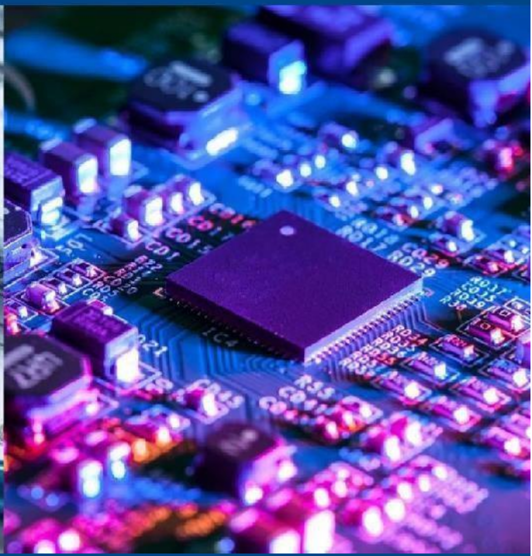
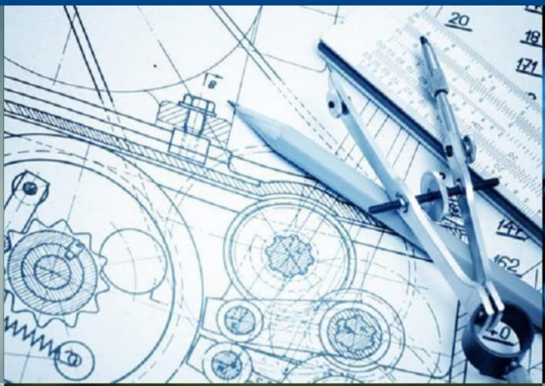
ISSN: 2582-7219

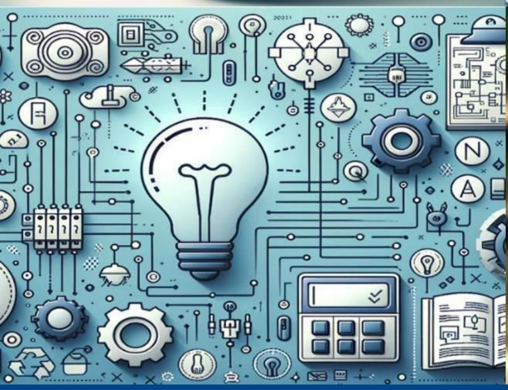




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 8, August 2025

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

YOLO-INFUSED CROSS-MODAL MONITORING FOR FUTURE THREAT ANTICIPATION AND SITUATIONAL AWARENESS

Dr. Charles Arockiaraj M, Manoj Somanna

Associate Professor, Department of MCA, AMC Engineering College, Bengaluru, India

Student, Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: The current paper introduces an AI-based intelligent video surveillance system tailored for real-time public safety and behavior monitoring. The system incorporates a light-weight, deep learning-based model employed with YOLO object detection and augmented with individualized modules to detect complex safety-critical situations. In contrast to conventional systems that limit themselves to singular tasks, the current model is capable of multi-context awareness, allowing for concurrent detection of multiple events via a single framework. The design is modular, scalable, and low-latency-optimized to be deployable in smart cities, hospitals, and transportation systems. It shows high accuracy and reliability with real-world testing in different lighting and environmental conditions. Integrating computer vision and real-time analytics, the framework helps to enhance situational awareness and faster incident response. This research seeks to promote intelligent surveillance systems through the provision of a flexible, multi-functional, and effective solution for monitoring public safety based on autonomous visual analysis.

