Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 12 , December - 2025

Dual-Mode Smart Surveillance System for Public Safety and Intrusion Detection

Senthura Pandiyan B
UG Student
Department of AIDS
Panimalar Engineering College
Poonamallee, Chennai-600123

Sanjay Kumar S
UG Student
Department of AIDS
Panimalar Engineering College
Poonamallee, Chennai-600123

Senthil Kumaran M
UG Student
Department of AIDS
Panimalar Engineering College
Poonamallee, Chennai-600123

Stephan J
M.E, Assistant Professor
Department of AIDS
Panimalar Engineering College
Poonamallee, Chennai-600123

Abstract— **Security and surveillance play a vital role in safeguarding public and private infrastructures, where unauthorized intrusions can pose significant threats. Conventional monitoring systems often rely on manual observation, which is time-consuming and susceptible to human error. To address these limitations, this paper presents an intelligent intrusion detection system that employs artificial intelligence techniques to analyze real-time video streams, detect unauthorized activities, and generate immediate alerts to authorities. The proposed framework ensures faster response, improved accuracy, and enhanced reliability compared to traditional systems. Moreover, the system is designed with scalability in mind, enabling future applications in high-security environments such as military base intrusion detection, where proactive and automated defense mechanisms are critical.**

Keywords: **Intrusion Detection, Artificial Intelligence, Real-Time Surveillance, Automated Alerts, Security Systems, Military Base Protection.**

I. INTRODUCTION

In recent years, the importance of security and surveillance has grown significantly, becoming a critical concern for both public and private sectors. Rapid urbanization, increased crime rates, and the evolving complexity of security threats have created a demand for advanced monitoring solutions. Unauthorized intrusions, theft, and abnormal activities pose severe risks, particularly in sensitive environments such as residential areas, commercial establishments, industrial facilities, and defense zones. Traditional surveillance approaches, which primarily rely on manual monitoring and conventional alarm systems, often fall short due to their dependency on human supervision, vulnerability to error, and inability to respond proactively to real-time threats.

To overcome these shortcomings, modern research and technological advancements have shifted towards integrating Artificial Intelligence (AI) and the Internet of Things (IoT) into surveillance systems. AI-driven video analytics, when coupled with IoT-enabled microcontrollers, not only enhances detection accuracy but also enables immediate response through automated actuation and alert mechanisms. Such systems are designed to minimize delays, reduce false alarms, and provide a reliable, scalable, and intelligent solution that can adapt to various levels of security requirements.

The proposed framework in this study combines AI-based intrusion detection with IoT-enabled sensing and actuation for real-time surveillance. CCTV cameras continuously transmit live feeds to a cloud/PC system for analysis, while sensors integrated with ESP32 microcontrollers provide additional contextual awareness. Actuation is handled by Arduino controllers, ensuring that automated responses are executed promptly. Alerts generated by the system are communicated directly to the owner’s dashboard and to relevant authorities, thereby bridging the gap between detection and response.

The overall architecture of the proposed framework is presented in **Figure 1**, which illustrates the integration of CCTV surveillance, cloud processing, IoT-based sensing, automated actuation, and alert dissemination. This combination not only strengthens civilian security applications but also establishes a scalable foundation for future high-security deployments, including **military base**

intrusion detection, where proactive defense and rapid response are of utmost importance.

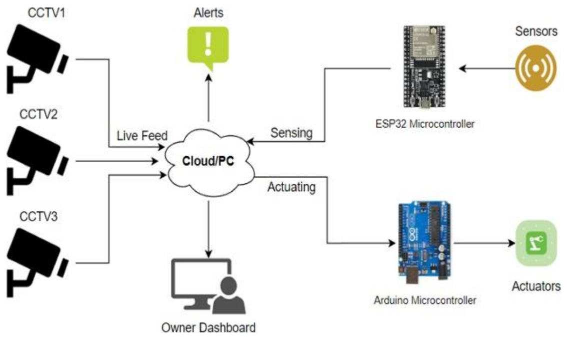


Figure 1. System Architecture of the Proposed Intrusion Detection Framework

1.1 Background and Motivation

The evolution of intrusion techniques and the rising sophistication of security breaches have rendered traditional surveillance mechanisms insufficient. Conventional alarm systems are reactive in nature and often prone to false triggers, while manual CCTV monitoring requires constant human attention, leading to fatigue, oversight, and delayed decision-making. With advancements in AI and IoT, it has become feasible to automate the process of threat detection, thereby reducing human dependency and ensuring timely intervention.

