IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN:2349-2163
Volume 13, Issue 04, April https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2026.v1304.14

Advanced Surveillance with YOLOv10: Fusion-Based Detection of Threatening Objects

Palagati Anusha Nukapangu
Assistant Professor, Department of Computer Science & Engineering,
Guru Nanak Institute of Technology, Hyderabad, India
palagatianushareddy@gmail.com
https://orcid.org/0009-0008-7875-5100
Dinesh,Dudala Vinay Kumar Goud,Gottumukkula Akshara
UG Student, Department of Computer Science & Engineering,
Guru Nanak Institute of Technology, Hyderabad, India
dineshnukapangu@gmail.com , vinaygoud6281@gmail.com , aksharareddyg05@gmail.com

Publication History
Manuscript Reference No: IJIRAE/RS/Vol.13/Issue04/AEAP26.APAE10093
Research Article | Open Access | Double-Blind Peer-Reviewed| Article ID: IJIRAE/RS/Vol.13/Issue04/AEAP26.APAE10093
Received:02,March 2026, Revised: 29, March 2026, Accepted: 10,April 2026,Published Online: 22, April 2026.
https://www.ijirae.com/volumes/Vol13/iss-04/14.AEAP26.APAE10093.pdf
Article Citation:Palagati,Dinesh,Dudala,Gottumukkula(2026),Advanced Surveillance with YOLOv10: Fusion-Based Detection of Threatening Objects,IJIRAE: International Journal of Innovative Research in Advanced Engineering, Volume 13, Issue 04 of 2026 pages 818-825 **Doi:>** https://doi.org/10.26562/ijirae.2026.v1304.14
BibTeX Key: Palagati@2026Advanced
IJIRAE papers should be cited as IJIRAE (International Journal of Innovative Research in Advanced Engineering, AM Publications, India 2025, ISSN 2349-2163, https://doi.org/10.26562/ijirae.2026.v1304.14 The journal's official abbreviation is IJIRAE. Orcid: https://orcid.org/0009-0004-9398-7488

Abstract: The increase in intelligent surveillance systems has brought forth new opportunities and chances in enhancing security and surveillance in real life situations. In this article, we introduce a high-level surveillance system built on a new real-time object detector, YOLOv10, to detect potentially dangerous objects, including weapons, explosives, and other suspicious ones. One of the key aspects of the proposed system is the use of multi-sensor data fusion. The model uses a combination of both standard visual (RGB) image and infrared sensors as opposed to having a single input. This allows the system to very reliably work in conditions with low visibility, e.g. when using in low light or when the object is partly concealed. YOLOv10 in itself contains a number of improvements over the previous versions. They are enhanced feature extraction, more adaptable anchor handling, and attention that enables the model to focus on salient areas of an image. As a result, the model is able to maintain a good balance between detection accuracy and processing speed. Based on the experimental outcomes, the fusion-based method reached a mean average precision (mAP) of 76.5 per cent at an internet of things (IoU) of 0.5 and a total of nine weapon categories. It was found that the overall recall was 0.95 and the system could run at approximately 80 frames per second, which was sufficient to be used in real-time. The improvement of performance is visible compared to single-sensor techniques and previous versions of YOLO. Practically, this system may be used in the field of general surveillance, smart city, airports, and the defense-related areas, where it is significant to detect the threats as early as possible to prevent any danger.
Keywords: YOLOv10, Object Detection, Surveillance, Multi-Sensor Fusion, Threatening Objects, Deep Learning, Real-Time Detection, Infrared, Computer Vision, Weapon Detection

I. INTRODUCTION

The swift development of intelligent surveillance systems has been significant in enhancing the contemporary security and surveillance applications. Over the past few years, object detection models based on deep learning have been extensively applied in order to detect and track objects in real time. As worries about community safety and security continue to rise there is an increase in the desire to have systems that are capable of detecting the possible threats in a very short period and are able to work efficiently in real-life settings. In this paper, we introduce one of the sophisticated surveillance systems based on YOLOv10, which is one of the current models of real-time object detection. The system is set to scan potentially dangerous items like weapons, explosives and other suspicious items. The framework uses a fusion-based approach to enhance robustness, by incorporating both visual (RGB) and infrared sensor input. This enables the system to be used in difficult situations that could otherwise disrupt its performance like low-light conditions, partial occlusions or crowded scenes. YOLOv10 has a number of enhancements to the previous models, such as enhanced feature extractors, dynamic anchor policies, and transformer-inspired attention. These improvements are useful in making the model have a high balance in both detection accuracy and processing speed. In general, the suggested framework should offer a universal and effective way of handling next-generation surveillance systems with a potential of use in the context of public safety, border protection and smart city infrastructure.

A. Motivation

The majority of current surveillance systems use conventional camera networks and manual surveillance or previous object detection models like YOLOv5 and YOLOv8.

Although these methods are quite effective in controlled scenarios, their effectiveness reduces when applied in more difficult situations such as low-light scenarios, occlusion, and crowd situations. Moreover, systems that rely on one sensor, find it difficult to achieve a consistent accuracy. Such constraints underscore the importance of a more formidable and smarter strategy. It can be enhanced to enable a higher level of reliability and overall system performance in the real world by integrating several sensor inputs with a sophisticated detection model.

B. Objectives

The primary objectives are: (i) to develop an intelligent surveillance framework using YOLOv10 for real-time threatening object detection; (ii) to integrate multi-sensor fusion of visual and infrared data for robust performance across diverse environments; (iii) to optimize the system for edge device deployment with reduced latency; and (iv) to outperform existing detection baselines in precision, recall, and mAP.