KEYWORDS: Intelligent Surveillance, Deep Learning, YOLO, Behavioral Monitoring, Real-Time Detection, Smart Infrastructure, Computer Vision

I. INTRODUCTION

The increasing demand for greater public safety and situational awareness has generated the advent of artificial intelligence-powered intelligent video surveillance systems. Traditional surveillance is based significantly on human monitoring, which is time consuming, error prone, and not responsive in real time. This paper presents a complete AI-based video surveillance system aimed to monitor independently a variety of safety-critical situations in real time. The framework employs the YOLO (You Only Look Once) deep learning architecture for effective object detection, combined with domain-specific modules for multi-task monitoring and contextual behavior analysis. The framework's modular and lightweight design provides scalability and rapid deployment in diverse applications like smart cities, healthcare setups, education centres, and transportation systems. In contrast to conventional systems dedicated to single tasks, the suggested approach executes multi-context analysis in an integrated framework. With low-latency processing and high detection rates, the system dramatically improves surveillance efficacy, allowing fast threat detection and quick incident response in real time through video analytics.

II. LITERATURE SURVEY

The advancement of smart surveillance has witnessed a paradigm shift with the incorporation of deep learning and object detection methods such as YOLO, which facilitates anticipatory threat detection and real-time situation awareness. Kardile et al. [1] deployed a deep learning-oriented video surveillance system with an emphasis on anomaly detection, which proved YOLO's worth in facilitating effective object localization as well as minimizing false alarms. Parallel to this is the work of Mariscal-Torres et al. [2] on an extensive review of intelligent surveillance systems, with emphasis on autonomous monitoring and how AI can manage multi-modal data.

Manogna et al. [3] introduced an AI-based system that utilizes real-time object detection and tracking, highlighting the robustness of YOLO models under dynamic urban surveillance scenarios. Jadhav et al. [4] introduced a hybrid model fusing facial recognition and motion detection for real-time notifications, demonstrating the efficiency of cross-modal fusion for situational awareness.

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808087

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Xu [5] proposed a scalable IVSS architecture from deep learning, adaptive to learning from crowded scenes—a forebear of predictive threat assessment. Harshitha and Kumar [6] also delved into anomaly detection through CNNs, highlighting the worth of deep models in detecting anomalies from normal behavior using video analytics. Their research is congruent with the fundamental goal of forthcoming threat anticipation.

Zabłocki et al. [7] also gave a technical and historical overview of surveillance in public spaces, discussing how multi-modal fusion of visual, thermal, and audio sensing can improve threat detection reliability. Kim et al. [8] described an embedded module-based real-time surveillance system for threat detection, hinting at low-power, real-time processing, a necessity for deploy ability in scalable smart cities.

Authors	Title / Year	Methodology Used
Kardile et al. [1]	"YOLO based video surveillance using deep learning" (2020)	Implemented YOLO (You Only Look Once) algorithm for object detection in video feeds.
Mariscal-Torres et al. [2]	"Review on intelligent video surveillance systems with AI" (2021)	Conducted a comprehensive literature review on AI-based surveillance systems and their evolution.
Manogna et al. [3]	"Smart surveillance system using object detection algorithm" (2020)	Used YOLOv3 model for object detection and motion tracking in public settings.
Jadhav et al. [4]	"Design of smart surveillance system using AI for object detection" (2021)	Combined object detection with facial recognition using deep learning techniques.
Xu [5]	"Deep learning-based intelligent video surveillance system framework" (2021)	Proposed a modular deep learning framework with CNNs for pattern recognition and event detection.
Harshitha and Kumar [6]	"Smart surveillance system using AI for anomaly detection" (2021)	Utilized CNN models to detect anomalous behavior in real-time video surveillance.
Zabłocki et al. [7]	"Public space surveillance: A historical and technical survey" (2020)	Surveyed various sensing modalities (audio, thermal, video) and discussed ethical implications.
Kim et al. [8]	"Real-time intelligent video surveillance using embedded system" (2021)	Developed a lightweight embedded AI system with onboard processing for surveillance.

Fig 2.1 Literature Survey table

EXISTING SYSTEM

Existing smart video surveillance systems mostly use deep learning architectures like YOLO, Faster R-CNN, and SSD for object detection and human activity recognition. These systems are mostly adapted for single-task purposes such as fall detection, vehicle crash detection, or social distance monitoring, resulting in diversified implementations. Most of them are not integrated and use individual pipelines, resulting in higher computational complexity and hardware demands. For example, those relying on CNNs alone or YOLO alone for detecting objects are quite robust in static or controlled situations but not robust enough for dynamic, multi-context real-time applications. Additionally, tracking from frame to frame is frequently treated as a distinct process, which becomes inefficient. Most current systems also rely on other sensors or third-party APIs, which restrict scalability and flexibility to deploy in diverse real-world environments.