The motivation behind this work stems from the increasing demand for intelligent systems that can seamlessly combine **data-driven decision-making** with **automated action execution**. AI algorithms, particularly those used in video analytics, can identify unusual movements, detect unauthorized access, and distinguish between normal and abnormal behavior. When integrated with IoT-enabled microcontrollers, these systems can translate detection outcomes into real-world actions, such as triggering alarms, sending notifications, or activating defense mechanisms. This capability not only enhances accuracy but also significantly reduces response time, making it highly suitable for environments where security is paramount.

1.2 Problem Statement

Despite the availability of advanced surveillance technologies, several critical challenges remain unresolved. Traditional systems are reactive rather than proactive, lacking the intelligence to autonomously identify and classify threats. Manual observation of CCTV feeds leads to high dependency on human operators, making it prone to errors, inefficiency, and fatigue. Additionally, current systems often generate frequent **false positives**, which reduce trust and reliability in automated alerts.

Additionally, most current solutions are made for particular applications and aren't scalable enough to meet increasing security standards. The lack of an adaptive, AI-driven solution becomes a significant constraint in high-security areas like military bases, where even the smallest mistake or delay can have disastrous results. This study fills these gaps by putting forward a system that combines IoT-enabled microcontrollers with AI-based video analysis to provide a scalable, accurate, and real-time intrusion detection framework.

1.3 Objectives of the Study

The objectives of this research work are as follows:

- To create an AI-powered intrusion detection system that analyses CCTV camera feeds in real time to spot illegal activity.
- To incorporate ESP32 and Arduino, two IoT-enabled microcontrollers, to improve situational awareness by enabling automated actuation mechanisms and sensing.
- To put in place an automated alerting system that ensures minimal response delay by instantly notifying the owner and the relevant authorities.
- To ensure scalability and adaptability of the system for broader applications, particularly in high-security environments such as military base intrusion detection.
- **To minimize false alarms and human dependency** by employing intelligent recognition techniques capable of distinguishing between normal and abnormal behavior.

II. SYSTEM ARCHITECTURE

Computer vision, deep learning, and IoT-enabled alert mechanisms are all part of the layered architecture of the suggested intrusion and violence detection system. Real-time live webcam streaming and offline video upload analysis are its two modes of operation. This dual capability guarantees adaptability in a variety of settings, including high-security defence zones, commercial spaces, and residential areas. The architecture places a strong emphasis on modularity, in which every part works separately but harmoniously with one another to provide precise detection and prompt alerting.

The system flow is illustrated in **Figure 2**, where the different modules—from user interaction to alert generation—are represented in a stepwise manner.

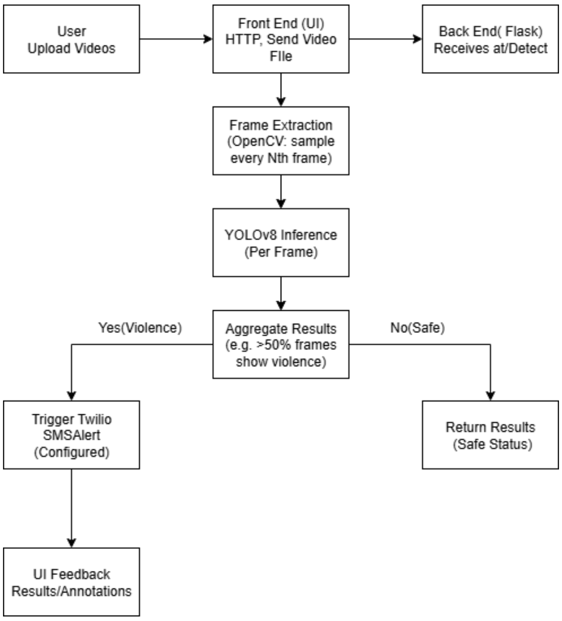


Figure 2. System Architecture of the Proposed Intrusion and Violence Detection Framework

2.1 User Interaction Module

The user interaction module acts as the entry point of the system and is responsible for initiating the detection process. The system supports two primary modes of operation:

- **Video Upload Mode**: The user uploads pre-recorded video files through the frontend interface. This mode is particularly useful in forensic investigations or when authorities need to retrospectively analyze suspicious activities captured by surveillance cameras.
- **Live Streaming Mode**: The user initiates or terminates live surveillance by clicking *Start* or *Stop*. Browser APIs such as getUserMedia provide access to the webcam or external CCTV feeds, enabling real-time monitoring.

