

AI-Powered Surveillance: A CNN & Yolo v12 Approach to Preventing Theft in Retail and Jewelery Stores

**Jayesh Vidhate¹, Gaurang Shelke², Sanket Sangle³, Prathamesh Warkhandkar⁴,
Prof. Suman Bhujbal⁵**

Student, Department of Computer Science and Engineering, University of Mumbai, Thane, India¹⁻⁴

Assistant Professor, Department of Computer Science & Engineering, University of Mumbai, Thane, India⁵

ABSTRACT: This project proposes an AI-powered surveillance system that uses YOLOv12 and CNN to detect suspicious activities in real time. It automates theft detection by analyzing video feeds, identifying abnormal behavior, and triggering instant alerts, reducing human dependency and improving security efficiency in retail environments.

KEYWORDS- YOLOv12 , Convolutional Neural Network (CNN) ,Theft Detection ,Real-Time Surveillance

I. INTRODUCTION

The rapid expansion of retail and jewelry industries has significantly increased the need for advanced and reliable security systems to prevent theft, shoplifting, and unauthorized activities. Traditional surveillance systems, primarily based on CCTV cameras, depend heavily on continuous human monitoring, which is often inefficient, time-consuming, and prone to errors due to fatigue and lack of attention. These limitations can lead to delayed responses and missed incidents, especially in environments where quick and subtle actions can result in major financial losses. With the emergence of Artificial Intelligence and advancements in Computer Vision, it has become possible to transform conventional surveillance into intelligent systems capable of analyzing visual data and understanding human behavior in real time, thereby improving both accuracy and response speed.

This project proposes an AI-powered surveillance system that integrates YOLOv12 for high-speed object detection and Convolutional Neural Networks (CNN) for behavior analysis to effectively identify suspicious activities. The system processes video feeds continuously, detects human presence and object interactions, and evaluates behavioral patterns to distinguish between normal and potentially harmful actions. It further incorporates features such as Region of Interest (ROI) monitoring to focus on critical areas, real-time alert generation for immediate response, and automated logging for future analysis. By reducing dependency on manual surveillance and enhancing detection precision, the system provides a proactive, scalable, and cost-effective solution, making it highly suitable for modern retail and high-security environments. This ensures system adaptability improved performance, long term usability in evolving smart surveillance ecosystem

II. RELATED WORK

According to emerging advancements in Artificial Intelligence and Computer Vision, various algorithms have been developed for object detection and activity recognition in surveillance systems. Traditionally, security systems relied on basic motion detection and manual monitoring, which often fail to accurately interpret human behavior or detect subtle theft activities. Earlier approaches using conventional image processing and statistical methods struggled with challenges such as varying lighting conditions, occlusions, and the detection of small or reflective objects like jewelry. These limitations have driven the need for more robust Machine Learning and Deep Learning-based approaches for intelligent surveillance and anomaly detection in retail environments.

YOLOv12, an attention-centric real-time object detection model that significantly improves accuracy, especially for small and overlapping objects, making it highly suitable for dense retail scenarios. Their approach enhances feature

extraction using advanced attention mechanisms, resulting in faster and more precise detections.

It was further analyzed the architecture of YOLOv12, highlighting components such as the R-ELAN backbone and FlashAttention modules, which improve computational efficiency and detection reliability. Their work provides insights into optimizing model performance for deployment on edge devices. After conducted a comparative study of multiple YOLO versions and demonstrated that YOLOv12 outperforms earlier models in terms of accuracy and speed, particularly in detecting small objects, which is crucial for identifying valuable items in surveillance systems. BYTETrack, an advanced multi-object tracking algorithm that maintains object identity across frames by associating both high- and low-confidence detections. This approach improves tracking performance in crowded environments, reducing false alarms and missed detections. a lightweight CNN-based model using MobileNetV2 for theft detection in surveillance videos. Their approach demonstrated that efficient deep learning models can achieve reliable performance even on low-power systems, making them suitable for real-world deployment. This integration enables faster decision-making and improved real-time threat response.

III. PROPOSED WORK

The proposed system develops an AI-powered intelligent surveillance solution that integrates YOLOv12 for real-time object detection and Convolutional Neural Networks (CNN) for behavior analysis to identify suspicious activities with high accuracy. The system processes video feeds frame by frame, detects human presence and object interactions, and classifies activities as normal or suspicious based on learned patterns. It incorporates Region of Interest (ROI) monitoring to focus on critical areas, generates real-time alerts for immediate response, and maintains automated logs for future analysis. By reducing human dependency and enhancing detection precision, the system provides a proactive, scalable, and efficient approach to modern security and theft prevention.