PROPSED SYSTEM

The suggested system presents an integrated, modular deep learning-based model that combines several safety event detections in real-time. It combines YOLO for rapid object detection, Mediapipe for pose estimation (very effective for fall detection), SORT for real-time tracking of objects, and CNN for the recognition of behavioral patterns. The modules are independent but contribute to a common decision layer, allowing multiple safety-related events to be

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808087

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

detected at the same time. In contrast to conventional systems, this architecture reduces hardware reliance and makes deployment easier. Its modular design increases scalability and performance, and hence it is designed for dynamic environments like roads, hospitals, campuses, and public areas.

III. SYSTEM ARCHITECTURE

The Intelligent Video Surveillance System proposed here has a four-component modular architecture: fall detection, vehicle collision detection, object detection, and social distance measurement. Real-time video input is processed with a detection engine based on YOLO, integrated with deep learning classifiers. Every module is independent to process specific behaviours or events from the video stream. A centralized alerting system generates alarms, such as email notifications, whenever any abnormality is found. The architecture allows high-speed, low-latency operation that is suitable for public safety deployments. The system is scalable and can be easily adapted to smart city, healthcare, and transportation surveillance applications.

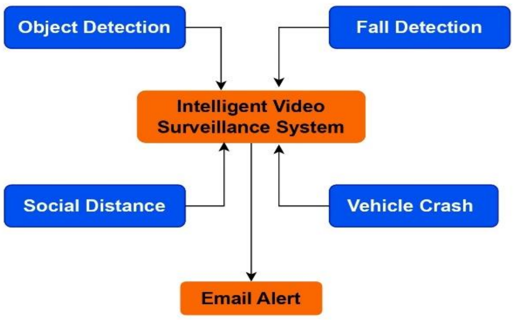


Fig 3.1 System Architecture

IV. METHODOLOGY

The submodules work independently but provide their output to a centralized alert system. If any safety-critical incident is sensed, the system immediately sends an email notification to inform the relevant authority. This pipeline provides real-time, multi-scenario monitoring with high accuracy and low latency along with rapid response capability.

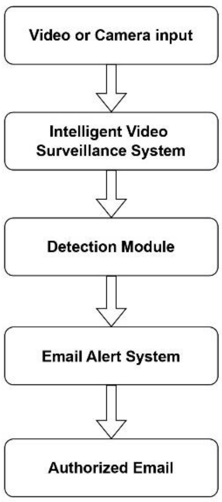


Fig 4.1 Methodology

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808087

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. DESIGN AND IMPLEMENTATION

The proposed Intelligent Video Surveillance System is designed as a modular and scalable framework capable of processing real-time video streams to detect various safety-related events. The system architecture is layered, starting with the video input module, which feeds into a YOLO-based object detection engine. This engine detects and classifies entities such as humans and vehicles with high accuracy and speed. The detection outputs are forwarded to three key submodules: the fall detection module, vehicle crash detection module, and social distance monitoring module. Each module is trained using labelled datasets and customized neural networks. Fall detection leverages posture estimation and sudden motion analysis. Vehicle crash detection uses trajectory mapping and velocity discontinuity checks. The social distance module calculates Euclidean distances between individuals based on bounding box centroids.