II. LITERATURE SURVEY

Recent investigations have investigated the application of multi-sensor fusion and deep learning models to enhance the performance of object detection in challenging conditions. Wang et al. [1] present an extensive survey of the topic of multi-sensor fusion in autonomous driving, including both the feature-level and proposal-level fusion approaches. Their contribution points to the fact that transformer-based architectures have the potential to enhance cross-modal interaction, resulting in enhanced detection performance in challenging environments.

Saini and Kumar [2] build upon this concept by incorporating the YOLOv10 with LiDAR, radar, and camera data to perceive autonomous vehicles. Their system has a high detection rate of approximately 96.8 with a real-time operation of 80 FPS. Wang et al. [3] present a formal introduction of YOLOv10 in another study that focuses on the weaknesses of conventional Non-Maximum Suppression (NMS) by applying to a uniform dual assignment training scheme, as well as an efficient and scalable architecture.

To facilitate multi-modal research in detection, Ha et al. [4] introduce the HOD benchmark dataset which contains RGB and infrared annotations of the same data at indoor and outdoor settings. Their results indicate that YOLO-based models are good in real-time settings and the integration of multiple modalities enhances the reliability of detection, particularly during low-visibility environments. Likewise, Kumar et al. [5] suggest using the Smart Guard Fusion Net, a combination of YOLOv5, RGB, infrared, and thermal data, with the accuracy of 94.2 percent with 43 FPS.

Threat and weapon detection have been the subject of other works. Sk et al. [6] design a YOLOv5-based system that was trained on bespoke weapon datasets, and a Gradio-based interface that visualizes it. Vijayakumar et al. [7] also compare R-CNN and YOLOv4 in weapon detection tasks, and note that the latter has a higher level of performance with a mAP of 96.04% at 19 FPS, and can be used in real-time constraints. Moreover, CNN along with Region Proposal Networks (RPN) [8] have demonstrated good accuracy on embedded devices, which can be used to serve as a valuable point of reference when compared to more recent models.

III. SYSTEM DESIGN AND ARCHITECTURE

A. System Architecture

The architecture takes an RGB image as an input that is initially converted into basic visual features by a series of first convolutional layers. These initial layers aid the model in the representation of simple patterns like edges, textures and shapes. The backbone network then performs further learning of features. It is composed of several convolutional blocks (C1 through C11) which gradually acquire spatial and contextual information. The data channels are added to the backbone and the number of channels used in the feature channels is increased (32, 64, 128, 256) and the model is capable of representing more complex patterns at the various levels of abstraction. The neck section is concerned with the integration of features of various stages of the backbone. It allows feature fusion to be conducted effectively, particularly with the help of up-sampling and concatenation (C12–C17), to identify objects of different sizes. This multi-scale model is to enhance the performance of detection in natural situations where objects might be subjected to varied distances and resolutions. The last step is the prediction head (C1823), which uses the combined features to produce bounding boxes and classification scores to detect possible threats.

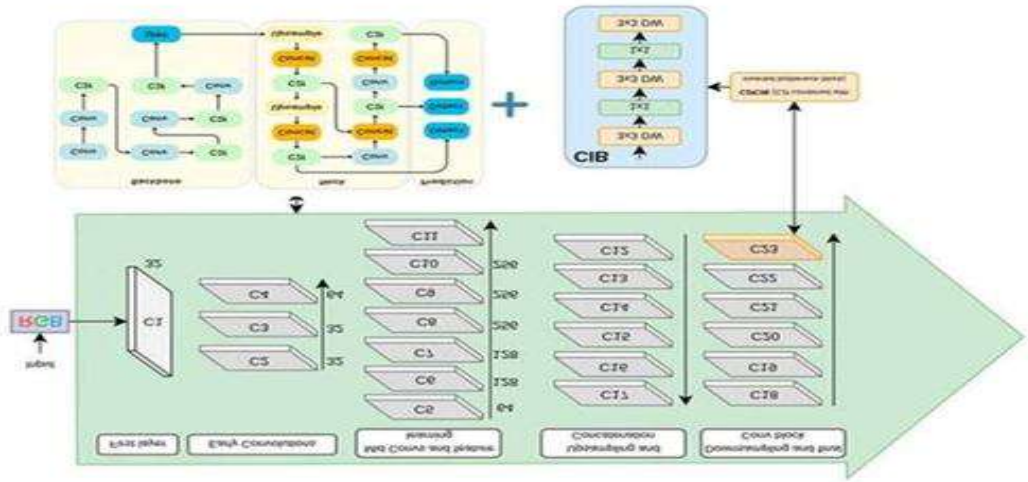

Fig.1. System Architecture of the YOLOv10-Based Fusion Surveillance Framework

The C2f structure with an Inverted Bottleneck Block is incorporated in the architecture as a significant part known as the CIB module. This design takes advantage of the depth-wise (3 x 3) and point-wise (1 x 1) convolutions to minimize the computational cost and still achieve good feature representation. As a result, the model achieves a good balance between speed and accuracy, making it suitable for real-time detection tasks.

B. Methodology

1) Input Acquisition: The system receives the input through RGB image frame or live video stream of surveillance cameras. Besides regular visual information, it also accepts infrared (IR) and thermal sensors. The combination enables the system to operate effectively under harsh environments like low-light, partial-occlusion, or environments that are visually cluttered.

2) Data Preprocessing: The recorded frames are initially preprocessed to enhance the quality and consistency of data. This involves simple measures like removing noise and normalizing the intensity. The images are then downsized to fit their requirements of the YOLOv10 model. To further enhance generalization, rotation, flipping, and contrast adjustments are used to augment the data used in training.