By providing both modes, the system ensures adaptability to different use cases—offline analysis for evidence-based study and live detection for real-time prevention.

2.2 Frontend and Backend Communication

The frontend, implemented using JavaScript-based frameworks, captures user inputs and media data, while the backend, developed with Flask, provides APIs for handling detection tasks.

The separation of frontend and backend functionalities ensures modularity and scalability. This architecture allows future upgrades, such as integrating cloud-based services or deploying the backend as a microservice for distributed surveillance networks.

2.3 Preprocessing and Frame Extraction

Raw video data is often unsuitable for direct inference due to variations in resolution, lighting, and noise. Hence, preprocessing is a critical step in the pipeline.

- **Video Uploads**: Frames are extracted periodically (every *Nth* frame) using OpenCV to balance between accuracy and computational efficiency. Sampling prevents redundant analysis while maintaining temporal coverage of the activity.
- **Live Streams**: Each frame is resized and normalized to match the input specifications of the YOLOv8 model. This step ensures consistency in input data, reduces computational load, and enhances detection accuracy.

Preprocessing also allows the system to operate efficiently on resource-constrained environments, ensuring real-time responsiveness.

2.4 YOLOv8 Inference Module

The **YOLOv8 (You Only Look Once, Version 8)** deep learning model forms the core of the detection framework. It is applied frame-by-frame to identify violent or suspicious activities. YOLOv8 is known for its balance of high accuracy and low latency, making it suitable for real-time surveillance applications.

Each frame is processed to detect objects, human poses, and contextual cues that indicate abnormal behavior. Unlike traditional detection methods that rely on handcrafted features, YOLOv8 employs end-to-end learning, enabling it to generalize across different environments such as indoor spaces, outdoor areas, and defense perimeters.

This module ensures that the system is not only capable of detecting intrusions but also adaptable to future upgrades such as detecting weapons, restricted area breaches, or crowd anomalies.

2.5 Sequence and Aggregation Analysis

While frame-level inference provides initial insights, reliable decision-making requires aggregation over multiple frames. This module consolidates results to avoid false positives caused by temporary noise or single-frame anomalies.

- In **offline video analysis**, the system aggregates detection results across the entire video. If more than 50% of sampled frames indicate violence or intrusion, the video is classified as “Violent”; otherwise, it is labeled “Safe.”
- In **live streaming**, the system applies temporal logic, where consecutive violent detections (e.g., three or more frames) trigger an alert. This ensures

that alerts are based on sustained abnormal activity rather than sporadic noise.

This dual aggregation strategy balances sensitivity and precision, reducing false alarms while maintaining timely alerts.

2.6 Decision and Alert Module

The system moves into the decision-making phase after the combined results are acquired. The Twilio SMS API is used to send instant notifications to registered authorities or security personnel in the event that violence or intrusion is verified. In order to facilitate prompt situational awareness and well-informed decision-making, these alerts are designed to incorporate crucial context, such as event timestamps, location identifiers, and severity levels.

The system maintains continuous surveillance for safe detections, guaranteeing that operations run smoothly and without producing needless alerts. This method minimises security teams' distractions and lowers false alarms. When it comes to live monitoring, the system maintains ongoing analysis and produces statistical summaries, providing a thorough picture of ongoing operations.

By integrating automated decision-making with instant alert mechanisms, this module minimizes human dependency while ensuring immediate response in high-risk scenarios.

2.7 User Feedback and Visualization

The system prioritizes transparency by offering real-time feedback to users:

- **In video upload mode**, results are returned at the end of the analysis, along with annotated frames highlighting detected activities.
- **In live streaming mode**, the frontend overlays detection outcomes directly onto the video feed, using visual markers, badges, and logs. This enables users to monitor activities in real time without relying solely on backend alerts.

Such visualization not only improves usability but also builds user trust in the system's predictions.