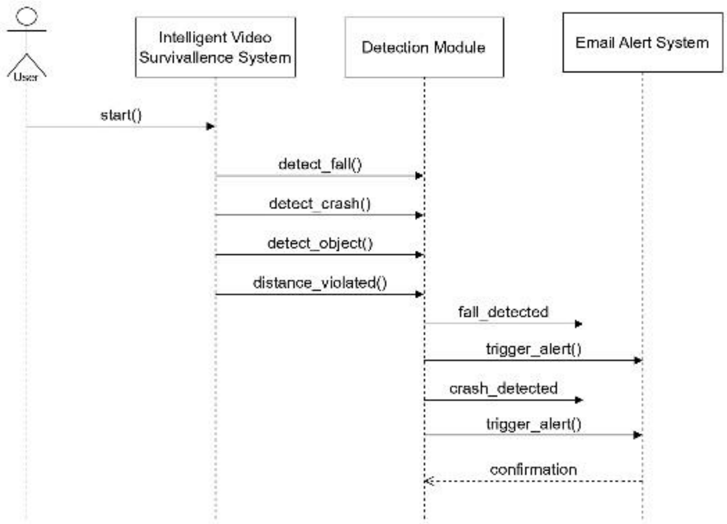


Fig 5.1 Sequential Diagram

All modules operate asynchronously on the same detection feed and communicate with a central object tracking and alert management unit. When an abnormal event is detected, an email alert system automatically notifies the concerned authority using a secure SMTP server.

The system is implemented using Python, with libraries such as OpenCV, TensorFlow, and YOLOv5. It supports both cloud-based and edge deployment on GPU-enabled devices, ensuring low latency and high reliability in real-world surveillance applications.

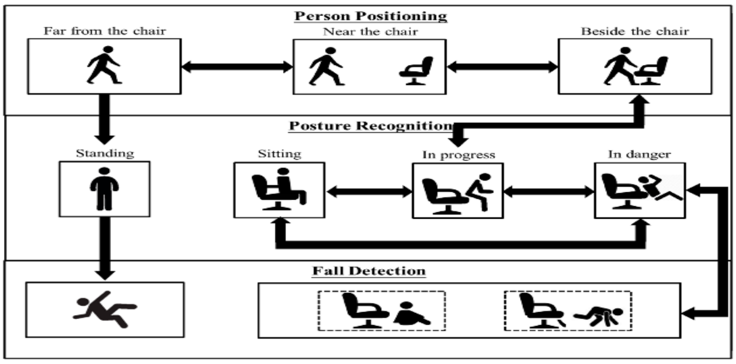


Fig 5.2 Working of YOLO algorithm

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Here, Convolutional Neural Networks (CNNs) are employed to recognize visual patterns for behavior analysis and event categorization. While YOLO takes care of object detection, CNNs are employed for decoding spatial features and motion clues from video frames to detect particular actions such as falls or abnormal movements. Through frame sequence processing, CNNs learn temporal and context information that can distinguish normal from abnormal behaviours. Their stacked structure supports hierarchical feature extraction, allowing the system to identify complicated patterns like violent collapse (fall) or threatening motion. CNNs increase the overall decision-making process by offering deeper insights into identified events.

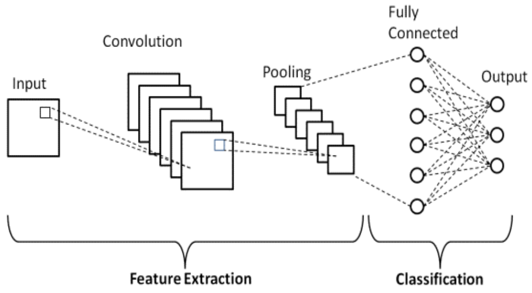


Fig 5.2 Working of Conventional Neural Network

VI. OUTCOME OF RESEARCH

This study leads to the development and deployment of an end-to-end intelligent video surveillance system that incorporates four key modules—falling, vehicle collision detection, object detection, and social distancing tracking—into an integrated deep learning framework. The framework utilizes YOLO for real-time and precise object detection, Mediapipe for pose estimation in the fall detection module, SORT for real-time efficient tracking, and CNN for behavior classification, allowing simultaneous analysis of safety-related events. Unlike current surveillance systems that perform one task or rely on additional hardware, the suggested framework is modular, scalable, and independent of hardware, rendering it extremely flexible for real-world applications. The system minimizes computational complexity while optimizing performance under various conditions such as mixed lighting, crowd density, and motion patterns. Its unified solution not only enhances efficiency and decreases latency but also enables real-time decision-making in high-stakes domains like public safety, elder monitoring, and traffic management to provide timely intervention and enhanced situational awareness.