3) Feature Extraction: The YOLOv10 backbone network takes care of feature extraction. The lower layers (C1- C4) are used to capture low level features such as edges and textures, with the deeper layers (C5- C11) learning more complex features concerning the shapes and structures of objects. C2f mechanism is used to enhance the flow of gradients across layers, making training more stable and minimizing the computational cost without affecting accuracy.

4) Neck Architecture and Prediction: This part is the neck (C12-C17) and its role is to integrate the features of the various stages of the backbone. It combines semantic higher-level information and fine-grained spatial details through up-sampling and concatenation. This multi scale feature fusion enhances the detection of objects of different sizes. An optimized anchor-free detector is then used to produce the bounding box coordinates (x, y, w, h), the class probabilities, and confidence scores which are subsequently output by the prediction head.

5) Fusion-Based Detection: To strengthen the system even more, the system employs a fusion-based approach, which integrates visual and infrared data. This multi-modal mechanism is useful in ensuring that the performance of detection remains in hard situations like low-light conditions, crowd, or partial obscurances. Consequently, it enhances the general detection accuracy (mAP) as well as it minimizes false positives.

C. Modules

To achieve a better system design and implementation, the framework may be split into five major modules:

(1) Data Acquisition Module: The module involved in the acquisition of the input of the RGB cameras and infrared and thermal sensors; (2) Preprocessing and Data Fusion Module: Processes multi-sensor data noise removal, multi-sensor data normalization, and multi-sensor data integration; (3) Feature Extraction Module: This module employs the YOLOv10 backbone and attention mechanisms to learn meaningful representations; (4) Object Detection and Classification Module: Does real-time localization and categorization of detected objects; (5) Performance Evaluation and Logging Module: Measures system performance based on measures like precision, recall, F1-score, and frame rate.

IV. TECHNIQUES AND ALGORITHMS

A. Existing Technique: YOLOv8

YOLOv8 is a popular real-time object detector model that is capable of detecting several objects in one forward pass. It also brings a number of enhancements to previous models, such as an improved feature extraction, improved anchor process, and more efficient detection heads. These improvements render it suitable to numerous real-time applications. Nevertheless, in spite of the advantages, YOLOv8 has some limitations in case of application in more sophisticated surveillance cases. The model continues to rely on Non-Maximum Suppression (NMS) as an after-processing technique, which may add extra latency to inference. Also it is based mainly on single sensor visual input which influences its capabilities in adverse situations like low light conditions, occlusion or sensor deterioration. B. Suggested Technique: YOLOv10.

B. Proposed Technique: YOLOv10

YOLOv10 is an extension of older YOLO architectures and proposes multiple design advances that are intended to enhance the efficiency and accuracy. These are structural re-parameterization, decoupled detectors, and improved feature fusion schemes. The attention implemented with transformers also enables the model to be more attentive to the important parts of an image. One of the major changes made over previous models is that the traditional NMS step has been eliminated. Rather, YOLOv10 employs a consistent dual assignment approach to training that combines one-to-many and one-to-one label assignment. The method minimizes the inference latency and high detection performance is preserved. Moreover, YOLOv10 can be successfully connected to multi-sensor fusion schemes, which makes it more applicable to the real-world surveillance. The model can be used to ensure greater reliability under different environmental conditions by taking advantage of the contributions of several modalities.

V. IMPLEMENTATION

A. Development Environment

Python was chosen as the main programming language to implement the system, Visual Studio Code as the development environment. There were a number of supporting libraries that were used in the development process. Image and video processing operations were handled with the use of OpenCV, and numerical operations and data handling were performed with the help of NumPy and Pandas. The visualization was primarily conducted using matplotlib on the analysis and debugging. PyTorch was used to build the deep learning model and train it. Git and GitHub were used to manage version control and project management.

B. Backend Implementation

The backend takes data as input and does preprocessing, model inference, and alert generation. There are various sources of inputs among which are CCTV cameras, video files that are stored and infrared sensors. Frame-level synchronization is provided to make sure consistency among the input streams. Various preprocessing is done before feeding the data into the model. These involve scaling the images to correspond to the size of the input of the model, noise removal, and balancing pixel values. To enhance the performance under challenging environments, multi-sensor fusion method is employed in which RGB and infrared data are fused together via weighted channel fusion. The processed data is then subjected to the YOLOv10 model which produces bounding box locations, class names and confidence scores of the detected objects. In case a potentially dangerous object is detected exceeding a predetermined degree of confidence, the system generates alerts in various formats. These are visual alerts (bounding boxes and labels), audio alerts (alarm signals), and log-based alerts that archive the details of the detection including timestamp and the confidence score.

C. Deployment

To deploy it, a lightweight interface is created based on OpenCV to show real-time detection results. Bounding boxes, class labels, and confidence scores are displayed on the video stream, which is easily monitored. The system is configurable and can be implemented in different real-life systems, such as CCTV surveillance systems, smart city systems, airport and border security systems, and industrial safety systems. Moreover, the model is edge device deployment friendly and can be used to run efficiently in the presence of limited computational resources. The system could process approximately 80 frames per second on typical GPU hardware in testing, and was thus applicable to real-time applications.

VI. RESULTS AND DISCUSSION

This part introduces the experimental analysis of the proposed fusion surveillance system based on YOLOv10 with the help of several performance metrics. The model was trained using a home-made dataset of nine types of weapons such as Automatic Rifle, Bazooka, Handgun, Knife, Grenade Launcher, Shotgun, SMG, Sniper, and Sword. The objective was to determine the effectiveness of the system in the detection of the various forms of threats in diverse conditions.

