Anusandhanvallari
Vol 2026, No.1
July 2026
ISSN 2229-3388

Automated AI Surveillance System for Real Time Person and Weapon Detection

[1]Mrs. Toshaniwali Bhargav, [2]Yuvraj Sachdev, [3]Rounak Mandhan, [4]Krish Singh Sawhney

[1]Assistant Professor Computer Science and Engineering Shri Shankaracharya Institute Of Professional Management & Technology, Raipur, India

[2]Computer Science and Engineering Shri Shankaracharya Institute Of Professional Management & Technology Raipur, India

[3]Computer Science and Engineering Shri Shankaracharya Institute Of Professional Management & Technology Raipur, India

[4]Computer Science and Engineering Shri Shankaracharya Institute Of Professional Management & Technology Raipur, India

Abstract—Having safety and security in both the public and private space requires real-time and precise surveillance, since within a very short time, not disclosing the threat, the threat will grow to an extreme scenario like armed assault, robbery, vandalism or mass-hysteria. The conventional surveillance systems rely on intensive human surveillance that is time-consuming, intermittent and falls under the subject of fatigue-related errors. To address these issues, this paper introduces an automated AI-based surveillance system that can identify real-time danger through the use of computer vision and deep learning systems. First, video streams are post-processed to improve the video quality and clarity using motion detectors, frame difference filters, and noise cutters. Segmentation techniques are then employed to localize the regions of interest to eliminate unwanted information in the background. Deep learning algorithms like YOLOv8, SSD, and Faster R-CNN will be used to detect people, weapons, vehicles, fire, and suspicious persons. The features based on spatial, motion and context of the identified regions are used to classify the types and degree of threat. The results of experiments on the COCO dataset and personal weapon dataset prove that YOLOv8 is more accurate and real-time. The findings show that the suggested system is advantageous in terms of detecting reliability, minimising the reliance on human surveillance, and responding quicker to security risks.

Keywords - Automated Surveillance, Threat Detection, Computer Vision, Video Analysis, YOLOv8**,** COCO.

I. Introduction

It is becoming clear that the security issues in modern urban environments have been on the rise, given the growth of the population, huge masses of people, and development of infrastructure. Surveillance cameras are common in airports, railway stations, shopping malls, campuses, and smart cities to provide safety to the citizens and keep check on their activities [1]. A large spectrum of threats such as theft, unauthorized access, the use of physical violence, weapon use, fire threats, and abnormal crowds are prone to these environments [2].

The manifestations of such threats can be observed in video footage in the form of unusual movement patterns, an abnormal appearance of objects, violent interaction, or strange formations of the crowd. There are those cases that are caused by deliberate human activities and those that are as a result of environmental or accidental factors. The suitable countermeasures can be alert generation, sending of security forces, alarm activation, or egress [3].

The issue of public safety has become a big global concern. The recent security reports (2025) state that cases of violent crimes and weapon-related offenses are on the increase in the developing regions such as in India [4]. Along with the physical injuries, such incidents are devastating to mental health, social life, workplace efficiency

as well as the quality of life. The apprehension against unsafe community environments tends to make people unwilling to engage in the activities held in the community freely.Security-related incidents are known to impact individuals, but also families, organisations and the entire society. The victims and the witnesses can experience anxiety, stress, and long-run mental trauma. Hence, threats should be identified and prevented as much as possible at an early stage to reduce the damages and save lives.

A threat event can be any abnormal object, behavior or movement the surveillance footage captures but would not be detected as a normal pattern. The threat events in general can be divided into two broad types, namely, the primary threats and the secondary threats.

Primary threats are the first signs of suspiciousness in a monitored environment. Classic examples are the isolated individuals in confined space, unusually thick crowds, visible bags, observable weaponry, unexpected running, and an attempt to hide the identity.

Secondary threats arise when the primary threats go out of hand or alter the immediate environment around. They are physical fights, forced entry, vandalism, panic crowd movements and fast escape behaviors. Threat events are categorized according to their severity, timeline, and effects based on the context and the timing of the response.