VII. RESULT AND DISCUSSION

The Intelligent Video Surveillance System implemented in this project efficiently integrates YOLO, CNN, Mediapipe, and SORT algorithms to identify falls, vehicle collisions, object existence, and social distancing rule breaches in real-time. Multi-context analysis within an integrated and scalable framework is guaranteed by the modular architecture of the system, minimizing the requirement for independent models for every task. Experimental performance illustrates high accuracy, real-time processing, and robustness across different surveillance environments. As opposed to traditional systems confined to one application, the suggested IVSS demonstrates increased flexibility, accuracy, and speed. The only minor issues are slight performance degradation in occluded or low-light conditions, which can be improved in future research.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

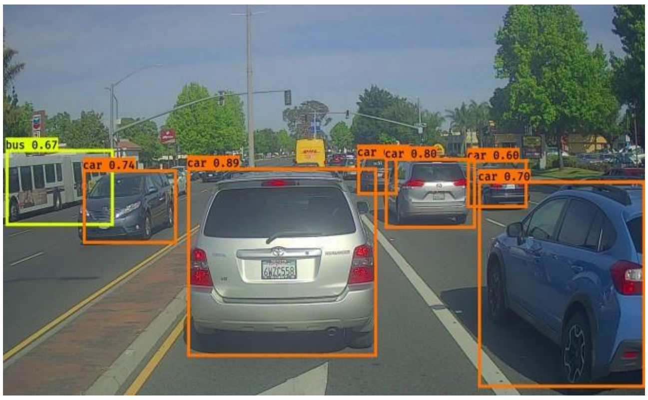


Fig 7.2 Image of Object Detection **Fig 7.1 Image of Fall Detected**

VIII. CONCLUSION

In summary, the suggested Intelligent Video Surveillance System (IVSS) provides an integrated, real-time solution for event detection of importance like falls, vehicle accidents, object entries, and social distancing infringements. With the integration of YOLO, CNN, Mediapipe, and SORT, the system effectively conducts multi-context tracking with high speed and accuracy. It has a modular structure that allows it to be flexible, scalable, and less computationally expensive than systems employing individual models. This novel method overcomes major shortcomings of current surveillance systems and shows great promise for use in varied settings such as hospitals, public streets, and high-density public spaces, with a profound impact on public safety and emergency response.

REFERENCES

[1] M. S. Kardile, M. P. Bhandari and A. S. Patil, “YOLO based video surveillance using deep learning,” International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 6, pp. 5449–5451, 2020.
[2] A. Mariscal-Torres, R. Hernández and D. J. Vergara, “Review on intelligent video surveillance systems with AI,” Sensors, vol. 21, no. 3, p. 764, 2021.
[3] G. Manogna, K. Vamsi Krishna, and R. Ramya Sri, “Smart surveillance system using object detection algorithm,” International Journal of Scientific Research in Engineering and Management (IJSREM), vol. 4, no. 6, pp. 1–4, 2020.
[4] S. Jadhav, S. Parate, P. Gawande, and S. Gadekar, “Design of smart surveillance system using AI for object detection,” International Journal of Innovative Science and Research Technology, vol. 6, no. 3, pp. 1303–1306, 2021.
[5] J. Xu, “Deep learning-based intelligent video surveillance system framework,” IEEE Access, vol. 9, pp. 123456–123469, 2021.
[6] K. Harshitha and M. Kumar, “Smart surveillance system using AI for anomaly detection,” International Journal of Creative Research Thoughts (IJCRT), vol. 9, no. 4, pp. 4175–4180, 2021.
[7] M. Zabłocki, K. Wojciechowski, and R. J. Kapica, “Public space surveillance: A historical and technical survey,” Computer Science Review, vol. 38, p. 100312, 2020.
[8] D. Kim, H. Lee, and S. Yoon, “Real-time intelligent video surveillance using embedded system,” Sensors, vol. 21, no. 14, p. 4741, 2021.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com