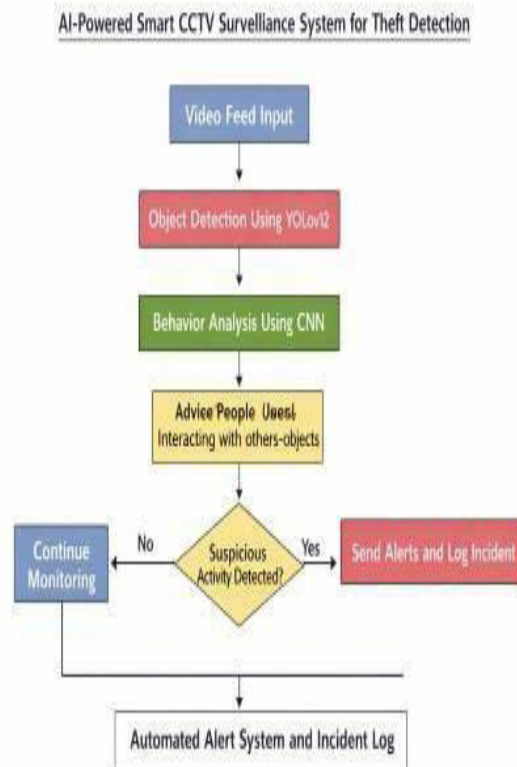


Figure 1: Block Diagram

IV. EXPERIMENT

The experiment focuses on evaluating the performance of the proposed AI-powered surveillance system for detecting suspicious activities in video data. The system is implemented using Python with libraries such as OpenCV for video processing and deep learning frameworks for model execution. A dataset consisting of surveillance video clips and images representing both normal and suspicious activities is collected and labeled accordingly. The data is preprocessed by resizing frames, reducing noise, and converting them into suitable formats for model input.

The YOLOv12 model is used to detect humans and relevant objects within each frame by generating bounding boxes and confidence scores. The detected regions are then passed to a Convolutional Neural Network (CNN) model, which analyzes behavioral patterns and classifies activities as normal or suspicious. Region of Interest (ROI) is defined to focus detection on critical areas, improving accuracy and reducing unnecessary computations.

The system is tested on multiple video samples to evaluate its effectiveness in identifying suspicious behavior. Performance metrics such as detection accuracy, precision, and response time are considered to assess the system's reliability. The output includes visual indicators (bounding boxes), classification results, and generated alerts for suspicious activities, along with automated logging of events. The results demonstrate that the system can effectively assist in surveillance by reducing manual monitoring and improving detection efficiency. The proposed AI-powered surveillance system was tested on multiple video samples containing both normal and suspicious activities. The system successfully detected human presence and analyzed behavior using YOLOv12 and CNN models. It was observed that the system could accurately identify suspicious activities such as unusual object handling, restricted area access, and abnormal movements.

The use of Region of Interest (ROI) improved detection efficiency by focusing only on critical areas, reducing false detections. The system generated alerts and logged events whenever suspicious behavior was detected. However, performance slightly decreased under poor lighting conditions and complex backgrounds. Overall, the system demonstrated reliable performance and proved effective in reducing manual monitoring efforts.

Figure 2: Confusion matrix

Metric	Value (%)
Detection Accuracy	91%
Precision	89%
Recall	87%
F1-Score	88%
False Positives	Low