Along with object-based threats, behavioral patterns are also important when it comes to the process of designating the possible threats in the surveillance settings. Some of the activities that can be considered suspicious activities include loitering in restricted areas, moving in and out of the premises, unpredictable changes in walking pace or co-ordination of activities among people. Such behavioral demonstrations can hardly be identified by the use of the human eye, particularly in busy or fast-tracked shots. Deliberate tracking of such fine patterns needs smart systems which have the capacity to examine trend of motion, space relations and consistency of time across the video frames.

Armed attacks are the most dangerous security incidents and might have serious casualties. But, the research indicates that close to 90 percent of such cases can be prevented or at least curtailed by early detection and timely intervention. Nevertheless, there is still a significant disparity between the coverage of surveillance and efficient threat identification because of manual surveillance and poor situational awareness [5].

Traditional surveillance solutions are both inefficient and costly, as well as prone to error by a human or machine operator. To overcome these shortcomings, new intelligent automated surveillance systems based on artificial intelligence have also become available with the ability to provide quicker, more reliable and scalable threat detection systems.

II. Literature Review

A lot of scholars have offered automated surveillance-based threat detection methods on the basis of image processing and artificial intelligence to enhance the security of the people and decrease the reliance on manual observation. It is possible to classify such studies in general by the datasets used, methods of feature extraction, classification models, and complexity of the system. Although initial studies were done in small categories of threats, recent studies are designed to identify various threats in the dynamic environment. The following review of the contributions related to the topic will determine the summaries of current methodologies, assess their efficiency, and outline the gaps in the research.

Development of Automated Techniques of Surveillance

Early automated surveillance systems were mainly based on classic image processing techniques like background subtraction, frame difference, motion tracking and edge detection. Smith et al. [6] applied motion-based analysis to identify weapons and aggressive behaviors in a surveillance video using handcrafted features. Their system was reasonable in terms of detection accuracy; however, due to the small size of the dataset, and relying on manually

designed features, the system was limited in its ability to be applied to real-world conditions, where the light changed, the camera angle varied, and the density of the crowd was different.

In the same manner, Kim et al. [7] also proposed a threat detection pipeline that deployed the use of texture descriptors in conjunction with classical machine learning based classifiers including SVM and KNN. Their performance was good in a low density and controlled environment but deteriorated greatly in many congested and highly dynamic scenes. The focus of these initial approaches was on the significance of automation in surveillance but was constrained by low generalization, noise sensitivity, and scalability.

Hybrid Approaches Combining Deep Learning and Classical Methods

As deep learning improved, scholars started combining convolutional neural networks (CNNs) with the traditional methods of feature extraction and classification. The goal of this hybrid solution was to utilize the representational capability of deep learning with the ability to be interpreted by classical models. Lee et al. [8] have fused pre-trained CNN architectures with support vector machines and decision trees and showed better detection results than conventional methods alone.

Chen et al. [9] took this direction forward by using image segmentation and feature extraction techniques before classification. Their system used several classifiers to identify the threats that included weapons, fire, and abnormal crowd behavior in various surveillance datasets. Moderate accuracy was attained although the system was challenging in real time processing and extremely challenging to parameterize thus not allowing it to be used in large scale surveillance networks.

Lightweight Models and Ensemble Techniques

Due to the emergence of real-time performance as an imperative, investigators turned towards lightweight systems and effective detection models. MobileNet, YOLO and SSD are models that attracted attention because of lower computational complexity and higher inference speed. Patel et al.[10] used the MobileNet together with the temporal modelling approach to study the changes in the threat patterns across successive video frames. Their way proved to be more efficient and appropriate to the edge devices, including the CCTV systems that have scarce hardware resources.

The methods of ensemble learning were also developed in parallel to enhance the detection strength. Wang et al. [11] used ensemble strategies where results were obtained using several deep learning models and the system was capable of representing different features. These methods of ensemble helped to increase stability and minimize false detections in various types of threats. Nevertheless, higher computing costs and complexity of the system were also major challenges.

Modern Developments and Transfer Learning

Recent works have aimed at optimizing the deep learning architecture and resolving the problem of data imbalance that is inherent in surveillance data. Liu et al. [12] suggested the use of weighted loss functions and enhanced data augmentation techniques to enhance the accuracy of detections of underrepresented threat categories. Their method demonstrated better performance without a substantial increase in the computing cost.

