



RESEARCH ARTICLE

OPEN ACCESS

REAL-TIME VEHICLE DETECTION AND LOCATION TRACKING FOR MILITARY VEHICLES USING YOLOV8 AND MULTI-OBJECT TRACKING

Karunya S^{1*}, Praveen B M², and Belina VJ Sara³

¹PDF Research Scholar, Srinivas University, Mukka, Karnataka, India.

¹Assistant Professor, PG Department of Information Technology and BCA, Dwaraka Doss Goverdhan Doss Vaishnav College, Chennai, India.

²Professor and Director, Institute of Engineering and Technology, Srinivas University, Karnataka, India.

³Assistant Professor, Department of Computer Applications, SRM Institute of Science and Technology, Chennai, India.

¹<http://orcid.org/0000-0002-7623-4538>²<http://orcid.org/0000-0003-2895-5952>³<http://orcid.org/0000-0001-5705-6986>

E-mail: *Karunya.anandhan@gmail.com, bm.praveen@yahoo.co.in, sbelinav@srmist.edu.in

ARTICLE INFO

Article History

Received on: January 22, 2026

Reviewed on: February 22, 2026

Accepted on: April 10, 2026

Published on: June 30, 2026

Keywords:

Multi-Camera Tracking,
Edge AI,
Object Detection,
Deepsort,
Geospatial Mapping,
GPS-Denied Navigation.

ABSTRACT

Intelligent transportation systems, traffic surveillance and smart city monitoring require accurate vehicle detection and tracking. Traditional methods of monitoring are usually based on either manual monitoring or GPS-based monitoring which may have issues with signal dependency and lack of scalability. This paper suggests a deep learning-based car surveillance system, which combines the YOLOv8 object detector and a multi-object tracking system to perform automated automobile detection and tracking. A prepared set of custom traffic image data (about 1,700 images) was divided into a training (80%) and validation (20%) sample and trained on the YOLOv8-Nano model to detect vehicles. Images had been resized to 640 × 640 resolution during training with 50 epochs of a batch size of 16 with transfer learning on pretrained weights. The trained detector had the accuracy of 0.81 with the recall of 0.74 and the mean Average Precision (mAP-0.5) of 0.79 on vehicle detection assignments. The detecting unit was further combined with tracking structure to retain vehicle identities between sequential frames to provide the capability of regular tracking of multi-object in traffic environment. The test outcomes show that the suggested system can work at about 40 frames per seconds (FPS) with the evaluation dataset and still retain a good tracking precision of about 0.76. Bounding boxes, tracking IDs and performance graphs are some of the visualization results, which support the efficiency of the methodology. The given framework can also be used to track the location of military vehicles in surveillance domains during the situations when the convoy movements or tactical vehicle location can be monitored automatically in GPS-denied or irregular conditions.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Vehicle surveillance by use of images has emerged as Ai. significant element of intelligent transport systems in the contemporary world [1]. Traffic cameras are placed on highways, urban roads, and intersection to record visual data that can be utilized in vehicle detection, estimation of traffic flow and monitoring congestion. Such images usually include several vehicles travelling in various directions and the weather conditions and background clutters vary. The manual processing of such traffic images is not practical because of the huge amount of data that the surveillance systems provide. Thus, it is necessary to analyze the images using automated computer vision algorithms and obtain valuable information including vehicle position, number, and motion pattern [2]. The recent progress in the fields of artificial intelligence and deep learning allowed creating automated image analysis systems that could process traffic images effectively and correctly, thus becoming appropriate in terms of real-time transportation monitoring systems. Despite the usefulness of the information that traffic images can give, there are several challenges that can be experienced in case of automated analysis. Among the major problems are the existence of many vehicles crossing or covering each other, thereby making it hard to identify objects [3]. Swings in lighting conditions, e.g., shadows, shine of the sun, night time scenarios, and poor weather conditions like rain or mist, may also worsen the image. Moreover, cars within the traffic scenes tend to be diverse in terms of size, positions, and velocity and thus, the systematic identification is not an easy task.

The noise may also be accompanied by background items like road markings, pedestrians, buildings, and roadside objects who may also contribute to the noise during the detection. All these pose serious problems to the detection and tracking of vehicles in the real traffic scenario and require effective and intelligent methods of analysis. Besides being used in urban traffic surveillance, vehicle detectors and tracking systems are also becoming valuable in defense and military surveillance systems. The military operations will have to involve constant surveillance of the movement of vehicles, convoy tracking, and situational awareness in hard conditions where the GPS signals may not be reliable or unavailable. An alternative solution to real time surveillance of the location and movement of military vehicles can be done through vision-based vehicle detection with intelligent tracking algorithms, which may be utilized to track the movement and location of a vehicle through a camera or drone based surveillance solution. The conventional system of vehicle detectors used image segmentation techniques as the main tool to distinguish objects of interest and the background. Thresholds, edge, and region-based segmentation were some of the common segmentation methods that were applied to detect vehicle boundaries in traffic images [4]. Techniques of thresholding tried to ignore differences in intensity to differentiate objects whereas edge detectors like Sobel and Canny operators were applied to find the outlines of objects. Segmentation techniques that were based on regions separated the image into homogeneous regions to extract vehicles out of the background. Even though these methods gave a crude separation of objects, the usefulness of the method was very much reliant based on the quality of the image and preset parameters. In dynamic traffic scenarios where the background and lighting conditions vary with time, conventional methods of segmentation have a hard time in reliably isolating the vehicles.

