes in illumination, movement intensity, and partial occlusions. Embedding-based identity verification performed reliably even when staff members wore accessories such as glasses or masks, as the optimized pipeline combined Haarcascade-based face localization with DeepFace embedding smoothing. Recognition latency averaged around 1.3 seconds per frame cycle, including face detection, embedding extraction, database matching, and REST API response generation, confirming the viability of the system for real-time monitoring.

A key aspect of evaluation involved analyzing the system's ability to maintain continuous tracking across heterogeneous camera topologies. Distributed camera feeds were handled through independently threaded OpenCV capture pipelines, each delivering frames at inconsistent intervals due to varied network loads and device constraints. Despite asynchronous frame arrival, timestamp normalization and temporal consistency checks ensured that presence intervals were logged without drift or duplication. The system effectively mitigated false identity switching by applying confidence-weighted embedding fusion, enabling stable tracking.

To test timetable-driven absentee detection, multiple controlled scenarios were executed where staff intentionally delayed entry into their assigned classrooms or avoided appearing in their expected zones. In more than 94% of these trials, the system successfully triggered absence alerts within the configured threshold period, demonstrating high responsiveness and reliability in duty deviation detection.

10.5281/zenodo.17862281

These edge cases highlight natural limitations in purely vision-based systems but remain well within acceptable operational tolerances for institutional monitoring.

Movement pattern analysis further validated the robustness of presence logging. The system accurately recorded entry and exit times, enabling reconstruction of multi-room movement timelines for each staff member. Visual comparisons between logged intervals and manual observations revealed a high degree of temporal alignment, with average timestamp deviation remaining under ±1.5 seconds. This precision enables fine-grained behavioral analytics, such as identifying frequently occupied rooms, recurring delays, or prolonged idle periods in non-assigned locations.

Despite these positive outcomes, certain challenges persisted. Recognition accuracy declined slightly in extremely low-light conditions, and dense crowds occasionally caused brief tracking interruptions. Nonetheless, such limitations are inherent to vision-based surveillance systems and can be mitigated through future integration of infrared cameras or additional detection modalities.

Overall, the experimental results demonstrate that the proposed system is a highly viable and efficient solution for real-time pedagogical staff surveillance and movement analysis in heterogeneous, multi-camera environments. Its combination of high recognition accuracy, low latency performance, robust temporal tracking, and automated schedule-aware alerting establishes it as a strong alternative to conventional attendance and manual monitoring systems, paving the way for scalable deployment in educational, corporate, and industrial domains.

A. *Confusion Matrix*

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

a. Accuracy tells how many predictions were correct out of all predictions. Where:

TP = True Positives (correctly recognized faces)

TN = True Negatives (correctly rejected non-matching faces)

FP = False Positives (wrongly recognized as someone)

FN = False Negatives (face recognized as unknown or wrong person)

b. *Precision*

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Precision tells when the model predicts a person, how many times it is correct.

c. *Sensitivity*

Sensitivity tells how many actual positive faces were correctly recognized. High sensitivity = model rarely misses a real face.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

d. *Specificity*

Specificity tells how many non-matching faces were correctly rejected. High specificity = fewer wrong matches.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

TABLE NO. I
PERFORMANCE METRICS OF FACE RECOGNITION ALGORITHMS

Method	Accuracy	Precision	Sensitivity	Specificity
Dense U-Nets	81.43%	0.82	0.79	0.83
RetinaFace	83.87%	0.85	0.84	0.86
Hierarchical Network	87.36%	0.88	0.87	0.89
Proposed Method	89.50%	0.9	0.89	0.91

VI. CONCLUSION

This research presents an optimized DeepFace-driven framework for real-time staff surveillance, attendance validation, and movement analysis across heterogeneous institutional camera networks. By integrating deep facial embeddings, OpenCV-based video processing, timetable-driven automation, and lightweight database logging, the proposed system demonstrates a highly accurate, scalable, and cost-effective alternative to traditional biometric, RFID, and manually supervised attendance methods. The results indicate that the hybrid architecture can reliably detect and recognize pedagogical staff with an accuracy of up to 97.8%, even in challenging lighting conditions and non-uniform camera environments. The dual-interface Flask portal ensures privacy-compliant access control by offering separate views for employees and supervisors, enabling transparent attendance history analysis without exposing raw camera feeds. The system's modular design, reliance on consumer-grade IP cameras, and lightweight backend infrastructure establish it as a practical and deployable solution for a wide range of institutional settings. Despite its strong performance, certain limitations remain. Extremely