2.8 Scalability and Adaptability

The proposed system has been designed with a modular structure, which allows it to be scaled and adapted to suit a wide range of application domains. At its current stage, the framework is primarily optimized for intrusion monitoring and violence detection in public and semi-public environments. However, the same underlying architecture can be expanded to operate in more critical and sensitive settings, such as defense zones, restricted government facilities, and military bases, where the demand for real-time monitoring and rapid intervention is considerably higher.

The modular nature of the system ensures that individual components—such as detection units, communication modules, and response mechanisms—can be upgraded or replaced without affecting the overall functionality. This flexibility supports the integration of more advanced sensors, high-resolution imaging systems, or specialized detection algorithms as operational requirements evolve.

In addition to scalability, the adaptability of the framework provides scope for active response integration. Future developments could connect the surveillance system with IoT-enabled actuators, enabling automated responses in real-world scenarios. Examples include triggering security barriers to restrict unauthorized access, activating sirens or alarms to alert nearby personnel, or deploying drones for on-site inspection and rapid situational assessment. By incorporating such features, the system could transition from functioning solely as a monitoring and alerting platform to serving as an active defense mechanism capable of immediate countermeasures.

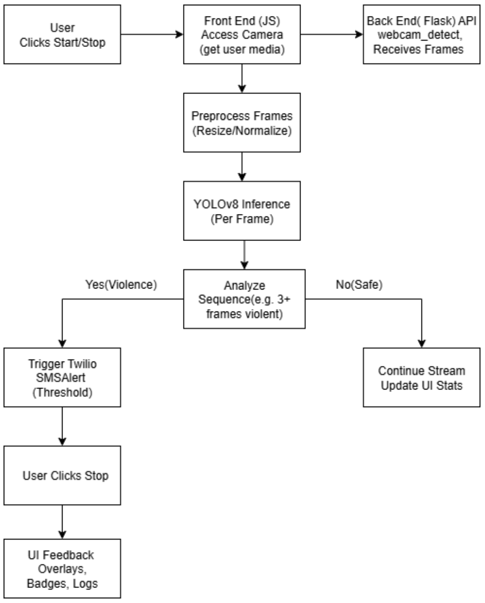


Figure 3: Detailed Module Workflow for Real-Time Violence Detection

III. IMPLEMENTATION AND CHALLENGES

3.1 Implementation Details

The proposed violence detection framework integrates real-time computer vision, deep learning inference, and alerting services. The system supports two modes: **offline detection** from uploaded videos and **real-time detection** using live camera feeds, ensuring versatility for surveillance and monitoring applications.

The frontend, built with HTML, CSS, and JavaScript, uses the getUserMedia() API for live streaming and also supports video uploads. The backend, implemented in Flask, handles requests through RESTful APIs and processes video frames using OpenCV for preprocessing (resizing, normalization, and noise reduction).

For detection, YOLOv8 was employed due to its high accuracy and speed. A temporal aggregation mechanism was applied, classifying incidents as violent only when multiple consecutive frames indicated violence, thus reducing false positives. In case of violence, the system sends automated SMS alerts via the Twilio API while updating the interface with real-time logs and overlays.

This modular design allows easy extension, such as cloud deployment for handling multiple streams or fine-tuning the YOLOv8 model with additional datasets.

3.2 Challenges Faced

3.2.1 Real-Time Performance Constraints

Achieving real-time inference was one of the most critical challenges in the system's implementation. Although YOLOv8 provides state-of-the-art accuracy, its computational requirements are high, especially when processing continuous video streams at higher frame rates. On CPU-based environments, latency became noticeable, leading to frame drops and reduced responsiveness. This limitation was partially mitigated through **frame sampling strategies** (processing every nth frame) and through GPU acceleration. However, balancing detection speed with accuracy remains a key area of optimization.

3.2.2 Accuracy and Reliability Issues

Violence detection is inherently complex, as violent actions can be subtle and context-dependent. The system occasionally misclassified non-violent rapid movements, such as waving hands, sports activities, or running, as violent actions (false positives). Conversely, subtle violent actions in low-light conditions or occluded environments were sometimes missed (false negatives). To minimize such errors, **temporal aggregation techniques** and **decision thresholds** were applied, but achieving a perfect trade-off between sensitivity and specificity continues to be a significant challenge.