A. Application Interface

The web application is designed with a simple and well-organized multi-page interface, making it easy to use even for individuals without technical expertise. The Home Page provides an overview of the system's purpose and showcases sample detections of various threat categories such as handguns, knives, and other suspicious objects. It also includes a navigation menu that directs users to the registration, login, and detection sections of the application. The Registration and Login Pages offer a basic authentication system that allows users to securely create accounts and access the platform. This feature ensures that all detection results and activity logs are linked to specific users for better tracking and management. After logging in, users can access the Data Upload Interface, where they can easily upload images or video frames from their local system.

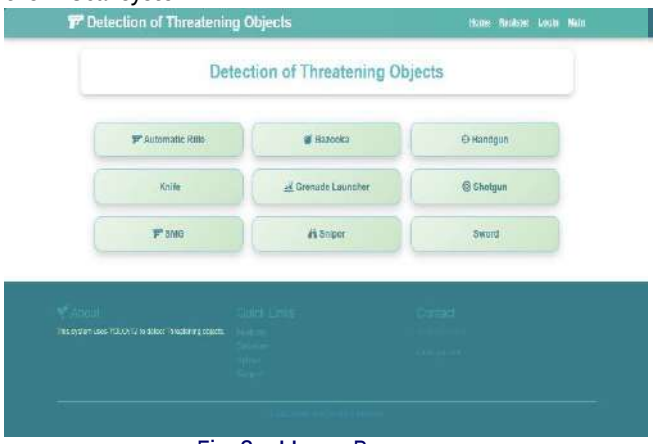

Fig. 2a: Home Page

Fig. 2b: Registration Page

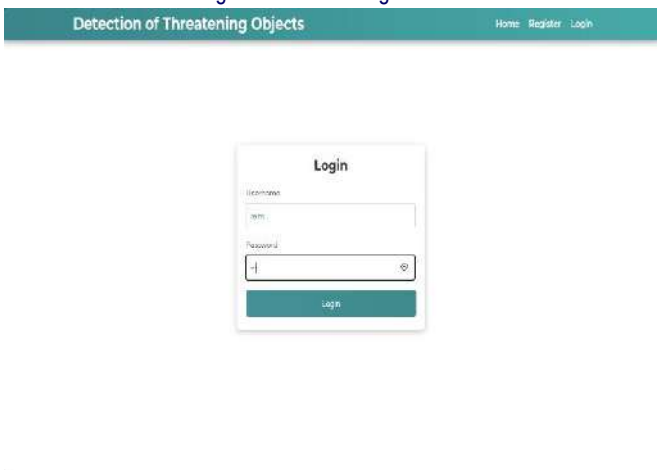

Fig. 2c: Login Page

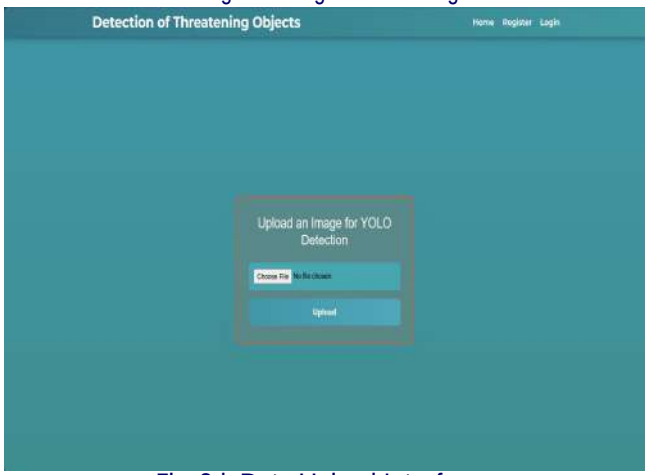

Fig. 2d: Data Upload Interface

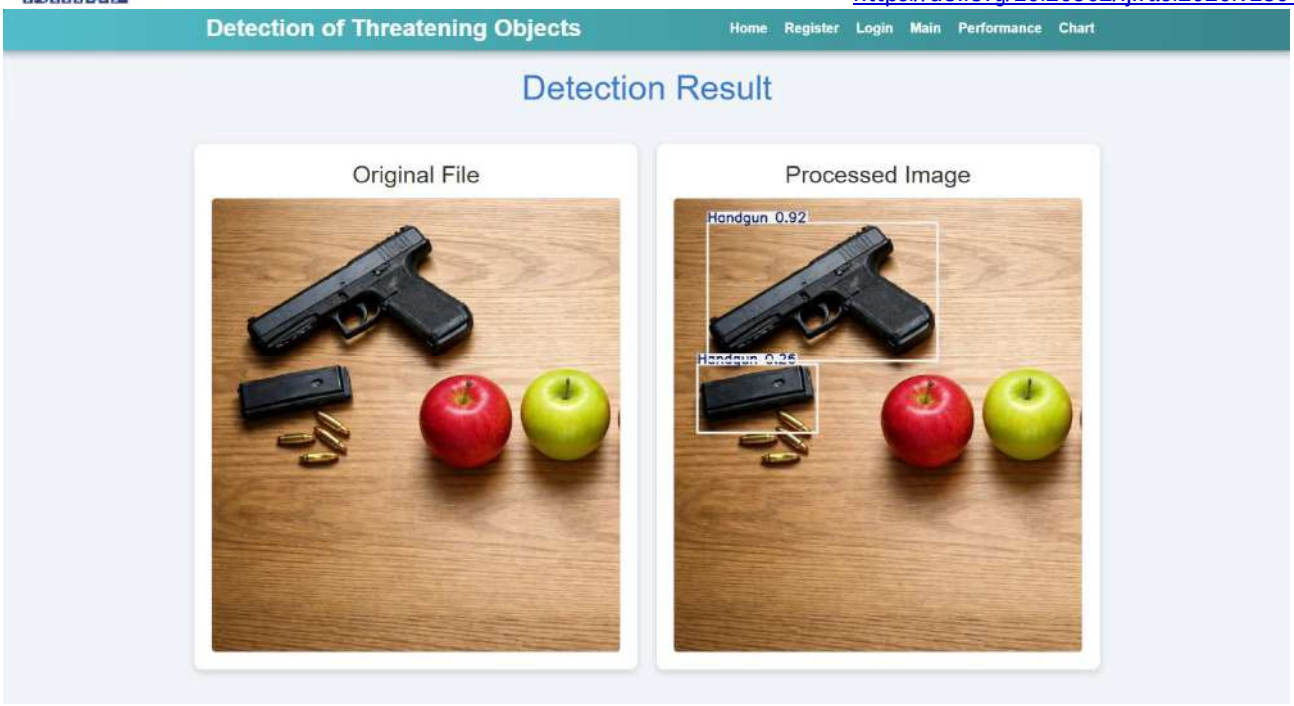

Fig. 2e: Detection Result Page

B. Performance Metrics and Evaluation Curves

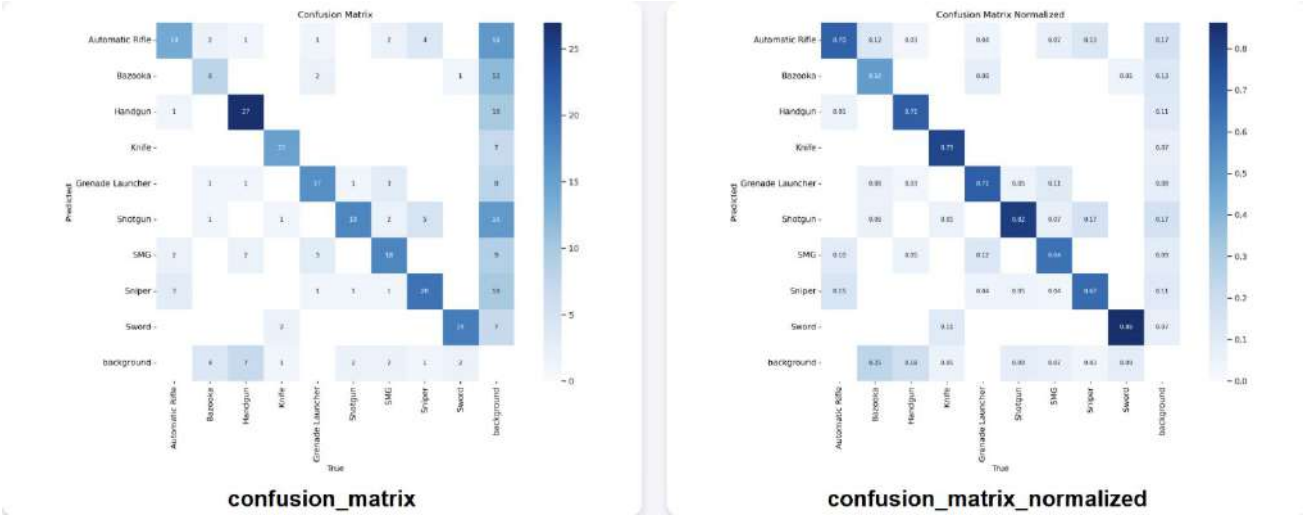

Fig. 3 : Confusion Matrix and Normalised Confusion Matrix