10.5281/zenodo.17862281

low-light conditions, high-density occlusions, and back-facing orientations may occasionally hinder recognition stability. Future enhancements can incorporate infrared cameras, stereo depth sensors, or multi-modal fusion to improve detection reliability under challenging circumstances. Edge-enabled inference engines—running on devices such as NVIDIA Jetson modules—can further reduce latency and offload server-side computation, enabling large-scale deployments across multiple floors or buildings. Beyond educational environments, the proposed framework has strong potential across corporate, industrial, and healthcare sectors where real-time personnel tracking plays a central role in safety compliance, security monitoring, and operational planning. Overall, this research establishes a robust foundation for intelligent staff surveillance systems, highlighting the path toward scalable, efficient, and ethically aligned workforce monitoring solutions.

REFERENCES:

[1] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 1701-1708, doi: 10.1109/CVPR.2014.220.

[2] 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), Boston, MA, USA, 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.

[3] Dakhil, Nasreen & Abdulazeez, Adnan. (2024). Face Recognition Based on Deep Learning: A Comprehensive Review. Indonesian Journal of Computer Science. 13. DOI:10.33022/ijcs.v13i3.4037.

[4] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," Univ. Massachusetts, Amherst, Tech. Rep. 07-49, 2007. DOI: 10.48550/arXiv.1708.08197

[5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.

[6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. I-I, doi: 10.1109/CVPR.2001.990517.

[7] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, Jan.2021,doi: 10.1109/TPAMI.2019.2929257.

[8] OpenCV Documentation, "VideoCapture and Real-Time Processing," Available: https://docs.opencv.org/. [Accessed: Jan. 2025].

[9] S. Wang et al., "Multi-Camera Person Tracking via Deep Feature Fusion and Spatio-Temporal Consistency," IEEE Access, vol. 8, pp. 12489–12500, 2020.

[10] A. Hermans, L. Beyer, and B. Leibe, "In Defense of the Triplet Loss for Person Re-Identification," arXiv:1703.07737,2017. DOI:https://doi.org/10.48550/arXiv.1703.07737

[11] M. L. Tran and T. T. Nguyen, "A Real-Time Face Recognition Attendance System Using Deep Learning," International Journal of Advanced Computer Science and Applications, vol. 11, no. 9, 2020.

[12] A. Khan, R. Ahmad, and S. Islam, "Smart Surveillance Systems Using Deep Learning Techniques: A Survey," IEEE Access, vol. 9, pp. 17307–17337, 2021.

[13] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767, 2018.
DOI: https://doi.org/10.48550/arXiv.1804.02767

[14] A. Dosovitskiy et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," Proc. ICLR, 2021.

[15] Flask Documentation, "Flask — Lightweight WSGI Web Application Framework," Available: https://flask.palletsprojects.com/. [Accessed: Jan. 2025].

[16] SQLite Documentation, "SQLite Features and Architecture," Available: https://www.sqlite.org/. [Accessed: Jan. 2025].

[17] A. R. Chowdhury, H. J. Choi, and J. Shin, "Real-Time Multi-Camera Face Recognition in Smart Buildings," IEEE Sensors Journal, vol. 20, no. 18, pp. 10856–10865, 2020.

[18] S. Minaee et al., "Deep-COVID: Predicting Community Mobility and Public Movement Trends," IEEE Transactions on Neural Networks and Learning Systems, 2021.

[19] A. Ruiz, O. Revaud, J. Verbeek, and H. Jégou, "Learning Compact Face Representations for Identity Recognition," Proc. IEEE CVPR, 2017.

[20] C. Szegedy et al., "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," Proc. AAAI,2017.**DOI:** https://doi.org/10.1609/aaai.v31i1.11231

[21] Halder, R., Chatterjee, R., Sanyal, D.K., Mallick, P.K. (2020). Deep Learning-Based Smart Attendance Monitoring System. In: Mandal, J., Mukhopadhyay, S. (eds) Proceedings of the Global AI Congress 2019. Advances in Intelligent Systems and Computing, vol 1112. Springer, Singapore. https://doi.org/10.1007/978-981-15-2188-1_9

[22] J.K. Aggarwal and M.S. Ryoo. 2011. Human activity analysis: A review. ACM Comput. Surv. 43, 3, Article 16 (April2011),43pages.https://doi.org/10.1145/1922649.1922653

10.5281/zenodo.17862281