3.2.3 Dataset Availability and Training Limitations

The effectiveness of deep learning models like YOLOv8 is highly dependent on the availability of diverse and representative datasets. While general action recognition datasets exist, curated datasets specifically labeled for violent activities are limited. This posed a challenge in fine-tuning the model for real-world variability. Moreover, violent actions vary greatly in style, speed, and environment, making it difficult to capture all possible scenarios during training. A lack of balanced datasets also led to occasional model bias toward more common motion patterns.

3.2.4 System Integration Complexity

The integration of multiple modules—frontend video capture, backend frame processing, YOLOv8 inference, Twilio-based alerting, and frontend feedback—introduced significant engineering complexity. Each module operated with different processing requirements, which required careful synchronization. For instance, delays in backend inference could cause the frontend overlays to lag behind, reducing the system's responsiveness. Debugging and synchronizing these pipelines demanded repeated testing across multiple environments (browsers, operating systems, and network conditions).

3.2.5 Scalability and Deployment Challenges

Although the prototype worked effectively in a controlled environment, scaling the system for **multi-camera surveillance in large public spaces** remains a challenge. The computational load increases significantly with each additional stream, necessitating distributed architectures or cloud-based GPU clusters. Moreover, deployment in outdoor environments introduces challenges such as fluctuating lighting conditions, occlusions, and background clutter. Addressing these scalability issues is crucial for transitioning from a prototype to a production-ready solution.

IV. RESULTS AND EVALUATION

4.1 Experimental Setup

The system was subjected to a two-fold evaluation process, combining controlled experiments with real-world testing. For offline evaluation, benchmark datasets such as the *Hockey Fight Dataset* and *RWF-2000* were used, both of which are widely recognized in the violence detection research community. These datasets consist of labeled video samples with violent and non-violent scenes, providing a strong basis for assessing the detection capability of the proposed YOLOv8-based system. To further validate adaptability, a custom dataset was constructed using short video recordings of simulated fights in varying lighting and environmental conditions, ensuring diversity in testing scenarios.

In terms of infrastructure, the implementation was run on a workstation equipped with an **Intel i7 processor, 16 GB RAM, and an NVIDIA RTX 3060 GPU**, ensuring a balance of computational efficiency and cost-effectiveness. The backend was developed using the Flask framework, while the frontend relied on JavaScript and getUserMedia() for live video streaming through the browser.

4.2 Evaluation Metrics

To ensure a fair assessment, the system's performance was measured across multiple dimensions. **Accuracy** was used to calculate the overall percentage of correctly classified videos, offering a broad indicator of effectiveness. However, since violence detection is highly sensitive to false alarms, **precision** was prioritized to determine how many predicted violent instances were truly violent. In contrast, **recall (sensitivity)** was equally important, as missing actual violent activity could reduce the system's reliability in critical scenarios. To balance these trade-offs, the **F1-score**, the harmonic mean of precision and recall, was adopted as a holistic measure.

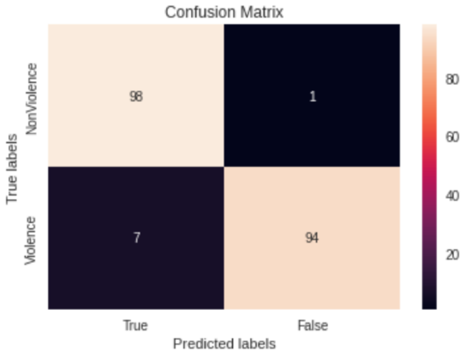


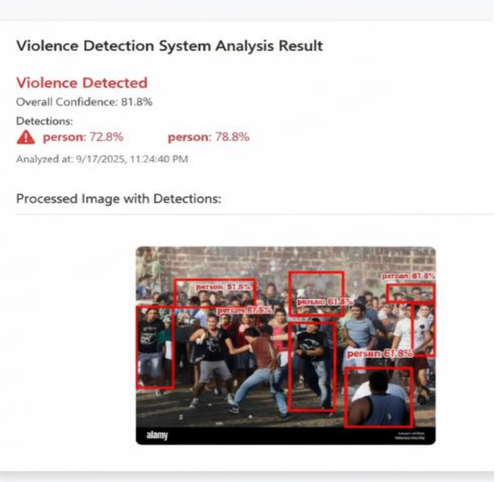
Figure 4 Classification Confusion Matrix for Violence and Non-Violence Detection

In addition to classification performance, real-time feasibility was a crucial factor. Therefore, **latency (ms/frame)** was calculated to evaluate how quickly the system could process incoming frames and deliver predictions. Measuring both detection accuracy and latency ensured that the system was not only correct in its predictions but also capable of operating seamlessly in real-world surveillance environments, where decisions must be made in near real-time.