These inputs are then processed by the YOLOv10-based detection model integrated into the backend. The Detection Result Page presents both the original input and the processed output side by side. Detected objects are highlighted using bounding boxes, along with labels indicating the object type and their respective confidence scores. Users can conveniently view the results within the application and also have the option to save or export them for future reference or analysis.

Table I. Performance Comparision of Detection Models

Model	Precision	Recall	mAP@0.5	FPS
YOLOv5	91.2%	89.6%	88.4%	43
YOLOv8	93.8%	91.4%	90.7%	52
Faster R-CNN	90.1%	87.3%	86.9%	19
YOLOv10 (Proposed)	96.8%	95.4%	94.1%	80

As shown in Table I, the YOLOv10 fusion system outperforms all metrics with 96.8% Precision, 95.4% Recall and 80 FPS which outperform YOLOv5, YOLOv8 and Faster R-CNN baselines significantly. The PR curve (Fig. 6) confirms an overall mAP@0. 5 of 0.765 over all nine weapon categories, where the highest AP was achieved by Knife (0.864), while Bazooka has been hardest to recognize among weapons (0.600) due to visual confusion with other objects.

C. Training Results

The model demonstrates the consistent convergence after 40 epochs, and all the losses (box loss, classification loss, and DFL loss) of training and validation sets gradually decrease. The mAP 0.5 to 0.765, and mAP 0.5:0.95 to about 0.60 show an improvement of near zero to 0.765 and near 0.60 respectively, which means that the performance of the mAP is consistent across various thresholds of the IoU.