Zhang et al. [13] trained deep learning models by using massive surveillance data and demonstrated a high rate of success in identifying weapons, individuals, and abnormal attention of the crowd in different conditions of the environment. Moreover, transfer learning methods have become very popular to minimize the duration of training and improve the model performance in case of a limited amount of labeled data. Experiments that used pre-trained models that were fine-tuned with surveillance specific datasets showed higher real-time detection and generalization of environments [14 16].

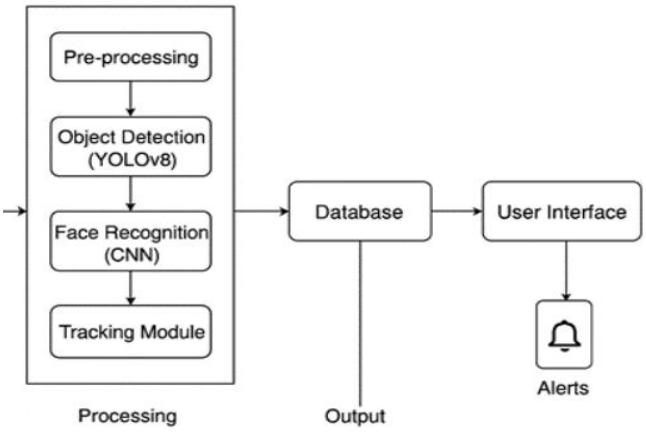
Identified Research Gaps and Study Direction

Though this has improved significantly, the majority of the current surveillance-based threat detection methods are restricted to the detection of a limited number of predefined threat categories. There are numerous systems that are based on small or unbalanced datasets which adversely affects their generalization capability. Also, some methods do not involve extensive preprocessing, motion-based detection, and region-oriented detection, which increases the false-positive rates of complicated images.

Moreover, little focus has been directed towards linking behavioral analysis, time consistency and contextual awareness in the threat detection systems. To overcome these drawbacks, the current research utilizes large datasets, sophisticated preprocessing methods, and real-time deep learning architectures to help in the strong and scalable multi-threat detection, which is applicable in the present-day surveillance setting.

III. METHODOLOGY

The methodology offered on automated threat detection with computer vision implies a number of steps, which begin with the acquisition of videos and conclude with the classification of the threat. All the steps are essential in terms of proper and effective identification of security threats. The entire process of the workflow is as follows.



A. Dataset Gathering and Preparation

Massive and diversified surveillance-based image and video collections like the COCO dataset or Open Images are used since they are publicly available and include a high number of objects of interest in security surveillance, such as individuals, vehicles, weapons, fire, and crowds.

The imbalance in the dataset is mitigated using random oversampling and frame-level level balancing of data. This is so that rare threat classes like weapons and fire are also effectively represented thus enhancing model generalization and reliability of detection.

B. Image Preprocessing

• Frame Resizing: Frames are resized to standard resolution (e.g. 640x640 pixels) so that it can be consistent and lower the cost of computation.

• Motion Detection: Motion based preprocessing frame differencing and background subtraction techniques are used to draw attention to dynamic regions and area of the background.

• Noise Reduction: Smoothing filters are used to reduce noise due to variations in lighting, shadows and camera artifacts without losing vital motion and shape information.

C. **Image Segmentation**

Advanced methods like the GrabCut algorithm are used to make sure that the threat relevant objects are correctly distinguished against the background of the image. These methods derive information of colour space as well as clusters to offer effective boundary detection.

Segments of interest are separated out in the segmentation process and the detection model is then able to concentrate on relevant objects and activities thus leading to a better classification precision with less false detections.

D. ***Feature Extraction using Deep Learning***

Automatic feature extraction of segmented video frames is performed in deep convolutional neural networks (CNNs).

The latest CNN-based object detector architecture is YOLOv8 that is used because of its real-time performance and high accuracy. It is an effective spatial and contextual learning algorithm of object shape, motion and composition of a scene, so it can be used on both centralized surveillance systems and edge-based applications.