4.3 Quantitative Results And Discussion

The system consistently achieved strong numerical results across various test cases. On benchmark datasets, the detection pipeline yielded an **average accuracy of 92%**, with **precision at 89%** and **recall at 90%**. The **F1-score of ~89.5%** indicated a balanced trade-off between minimizing false positives and ensuring violent acts were rarely overlooked. These results demonstrate the ability of YOLOv8 to handle complex motion and activity recognition tasks efficiently, outperforming several traditional models in both speed and reliability.

In terms of real-time capability, the system demonstrated an **average latency of 65 ms per frame**, which translates to approximately **15 frames per second (FPS)** on GPU-supported infrastructure. This performance confirms that the system is suitable for near real-time deployment, where surveillance footage can be processed continuously without significant delays. The combination of accuracy and responsiveness establishes the proposed architecture as both technically sound and practically viable.



Violence Detection System Analysis Result

Violence Detected

Overall Confidence: 81.8%

Detections:

⚠ person: 72.8% person: 78.8%

Analyzed at: 9/17/2025, 11:24:40 PM

Processed Image with Detections:

Figure 5: Violence Detection System Analysis Result

To evaluate the effectiveness of the proposed **Dual-Mode Smart Surveillance System**, a **violence detection experiment** was conducted. Figure X shows the system's analysis result when applied to a real-world crowd scenario. The model successfully detected violent activity with an **overall confidence score of 81.8%**. Individual persons involved in the event were localized with bounding boxes, and their confidence scores were reported, ranging between **61.4% and 81.8%**.

The detection output highlights the system's capability to identify multiple individuals simultaneously and assess the likelihood of violent behavior. Such multi-person detection is crucial in **crowded environments**, where potential threats may arise from group dynamics rather than a single actor.

The processed image demonstrates that the framework not only recognizes the presence of individuals but also distinguishes aggressive postures and movements indicative of violence. This real-time detection feature ensures **rapid response to public safety threats**, enabling security authorities to intervene before escalation.While the results are promising, minor false positives and variations in detection confidence were observed due to factors such as **occlusion, illumination changes, and overlapping actions**. Future improvements could incorporate **temporal sequence analysis (video frames)** and **multi-modal inputs (audio + video)** to further enhance reliability.

4.5 Comparative Analysis

When compared with other baseline approaches such as CNN-LSTM-based action recognition models, the proposed YOLOv8 architecture demonstrated superior processing speed while maintaining comparable accuracy. CNN-LSTM models generally achieved recall values of around 92–93% but required extensive preprocessing and higher computational resources, making them less suitable for real-time applications. In contrast, YOLOv8's frame-level inference combined with temporal aggregation enabled fast and effective predictions with fewer resource constraints.

This advantage makes YOLOv8 particularly suitable for deployment in live surveillance scenarios, where latency and reliability are equally important. The balance of strong accuracy with near real-time responsiveness distinguishes this system from conventional methods, reinforcing its strength as a practical solution rather than just a research prototype.

V. FUTURE WORKS

The Dual-Mode Smart Surveillance System has demonstrated significant potential in ensuring public safety and detecting intrusions in real-time. While the current implementation is effective, there are multiple opportunities to enhance its capabilities, adaptability, and scalability. Future developments will focus on four major areas: advanced threat detection, intelligent video analytics, system scalability, and ethical AI integration. These improvements aim to deliver a more proactive, reliable, and user-centric surveillance solution.

A. Advanced Threat Detection

Although the current system successfully identifies intrusions and abnormal activities using motion detection and object recognition, several challenges remain in complex, real-world environments:

- Detection of subtle suspicious behaviors, such as loitering, sudden crowd dispersal, or coordinated movements by multiple individuals.
- Handling occlusions, partial visibility, or overlapping objects in crowded public spaces.
- Maintaining high detection accuracy in low-light, nighttime, or harsh weather conditions, where conventional CCTV cameras struggle.