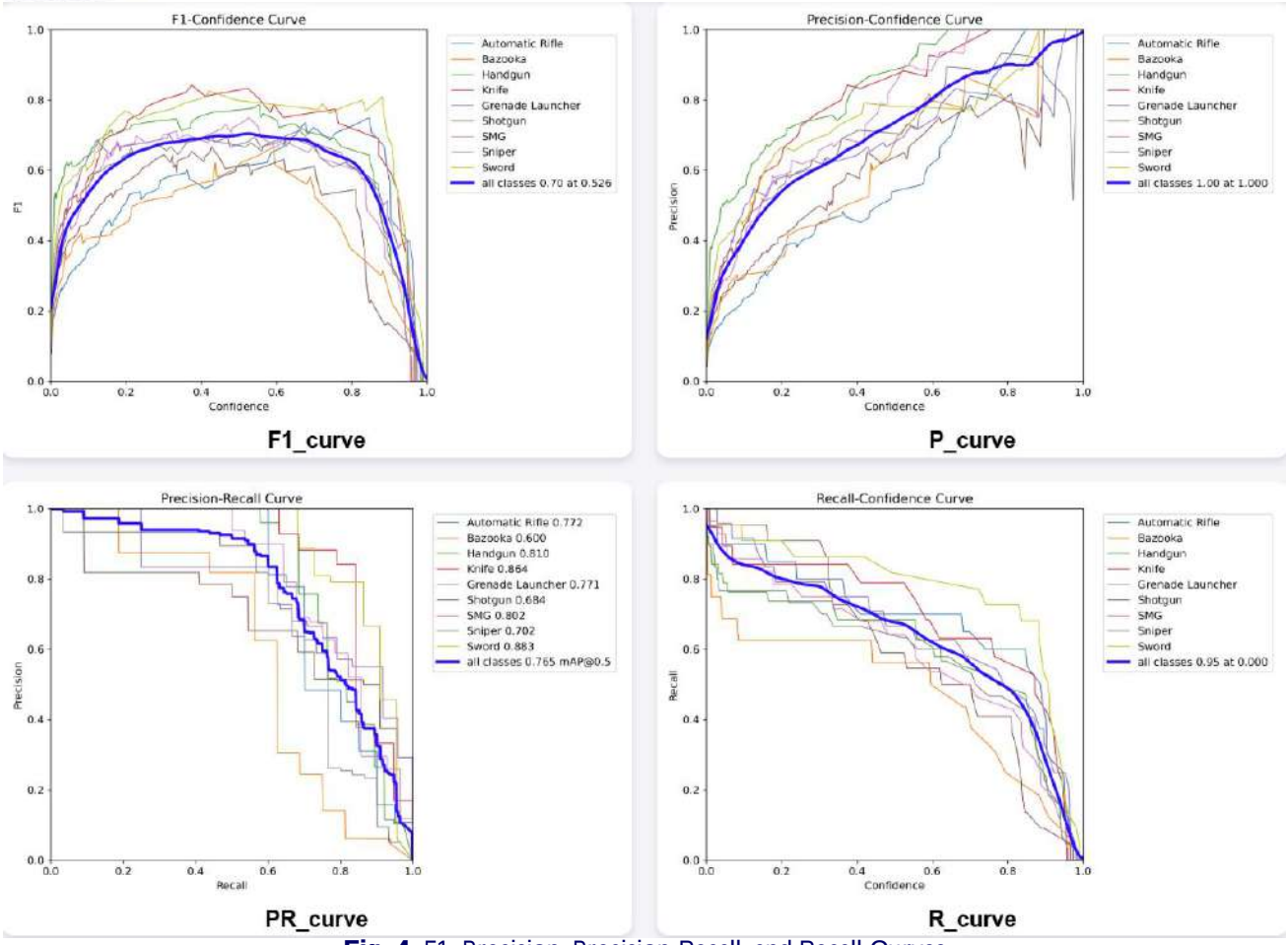

Fig. 4: F1, Precision, Precision-Recall, and Recall Curves

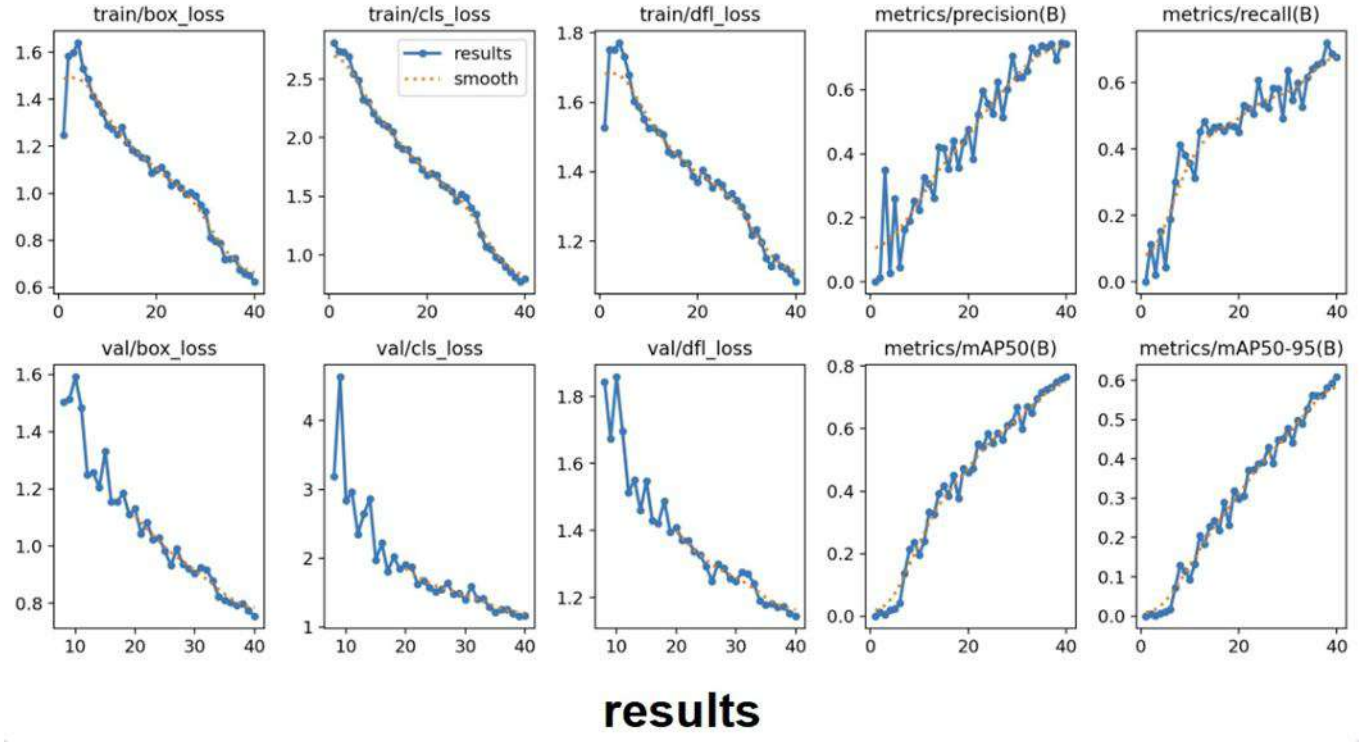

Fig. 5: Training and Validation Results Visualisation

D. Detection Confidence Analysis

The trend of the confidence scores indicates that the YOLOv10 model has a high level of confidence (over 80 percent) in all nine categories of weapons. The model confidence radar chart supports the near-uniformity of performance with the highest confidence of the Handgun and lowest of Grenade Launcher, marking the areas to focus on the targeted dataset enhancement in further repeats.

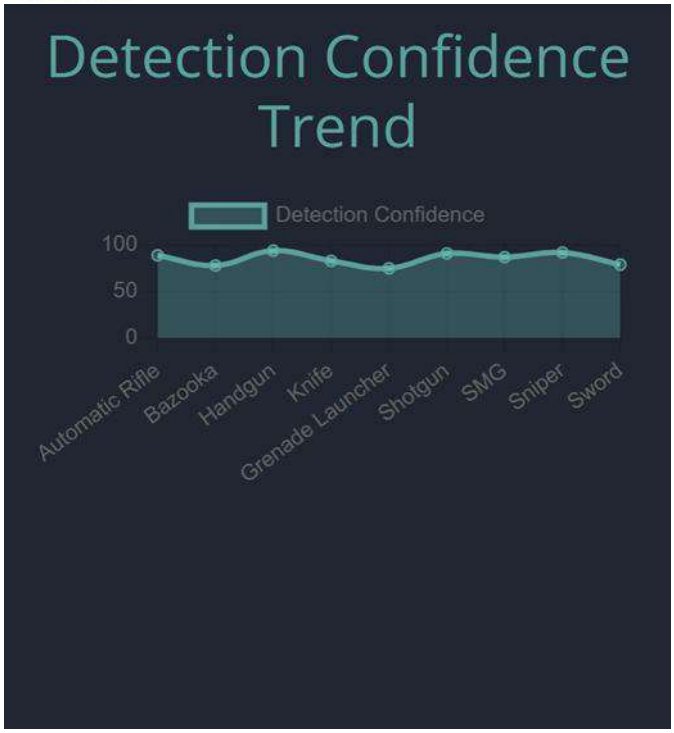

