International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Proposed Framework for a Multi-Scenario Intelligent Surveillance System in Educational Institutions

Mr. Rushikesh S. Bhalerao[1], Mr. Rahul D. Kakad[2], Ms. Chetana P. Suryavanshi [3], Mr. Devidas K. Tambe[4], Mr. Dinesh R. Thorat[5]

[1]Assistant Professor, Department of IT, SVIT
[2]Bachelors of Engineering, Department of IT, SVIT
[3] Bachelors of Engineering, Department of IT, SVIT
[4] Bachelors of Engineering, Department of IT, SVIT
[5] Bachelors of Engineering, Department of IT, SVIT

Abstract - In recent times, maintaining safety and academic integrity within educational institutions has become a critical challenge. While CCTV cameras are extensively deployed across campuses and examination halls, they rely heavily on manual monitoring, which is inefficient, prone to human error, and causes delayed responses to critical events. This paper proposes a novel, multi-scenario automated surveillance framework based on advanced deep learning and computer vision techniques. The proposed system is designed to operate in two distinct modes: a Campus Safety mode and an Academic Integrity mode. For violence detection, the architecture integrates YOLOv7 for rapid human detection with a MobileNet-BiLSTM classifier to accurately recognize violent actions. For examination monitoring, the system utilizes Deep Keyframe Detection combined with a Multilayer Perceptron (MLP) enhanced YOLOv8 algorithm (SE-YOLOv8) and a ResNet-based 3D CNN to identify subtle cheating behaviors like passing notes or unauthorized looking. By unifying these dual pipelines into a single deployable framework, this study outlines a scalable solution that minimizes manual invigilation, enhances real-time threat detection, and ensures a secure, fair educational environment.

Key Words: Multi-Scenario Surveillance, Deep Learning, Violence Detection, Smart Proctoring, YOLO Framework, Action Recognition, Deep Keyframe Detection, Computer Vision.

1.INTRODUCTION

The extensive deployment of high-definition video surveillance equipment in educational institutions aims to ensure public safety and maintain the fairness of talent selection during standardized examinations. However, public violence and academic misconduct are both dynamic, fast-occurring events. The current bottleneck in institutional security lies in the excessive reliance on the manual inspection of massive video feeds. This traditional mode of observation leads to a situation of "monitoring without detailed inspection," allowing sporadic cheating behaviors and sudden violent acts to go unnoticed until it is too late.

With the deep penetration of artificial intelligence in the educational sector, leveraging AI to empower electronic surveillance offers a robust solution. This paper proposes an intelligent, multi-scenario surveillance system. Rather than relying on singular object detection algorithms that fail to capture contextual complexity, this framework utilizes a hierarchical approach. It combines the rapid localization capabilities of YOLO-variant models with lightweight temporal classifiers to create an automated, highly responsive environment.

2. LITERATURE REVIEW

The application of deep learning in automated surveillance has seen significant growth; however, current research predominantly focuses on isolated, single-use applications rather than multi-scenario frameworks.

A. Violence Detection Systems In the domain of public safety, earlier research heavily explored models like Faster R-CNN, SSD, and YOLOv3. While models like Faster R-CNN demonstrated high accuracy, they suffered from significant computational overhead, rendering them unsuitable for real-time edge device deployment. To address speed limitations, recent frameworks have adopted YOLOv7, leveraging its extended layer aggregation networks to efficiently process dynamic environments and detect small objects. However, object detection alone is insufficient for action recognition. Recent studies propose pairing rapid detectors with lightweight classifiers, such as MobileNet or EfficientNet, to optimize performance without exhausting computational resources.

B. Smart Proctoring Systems Conversely, in the realm of academic integrity, automated invigilation systems have historically relied on conventional computer vision techniques or standalone deep learning networks. Implementations using standard algorithms often exhibit high confusion rates between normal academic behaviors and actual cheating activities. To overcome this, recent methodologies suggest enhancing the YOLOv8 architecture by incorporating Multilayer Perceptrons (MLP) and squeeze-and-excitation mechanisms (such as SENetV2) to significantly improve candidate localization. Furthermore, because cheating involves complex temporal sequences, researchers emphasize using 3D Convolutional Neural Networks (CNNs) backed by ResNet architectures to achieve precise pose estimation and action categorization.

C. Identified Gap

Despite the proven efficacy of these independent models, existing literature lacks a unified, hybrid framework capable of dynamically routing surveillance feeds to handle both distinct environments. The proposed architecture bridges this gap by integrating both pipelines into a cohesive, intelligent institutional monitoring system.