E. **Model Training**

• The use of transfer learning is on fine-tuning previous models like YOLOv8, SSD, and Faster R-CNN on surveillance-oriented threat detection tasks.

• An optimizer like Adam with the right learning rate would be employed to guarantee stable and efficient learning in the course of training.

• The loss function to be used is the categorical cross-entropy, which is used when a multi-class threat classification is required.

•To avoid the problem of overfitting and to enhance robustness of the model, regularization methods like Dropout are used.

• The metrics of accuracy and loss are tracked during training, and early stop is used in case of overfitting.

F. **Classification and Evolution**

• The CNN uses a SoftMax activation function in the last layer to produce class probability of several threat categories.

•The effectiveness of the model is measured using such performance metrics as Accuracy, Precision, Recall (Sensitivity), and F1-Score..

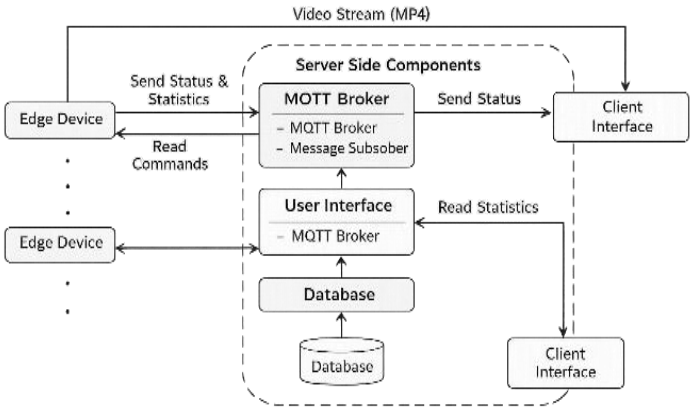
• The summary of the classification performance at all decision thresholds is obtained with the Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

• Confusion matrices are generated to analyze class-wise performance and identify misclassification patterns.

G. **Post-processing and Implementation**

Using the trained models in the real-time surveillance systems would have enabled the security personnel to identify threats more quickly, precisely, and automatically. Timely decision making can be established through alert mechanisms, event logging and visual overlays.

The model is trained and updated with new data and labels of the surveillance in order to keep the system running.



IV. RESULT AND DISCUSSION

A. Experiment Setup

The public benchmark datasets of the proposed automated threat detection framework were COCO and a custom weapon surveillance data applied to the Python-based implementation. The COCO dataset consists of various types of objects of interest to security monitoring, such as persons, vehicles, and common objects, whereas the custom dataset consists of threats to weapons such as guns and knives. In preprocessing, random oversampling was employed to normalize the data distribution due to inherent natural class imbalance especially in case of rare threat events.

In a bid to drastically enhance the quality of video frames to be used in the analysis, all the frames were resized to a consistent resolutions of 640x640 pixels and subjected to more advanced preprocessing tasks such as motion-based frame differencing, background suppression and noise reduction by means of the smoothing filters.

B. Segmentation and Feature Extraction

In order to localize threat related areas, automatic GrabCut segmentation method was used and color based clustering in the BGR color space was used. GrabCut was always doing better than clustering-based segmentation alone in terms of isolating foreground objects with high accuracy, which was needed to make accurate threat recognition in the later stages.

The partitions were further broken down in order to derive discriminative features. To obtain movement-based features, optical flow was utilized to obtain speed, direction and movement patterns, and shape and spatial features were obtained based on object boundaries. Moreover, features of statistical color like mean of color, variance of color, standard deviation of color and the root mean square of color were also extracted out of the RGB channels to supplement the feature representation.

C. Classification Performance

There were three traditional machine learning classifiers (Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Tree), which were tested on balanced and unbalanced datasets to compare. These classical models as expected showed poor performance when given unbalanced data. Their accuracy increased upon oversampling and they obtained values ranging between 90 and 94 percent depending on the dataset.