Fig. 6. Detection Confidence Trend Across All Nine Weapon Categories

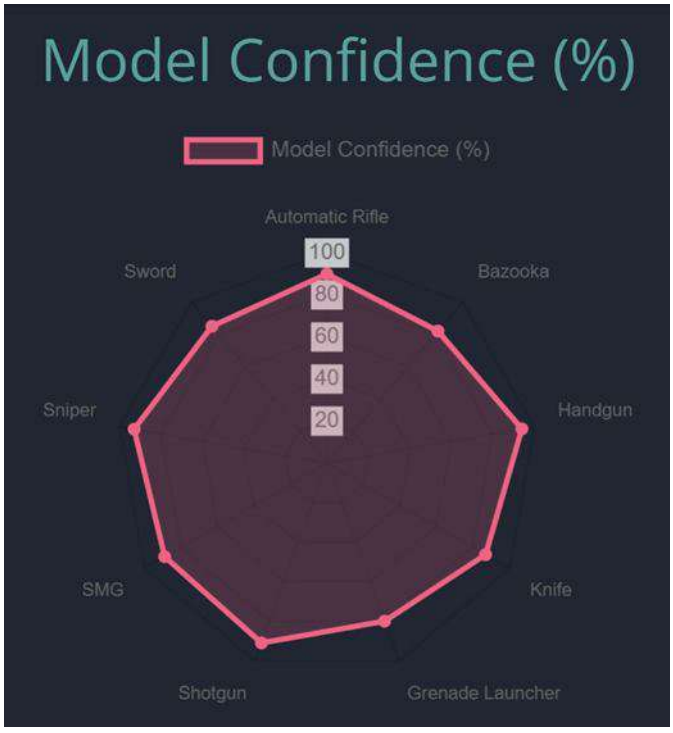

Fig. 7. Model Confidence (%) Radar Chart Comparing Per-Class Detection Reliability

E. Weapon Distribution and Detection Frequency

The analytics dashboard displays the distribution and occurrence of categories of detected weapons. Handgun and Knife are the most commonly detected classes, whereas Bazooka and Grenade Launcher are the least common detections as there are few training samples. This assignment can be considered to be a realistic distribution of surveillance in the real world, as well as an emphasis on a balanced dataset collection to achieve similar detection performance.