3. SYSTEM ARCHITECTURE

The logical flow of the proposed multi-scenario system is designed for continuous, uninterrupted monitoring.

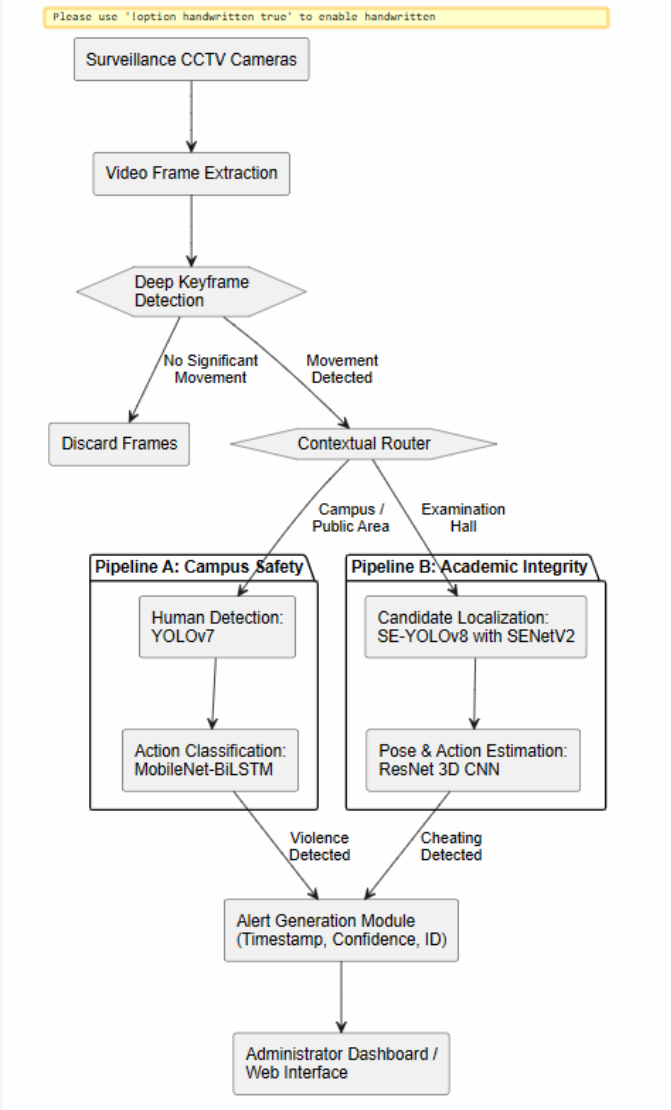

Fig-1 : System Architecture

1. **Video Input:** Real-time RTSP streams from strategically placed CCTV cameras act as the primary data source.
2. **Keyframe Extraction:** The raw feed undergoes inter-frame differencing to eliminate redundant, static frames.
3. **Contextual Routing:** Based on the camera's designated zone (Public Area vs. Examination Hall), the data is routed to the respective deep learning pipeline.
4. **Processing:** The frames undergo either YOLOv7 + MobileNet-BiLSTM processing (Violence) or SE-YOLOv8 + ResNet 3D CNN processing (Cheating).
5. **Alert System Integration:** Upon positive classification of anomalous behavior, the system generates an immediate output payload containing the camera ID, timestamp, and confidence score. This payload triggers an automated alert mechanism to notify relevant authorities for a rapid response.

4. PROPOSED METHODOLOGY

The proposed system is designed as a highly modular, hierarchical detection framework capable of processing diverse surveillance feeds. The architecture consists of three primary phases:

A. Deep Keyframe Detection and Preprocessing

Continuous surveillance feeds generate massive amounts of redundant data. In standardized environments, individuals often exhibit minimal movement, resulting in negligible inter-frame differences. Analyzing every frame computationally exhausts the system. Therefore, a Deep Keyframe Detection module is proposed to extract only the frames displaying significant behavioral changes. By computing the absolute difference of pixel values between consecutive frames and applying a dynamic threshold, the system actively filters out static moments and isolates instances of rapid, exaggerated movement indicative of cheating or violence.

B. Violence Detection Pipeline (Campus Mode)

When deployed in public areas, the filtered keyframes are routed to the violence detection module.