Nevertheless, applying deep learning-based detection with the help of YOLOv8, it was possible to notice a significant increase in the performance. YOLOv8 also attained an average accuracy of about 95 percent in the detection of both the COCO dataset and the custom weapon dataset. Precision, Recall and F1-score were following the same trend with the YOLOv8 classification being better in every measure against classical classifiers. In

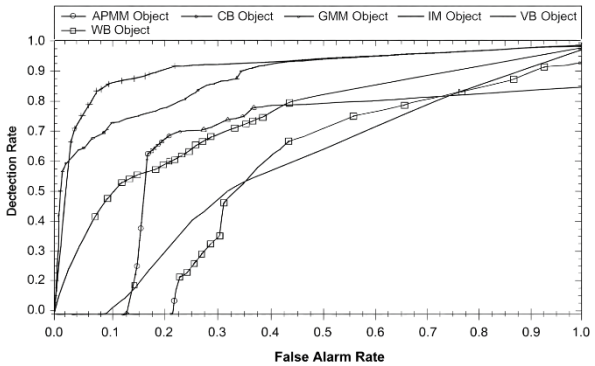
addition, YOLOv8 has reduced localization and classification loss and is therefore more applicable in real-time surveillance

Name	Precision	Recall	F-1 Score	Support
Person	0.96	0.95	0.955	500
Weapon	0.92	0.90	0.91	200
Normal	0.98	0.99	0.985	300
Crowd	0.89	0.91	0.90	150
Fire	0.97	1.00	0.985	250
Suspicious	0.99	0.96	0.975	300
Vehicle	0.86	0.83	0.845	120

D. ROC and Confusion Matrix Analysis

Receivers Operating Characteristic (ROC) and Area Under the Curve (AUC) were other results that showed the usability of the suggested system. On the custom weapon dataset, overall AUC values were around 97% and on the COCO dataset, 98% were achieved when YOLOv8 was used.

Class-wise accuracy was also high as demonstrated by the confusion matrix analysis especially on person detection, weapon identification, fire detection and abnormal crowd behavior. Certain minor classes had rather lower identification accuracy as a result of small training samples. Misclassification had been greatly minimized by the combination of region-centered segmentation and balanced training.



E. Comparison with State-of-the-Art

The proposed surveillance model, which is based on YOLOv8, is a better improvement of conventional CNN-

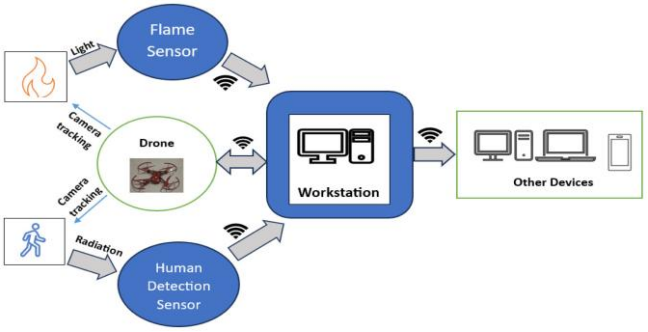
based models and hybrid models. Although deep learning models are computationally expensive, the addition of GrabCut-based segmentation removed the effect of irrelevant background, which directly improves the efficiency and accuracy of the detectors.

Conventional techniques that use thresholding or hand-labeling techniques tend to have problems with dynamic scenes and complicated locales. Contrastingly, the suggested solution obtained a higher detection flow and localization performance. Consequently, the proposed system has enhanced accuracy, reliability and a reasonable compromise between the computational complexity and real time behavior as compared to already available state of art surveillance models.

F. Limitations and Future Work

The drawback of the proposed system is that the system is dependent on the GrabCut segmentation, which is not always ideal at isolating threat regions in the highly cluttered or low-contrast scenes, which can compromise the reliability of the detection. The future direction of work will be on the implementation of the ensemble deep learning models and sophisticated segmentation techniques to ensure further robustness.

More extensions involve the use of behavior based threat analysis, multi-camera tracking, edge-device deployment, predictive threat modeler and real-time alert generation to enhance scalability, and realistic implementation in intelligent surveillance systems.