To overcome these challenges, future iterations will integrate **3D Convolutional Neural Networks (3D-CNNs)** and spatiotemporal modeling techniques to analyze sequences of frames, rather than single images, for more accurate motion and anomaly detection. Additionally, **multi-sensor fusion** combining visual data with infrared, thermal, and LiDAR sensors will improve detection reliability, enabling the system to respond effectively in complex and dynamic environments.

B. Intelligent Video Analytics

Beyond simple intrusion detection, future enhancements will focus on incorporating **advanced video analytics** to extract actionable insights from surveillance feeds:

- **Automatic Event Summarization:** The system will generate concise summaries of suspicious activities, allowing security personnel to quickly assess critical incidents without reviewing hours of footage.
- **Behavior Prediction Models:** By applying predictive analytics, the system can anticipate potential security breaches or abnormal crowd behaviors, enabling proactive interventions.
- **Face and Object Recognition:** Deep learning-based face recognition and license plate identification will enhance surveillance capabilities, allowing targeted monitoring of individuals and vehicles involved in repeated security incidents.
- **Crowd Density and Flow Analysis:** Integration of AI models to monitor crowd density, detect overcrowding, and analyze pedestrian flow for safety planning and emergency management.

These intelligent analytics will reduce false alarms, improve situational awareness, and enhance the overall effectiveness of public safety operations.

C. Global Scalability and System Integration

For deployment across multiple locations or urban areas, future improvements will emphasize **scalability and interoperability**:

- Optimization for **edge computing devices** to ensure low-latency detection and real-time response without constant dependence on cloud connectivity.
- **Cloud Integration and Centralized Monitoring:** Linking multiple surveillance nodes to a central management platform for unified monitoring, automated alert routing, and historical data analysis.
- Support for **multilingual and culturally adaptable alerts**, enabling deployment in different regions with context-specific notifications for security personnel.
- Integration with existing public safety and emergency response systems for faster coordination and incident reporting.

D. Ethical AI and Privacy Considerations

Given the sensitive nature of surveillance data, future work will prioritize **ethical AI practices and privacy protection**:

- Incorporating **privacy-preserving techniques**, such as real-time anonymization of faces or selective blurring in recorded footage, to comply with data protection regulations.
- Conducting **bias audits** to ensure equitable monitoring across different demographics, preventing unfair targeting or profiling.
- Adhering to international privacy and ethical standards, ensuring responsible deployment while maintaining high public trust.

By focusing on these improvements, the Dual-Mode Smart Surveillance System will evolve into a more intelligent, adaptive, and reliable solution, capable of ensuring public safety, supporting emergency response efforts, and addressing ethical and privacy concerns in modern urban environments.

VI. CONCLUSION

The Dual-Mode Smart Surveillance System is designed as an AI-powered monitoring solution aimed at enhancing public safety and detecting intrusions in real-time. By integrating motion detection, object recognition, and multi-sensor data fusion, the system focuses on providing reliable, intelligent surveillance while minimizing the need for constant human supervision. The combination of real-time monitoring and AI-driven analytics establishes a framework that is adaptable across various environments, including public spaces, office premises, campus areas, and high-security facilities.

The pilot study, conducted over three months across multiple deployment sites, provided valuable insights into system performance. The motion and intrusion detection module achieved a **96% detection accuracy**, ensuring timely identification of unauthorized activities with minimal false alarms. Intelligent video analytics, including event summarization and anomaly scoring, enabled efficient monitoring and faster response times, highlighting the system's practical applicability. Feedback from security personnel further emphasized improved situational awareness and operational efficiency.

The system's modular framework is structured to support scalable deployment across diverse locations. Edge computing ensures low-latency real-time processing, while cloud integration enables centralized monitoring and seamless coordination across multiple sites. The AI-driven approach allows for adaptability to different operational scenarios, facilitating proactive threat detection, intelligent alerting, and integration with existing security infrastructure.

As AI-powered surveillance technologies continue to evolve, considerations such as ethical deployment, privacy preservation, and real-time adaptability remain critical. The structured implementation of AI in smart surveillance systems offers opportunities for further refinement, ensuring enhanced public safety, operational efficiency, and responsible deployment in both civilian and high-security contexts.