1. *Rapid Human Detection:* The YOLOv7 algorithm is employed to scan the frames. YOLOv7 is selected for its superior trade-off between speed and accuracy, ensuring efficient human localization even in cluttered environments. Frames lacking human presence are immediately discarded.
2. *Action Classification:* The localized human features are passed to a MobileNet-BiLSTM classifier. MobileNet serves as the spatial feature extractor, while the BiLSTM network processes the temporal sequences of the extracted features. This dual-model approach guarantees that violent actions are correctly distinguished from non-violent interactions with minimal latency.
3.

C. Cheating Recognition Pipeline (Exam Mode)

For examination halls, the system requires granular analysis to detect discreet activities.

1. *Enhanced Candidate Localization:* The system utilizes SE-YOLOv8, an improved iteration of the YOLOv8 algorithm. By replacing the standard C2f module with a Squeeze Aggregated Excitation (SaE) module from SENetV2 in the network's neck, the algorithm gains enhanced global representation learning and adaptive scaling. This ensures highly precise bounding box localization around candidates.

2. *Pose Estimation and Recognition:* The localized data is fed into an advanced ResNet architecture enhanced with 3D convolutions. Unlike 2D convolutions that lack temporal awareness, 3D CNNs capture the precise spatial-temporal dynamics of the examinee, allowing the system to accurately categorize behaviors such as passing notes, looking around, or unauthorized leaning.

5. EXPECTED OUTCOMES

The implementation of this hybrid deep learning framework is expected to yield a highly scalable and computationally efficient surveillance model. By strictly processing dynamic keyframes and utilizing lightweight classifiers like MobileNet, the system is expected to perform reliably on resource-constrained edge devices. Ultimately, this dual-scenario approach will substantially reduce the dependency on human monitoring, drastically improve incident response times, and establish a verifiable, objective standard for maintaining institutional security and integrity.

6. CONCLUSION

The reliance on manual CCTV monitoring in educational institutions leaves critical vulnerabilities in both physical safety and academic fairness. This paper proposed an automated, multi-scenario surveillance system that integrates advanced computer vision and deep learning techniques to address these flaws. By synthesizing YOLOv7 and MobileNet-BiLSTM for rapid violence detection alongside an MLP-enhanced YOLOv8 and ResNet 3D CNN for precise cheating recognition, the framework provides a comprehensive solution optimized for edge deployment. Future iterations of this project will focus on the empirical implementation of these algorithms, testing them against custom datasets, and integrating the analytical backend with a real-time web dashboard for institutional administration.

6. ACKNOWLEDGEMENT

We would like to express our profound gratitude to our Project Guide, **Mr. Rushikesh S. Bhalerao**, for their valuable guidance, continuous support, and constructive feedback throughout the conceptualization of this research. We also extend our sincere thanks to **Dr. Pratibha V. Kashid**, Head of the Information Technology Department, and **Dr. Sarang Pande**, Principal of **Pravara Rural Education Society's Sir Visvesvaraya Institute of Technology**, for providing the necessary facilities and a conducive environment to carry out this work. Finally, we are thankful to Savitribai Phule Pune University for structuring a curriculum that fosters practical research, technological innovation, and academic growth.

REFERENCES

[1] S. Senthilkumar, S. Kolte, G. Agarwal, and A. Shirish, "Real Time Violence Detection System using YOLOv7 and Deep Learning Techniques," *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, 2025, pp. 1447-1454.

[2] J. Lu, J. Wang, N. Song, Z. Luo, W. Zhang, and Y. Wang, "Cheating Recognition in Examination Halls Based on Improved YOLOv8," *2024 International Conference on Artificial Intelligence of Things and Systems (AIoTSys)*, 2024.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[4] M. Malhotra and I. Chhabra, "Automatic invigilation using computer vision," *International Conference on Integrated Intelligent Computing Communication & Security*, Atlantis Press, 2021, pp. 130-136.

[5] L. A. Siddique, R. Junhai, T. Reza, S. S. Khan, and T. Rahman, "Analysis of Real-Time Hostile
Activity Detection from Spatiotemporal Features Using Time Distributed Deep CNNs, RNNs and Attention-Based Mechanisms," *arXiv preprint arXiv:2302.11027*, 2023.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[7] L. Rahmawati, S. Rustad, A. Marjuni, M. A. Soeleman, and P. N. Andono, "Foggy-Based Object Detection In Video Using Faster R-CNN, YOLOv3, and SSD," *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)*, IEEE, 2023, pp. 412-416.