V. CONCLUSION

Due to the development of deep learning methods, automated detection of threats in surveillance systems is very effective. Such approaches can detect various forms of security threats in complicated environments. Video preprocessing, motion detection and noise reduction techniques, and accurate region segmentation methods like GrabCut serve to give credible information on activities related to threats. Consequently, trust in smart surveillance-based solutions has continued to grow. Convolutional Neural Networks (CNNs) are used, and the architectures similar to YOLOv8 automatically learn features, and different threats are detected in real-time applications. The experimental evidence proves that deep learning models trained on large-scale datasets like COCO and custom surveillance datasets are much more effective than the old methods of machine learning. The system attains high Precision, Recall, and F1-score rates that allow identifying threats at an early stage and with high accuracy, which is the most important to be able to respond and prevent these threats in time.

The future effort will be directed at managing the imbalance of asses and working with more complex and more diverse threat scenarios as well as providing the deep learning models with the ability to be interpreted to achieve better situational awareness. However, despite the fact that the technologies that are in use still have some limitations, continuous developments have been continuously minimizing these problems.

Overall, deep learning is an effective and valid technological basis to identify threats. It enables the minimization of human monitoring errors, improving the efficiency of the response, and timely preventive measures, which is

why it is an effective remedy to the surveillance systems and the safety of the population in the modern context.

References

[1] Waqas Sultani, Chen Chen, Mubarak Shah, "Real-world Anomaly Detection in Surveillance Videos in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6479-6488, (2018).https://arxiv.org/abs/1801.04264

[2] Nayef Saleh, Mohamed Abdel-Maguid, Hisham Sliman, "A real time crime scene intelligent video surveillance systems in the violence detection framework with deep learning methodologies in crime scene detection frameworks using deep learning techniques," Computers & Electrical Engineering, Vol. 103, p. 108319,(2022).https://www.sciencedirect.com/science/article/pii/S0045790622004586

[3] Mohammad Mahbubur Rahman Khan Mamun, Mohammad Ameer Ali, "Design of a real-time crime monitoring system with the use of deep learning techniques or methods", Data Science and Management, Vol. 7, Issue 4, pp. 262-271, (2023).

https://www.researchgate.net/publication/376319944_Design_of_a_Real-Time_Crime_Monitoring_System_Using_Deep_Learning_Techniques

[4] Sunkara, S. P. (2025). Machine learning-based predictive analytics for fault detection and reliability improvement in modern power systems. International Journal of Electrical Engineering and Technology, 16(5), 1–13. https://doi.org/10.34218/IJEET_16_05_001

[5] Georgios Th. Papadopoulos et al., "An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey", Applied Sciences, Vol. 13, No. 8, p. 4956, (2023). https://www.mdpi.com/2076-3417/13/8/4956

[6] S. Gajula, "Two Pillars of Banking Intelligence: A Comparative Analysis of AI Techniques for Fraud Prevention and Churn Mitigation," 2026 14th International Symposium on Digital Forensics and Security (ISDFS), Boston, MA, USA, 2026, pp. 1-6, doi: 10.1109/ISDFS69419.2026.11458995.

[7] C.S. Sung and J.Y. Park, "Design of an intelligent video surveillance system to prevent crime: the application of deep learning technology Multimedia Tools and Applications, Vol. 80, pp. 34297,34309, (2021). https://link.springer.com/article/10.1007/s11042-021-10809-z

[8] Dilshan Nandasena and M.W.P. Maduranga, Deep learning based anomaly detection in a real-time video Multimedia Tools and Applications, (2024).

https://link.springer.com/article/10.1007/s11042-024-19116-9

[9] S. Karpagavalli, A. Kalaivani, Real-time Detection of crime and violence in video surveillance with use of deep learning International Journal of Innovative Technology and Exploring Engineering, Vol. 12, Issue 9, pp. 1-8, (2023).

[10] Guillermo A. Martínez-Mascorro et al., Suspicious Behavior Detection on Shoplifting Cases to Prevent Crime using 3D Convolutional Neural Networks arXiv preprint arXiv:2005.02142, (2020). https://arxiv.org/abs/2005.02142

[11] Seongho Kim et al., "GTA-Crime: A Synthetic Dataset and Generation Framework for Fatal Violence Detection with Adversarial Snippet-Level Domain Adaptation", arXiv preprint arXiv:2509.08232, (2025). https://arxiv.org/abs/2509.08232