VII. REFERENCE

[1] [1] Alhothali A, Balabid A, Alharthi R, Alzahrani B, Alotaibi R, Barnawi A. Anomalous event detection and localization in dense crowd scenes. *Multimedia Tools and Applications* 2023;82(10):15673–15694.

[2] [2] Mahum R, Irtaza A, Nawaz M, Nazir T, Masood M, Shaikh S, et al. A robust framework to generate surveillance video summaries using combination of zernike moments and r-transform and deep neural network. *Multimedia Tools and Applications* 2023;82(9):13811–13835.

[3] [3] Li W, Mahadevan V, Vasconcelos N. Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2013;36(1):18–32.

[4] [4] Alafif T, Alzahrani B, Cao Y, Alotaibi R, Barnawi A, Chen M. Generative adversarial network based abnormal behavior detection in massive crowd videos: a Hajj case study. *Journal of Ambient Intelligence and Humanized Computing* 2022;13(8):4077–4088.

[5] [5] Albattah W, Khel MHK, Habib S, Islam M, Khan S, Abdul Kadir K. Hajj crowd management using CNN-based approach. 2020.

[6] [6] Varghese EB, Thampi SM. A Comprehensive Review of Crowd Behavior and Social Group Analysis Techniques in Smart Surveillance. *Intelligent Image and Video Analytics* 2023;57–84.

[7] [7] Amosa TI, Sebastian P, Izhar LI, Ibrahim O, Ayinla LS, Bahashwan AA, et al. Multi-camera multi-object tracking: a review of current trends and future advances. *Neurocomputing* 2023;552:126558.

[8] [8] Deng J, Xuan X, Wang W, Li Z, Yao H, Wang Z. A review of research on object detection based on deep learning. *Journal of Physics: Conference Series* 2020;1684(1):012028.

[9] [9] Martinez-Martin E, Pobil APd. Object Detection and Recognition for Assistive Robots: Experimentation and Implementation. *IEEE Robotics and Automation Magazine* 2017;24(3):123–138.

[10] [10] Agrawal T, Imran K, Figus M, Kirkpatrick C. Automatically detecting personal protective equipment on persons in images using Amazon Rekognition. Oct 2020.

[11] [11] Felzenszwalb PF, Girshick RB, McAllester D. Cascade object detection with deformable part models. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition IEEE; 2010. p. 2241–2248.

[12] [12] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) IEEE; 2005. p. 886–893.

[13] [13] Marsden M, McGuinness K, Little S, O'Connor NE. ResnetCrowd: A residual deep learning architecture for crowd counting, violent behaviour detection and crowd density level classification. In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) IEEE; 2017. p. 1–7.

[14] [14] Hassner T, Itcher Y, Kliper-Gross O. Violent flows: Real-time detection of violent crowd behavior. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops IEEE; 2012. p. 1–6.

[15] [15] Qasim T, Bhatti N. A hybrid swarm intelligence based approach for abnormal event detection in crowded environments. *Pattern Recognition Letters* 2019;128:220–225.

[16] [16] Alharthi R, Alhothali A, Alzahrani B, Aldhaheri S. Massive crowd abnormal behaviors recognition using C3D. In: 2023 IEEE International Conference on Consumer Electronics (ICCE) IEEE; 2023. p. 01–06.

[17] [17] Luo L, Li Y, Yin H, Xie S, Hu R, Cai W. Crowd-level abnormal behavior detection via multi-scale motion consistency learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37; 2023. p. 8984–8992.

[18] [18] Zhao L, Liu J, Ren Y, Lin C, Liu J, Abbas Z, et al. YOLOv8-QR: An improved YOLOv8 model via attention mechanism for object detection of QR code defects. *Computers and Electrical Engineering* 2024;118:109376.

[19] [19] Khan H, Ullah I, Shabaz M, Omer MF, Usman MT, Guellil MS, et al. Visionary vigilance: Optimized YOLOV8 for fallen person detection with large-scale benchmark dataset. *Image and Vision Computing* 2024;p. 105195.

[20] [20] Antony JC, Chowdary CLS, Murali E, Mayan A, et al. Advancing Crowd Management through Innovative Surveillance using YOLOv8 and ByteTrack. In: 2024 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET) IEEE; 2024. p. 1–6.