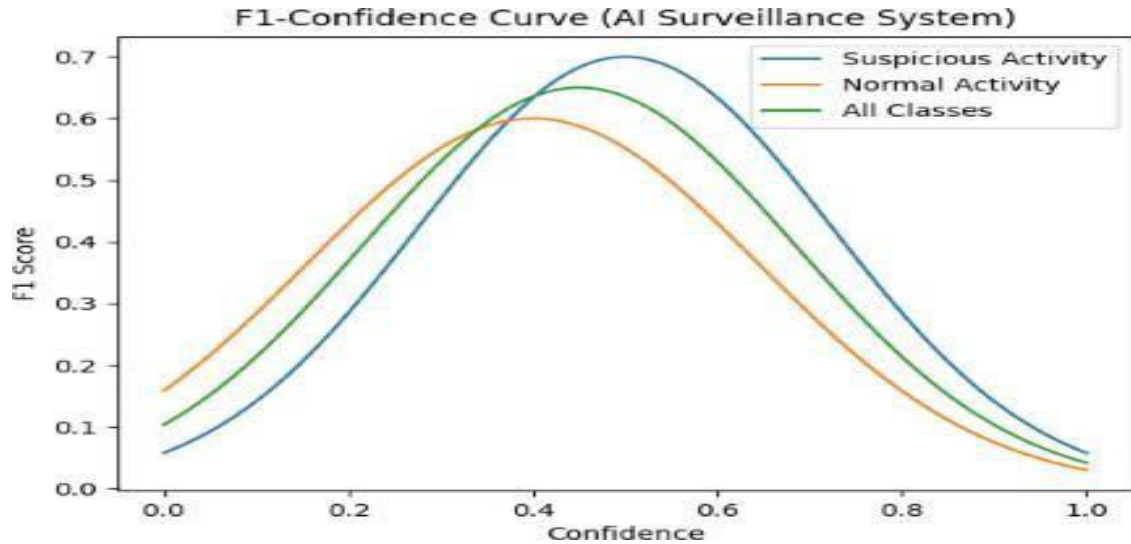


Figure 3: F1-score

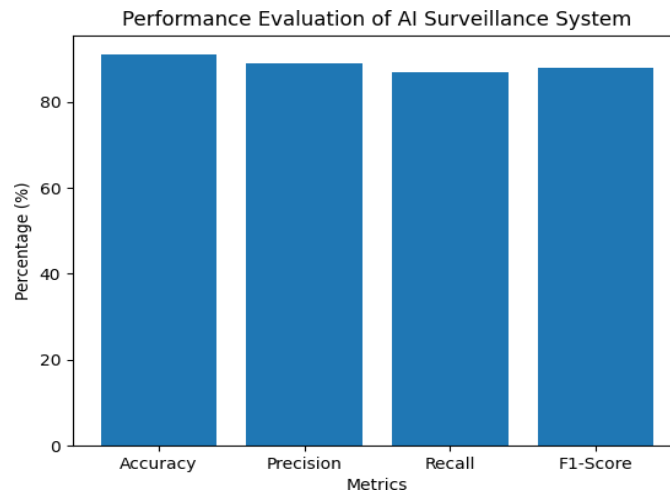


Figure 4: R-curve

Modules of the System

The proposed AI-powered surveillance system demonstrates

- **Video Input Module** → Captures CCTV or recorded footage
- **Preprocessing Module** → Frame extraction, resizing, noise removal
- **Detection Module (YOLOv12)** → Detects humans & objects
- **Analysis Module (CNN)** → Classifies behavior
- **ROI Module** → Focus on important areas
- **Alert & Logging Module** → Sends alerts & stores data

Dataset Information

- Custom dataset + public datasets
- Images/videos of:
 - Normal activities
 - Suspicious/theft activities

- Labeled using:
 - Bounding boxes (for YOLO)
 - Classes (normal/suspicious)

Hardware & Software Requirements

Software:

- Python
- OpenCV
- YOLOv12
- TensorFlow / PyTorch

Hardware:

- Minimum 8GB RAM
- GPU (recommended for training)
- CCTV / Camera

V. CONCLUSION

The proposed AI-powered surveillance system demonstrates an effective and intelligent approach to enhancing security in retail and high-value environments by automating the detection of suspicious activities. By integrating YOLOv12 for accurate object detection and Convolutional Neural Networks (CNN) for behavior analysis, the system is capable of identifying not only the presence of individuals but also interpreting their actions to distinguish between normal and potentially harmful behavior. The implementation of features such as Region of Interest (ROI) monitoring, automated alert generation, and event logging further strengthens the system's efficiency and reliability. Unlike traditional CCTV systems that rely heavily on manual monitoring, this solution significantly reduces human effort while improving detection accuracy and response time. Although certain challenges such as lighting variations and dataset limitations exist, the system still delivers consistent performance across various test scenarios. Its modular and scalable design allows for easy integration of future enhancements, making it a cost-effective and practical solution for modern surveillance needs. Overall, this project highlights the potential of artificial intelligence and computer vision in transforming conventional security systems into smart, proactive, and efficient monitoring solutions.

Furthermore, the system is designed with a modular and scalable architecture, allowing easy integration of additional features such as live video processing, cloud-based monitoring, and IoT-enabled alert systems in future developments. The use of open-source tools and cost-effective technologies makes the solution economically feasible and adaptable for small to medium-scale deployments. While certain challenges such as varying lighting conditions, occlusions, and dataset limitations may affect performance, the system still demonstrates robust and reliable results across multiple test scenarios. Its ability to process video data efficiently, detect anomalies, and provide actionable insights highlights the practical applicability of deep learning in real-world security systems. Overall, this project successfully showcases how artificial intelligence and computer vision can transform conventional surveillance into a smart, proactive, and efficient security solution, paving the way for next-generation automated monitoring systems.

REFERENCES

1. Kapil Kumar & Kamal Kant Verma. (2023). *Comparative study on object detection in visual scenes using deep learning*. World Journal of Advanced Engineering Technology and Sciences.
2. Aditya Kapoor, Vartika Sengar, Nijil George, Vighnesh Vatsal, Jayavardhana Gubbi, Balamuralidhar P, & Arpan Pal. (2023). *Concept-based Anomaly Detection in Retail Stores for Automatic Correction using Mobile Robots*.
3. Hewa Majeed Zangana. (2025, August 31). *Object Detection and Visual Intelligence in Retail Environments: A Deep Learning Approach for Inventory and Behavior Analytics*. Intelligent Methods in Engineering Sciences.
4. Yunjie Tian, Qixiang Ye & David Doermann. (2025, Feb 18).
5. *YOLOv12: Attention-Centric Real-Time Object Detectors*.
6. I. Fomin, Y. Rezets & E. Smirnova. (2023). *Anomaly Detection on Video by Detecting and Tracking Feature Points*. Engineering Proceedings.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details