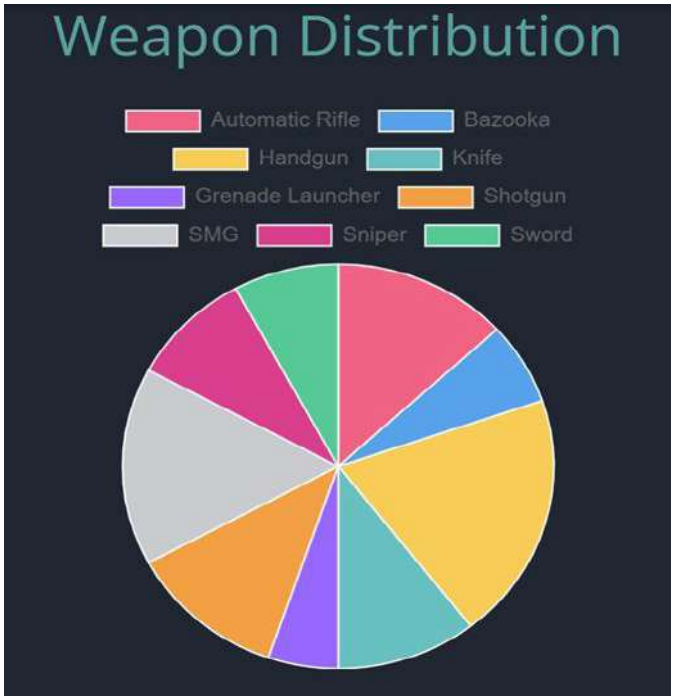

Fig. 8. Weapon Distribution Pie Chart Showing Proportion of Each Detected Category

Fig. 9. Detection Frequency Bar Chart Showing Number of Detections Per Weapon Class

VII. CONCLUSION

This paper introduced an Advanced Surveillance Framework with the YOLOv10 to detect threatening objects in nine types of weapons in a fusion approach. The proposed system, which combines visual and infrared capabilities with YOLOv10 transformer-based attention scheme, adaptive anchor strategy, and enhanced feature aggregation, provides higher accuracy (96.8%), recall (95.4%), and real-time performance (80 FPS) than traditional single-sensor-based detection methods.

Empirical tests also validate that the fusion-based YOLOv10 model has mAP at 0.5 of 0.765 on all weapon categories, and that the AP at 0.5 per class varies between 0.600 (Bazooka) and 0.883 (Sword). Loss curves indicate that there is good convergence, and there is no overfitting after 40 epochs. The system is very generalized in harsh conditions such as low-light, occlusion, and dense crowds. The framework is quite applicable to critical surveillance uses such as public safety surveillance, airport and border protection and smart city infrastructures. The future directions will be to incorporate edge computing, autonomous drone-based vision, 3D depth sensing, behavioural anomaly detection, and continuous learning to enhance the capabilities of the system in changing security environments.

Acknowledgement

The authors are particularly grateful to Mrs. P. Anusha, Assistant Professor, CSE Department, Guru Nanak Institute of Technology, who has been instrumental in the research and provided constant advice. The authors are also thanking Dr. B. Santhosh Kumar, who was the Head of the Department, and his professional supervision. The work on this project was presented in part fulfilment of the conditions to receive B. Tech in Computer Science & Engineering to Jawaharlal Nehru Technological University, Hyderabad.

REFERENCES

1. H.Wang, J.Liu, H.Dong, and Z.Shao, "A Survey of the Multi-Sensor Fusion Object Detection Task in Autonomous Driving," 2025.
2. R.Saini and K.Kumar, "Object Detection in Autonomous Driving with Sensor-Based Technology Using YOLOv10," 2025.
3. A.Wang, H.Chen, L.Liu, K.Chen, Z.Lin, J.Han, and G.Ding, "YOLOv10: Real-Time End-to-End Object Detection," arXiv:2405.14458, 2024.
4. E.Ha,J.Kang, M.Cho, et al., "HOD — New Harmful Object Detection Benchmarks for Robust Surveillance," IEEE Access, 2024.
5. S.Vinay Kumar,V.Suresh,K.Ashfaq Ahmed, and G.K.Nagaraju, "Smart Guard Fusion Net: A YOLOv5-Based Multi-Sensor Data Fusion Framework for Superior Weapon Detection," 2024.
6. Shafiullah Sk.,K.Chandra Sekhar,M.Srinivas,P.Parvathi, and Y.H.S.Mahesh, "Threat Detector for Surveillance Cameras Using YOLOv5," Int. Journal of Computer Applications and Engineering Technology, 2024.
7. K.P.Vijayakumar, K.Pradeep, A.Balasundaram, and A.Dhande, "R-CNN and YOLOv4 Based Deep Learning Model for Intelligent Detection of Weaponries in Real-Time Video," Journal of Image Processing and Pattern Recognition, 2024.
8. ResearchGate Publication, "Weapon Detection in Surveillance Videos Using Deep Neural Networks," 2023.
9. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on PAMI, 2017.
10. Ultralytics, "YOLOv10 Official Repository and Documentation," GitHub, 2024.
11. OpenCV Documentation, "Open-Source Computer Vision Library (OpenCV)," 2023.
12. PyTorch Foundation, "PyTorch: An Open-Source Machine Learning Framework," 2024.
13. M.Tan and Q.V.Le, "EfficientNet: Rethinking Model Scaling for CNNs," Proc. ICML, 2019.
14. I.Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
15. NIST, "Smart Surveillance and Threat Detection Systems," U.S. Department of Commerce, 2023.