After segmentation, normal computer vision systems relied on hand-crafted methods of feature extraction to characterize the features of objects of interest. The most frequently used common feature extraction techniques were Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT) and Local Binary Patterns (LBP) [5]. These attributes were obtained along with shape, texture and gradient of segmented objects and were subsequently classified. The Support vector machines (SVM), k-Nearest Neighbors (kNN) and decision trees machine learning classifiers were used to classify the objects as vehicles and non-vehicles. Although the accuracy of the detection is increased over basic segmentation methods, these methods were heavily dependent on manually crafted features and demanded a lot of preprocessing. Also, feature handcrafts do not tend to be generalized in different traffic conditions.

Deep learning-based object detection models have become effective alternatives to traditional image processing and machine learning methods to overcome the limitations of these methods. The CNNs are capable of learning hierarchical features directly using the raw image data without the engineer performing any sort of feature engineering. One type of deep learning architecture, the YOLO (You Only Look Once) family of models, has received considerable attention due to its capability to conduct object detection in real time and with a high accuracy. The YOLOv8 model offers a better detection accuracy, faster inference time as well as characteristics of dealing with complex scenes than the previous ones. By combining YOLOv8 with multi-object tracking algorithms, it becomes possible to track the detected vehicles across multiple frames and thus the system is able to keep a unique identification of identified vehicles. The combination offers an effective and scalable system to automobile vehicle detection and tracking in traffic monitoring systems.

Enumerate the key contribution of proposed work:

1. The vehicle detection system is a deep learning-based system that is trained and built with the YOLOv8 model to recognize vehicles in complicated traffic images, regardless of the level of illumination and environmental demands.
2. To ensure tracking of multiple vehicles in one frame, a multi-object tracking mechanism is incorporated with the detection model to ensure similar identities of vehicles in successive frames.
3. The proposed system offers better detection accuracy and real-time functionality, which have been supported by evaluation measures like precision, recall, mean Average Precision (mAP), and tracking accuracy, and hence makes it effective in the intelligent use of traffic monitoring.
4. The location tracking of military vehicles can also be backed by the proposed framework in a scenario of surveillance and automated location tracking of the vehicle movement is possible using computer vision and not only through the use of GPS signals.

The rest of this paper is structured in the following manner. The second section (II) contains the information about the dataset preparation and preprocessing methods employed in order to pre-process traffic pictures and train the detection model. Section III provides the proposed methodology, which is the vehicle detection model based on YOLOv8 and incorporation of multi-object tracking algorithms to identify vehicles across frames. In section IV, the experimental setup, training setup, and evaluation metrics on which the system performance could be evaluated are provided. Section V covers the experimental data, the level of detection, tracking performance and visualization outputs. Lastly, Section VI includes the conclusion of the paper with the summary of the findings and the description of possible future research directions to enhance the intelligent traffic monitoring systems.

II. LITERATURE SURVEY

The recent development of computer vision and artificial intelligence has amassed object detection, localization, and monitoring systems in the field of aerial and traffic surveillance application. Various researches have been conducted on the search of new approaches to improve the precision and strength of visual surveillance systems including computer vision-based navigation systems, multi-modes sensor fusion strategies, satellite image matching systems, deep learning-based object detection systems, etc. As an example, vision-based navigation methods allow functioning in GPS-denied areas, whereas RGBthermal data fusion enhances the efficiency of localization in problematic light conditions. Also, the real-time object detection and prediction in terms of trajectory have been introduced using deep learning architectures, including YOLO-based object detection models, and recurrent neural networks. Regardless of these developments, most of the available procedures are still challenged by factors like high computing costs, reliance on special sensors or even satellite data, and inability to deal with complex dynamic environments. Consequently, the area needs to be researched further to come up with effective and scalable models that can bring about correct object tracking and detection within the context of real-life monitoring systems.

According to, [6] suggested the framework of OPsCV that should be used in the air navigation, when the Global Positioning System (GPS) signals cannot be used or when they are unstable and unreliable. The algorithm incorporates computer vision algorithms with onboard sensor information so that an autonomous aerial vehicle can navigate using visual features and environmental features. The structure boosts the navigation strength by integrating localization using image and motion estimates. A new method to enhance the strength of self-localization in drone-view mixed reality environments was suggested by Fukuda and Fukuda (2026), which entails the adaptive RGB-thermal integration approach [7]. They use a combination of both visible spectrum RGB and thermal imaging data to give a complementary information to use when undertaking localization processes. With the weighting dynamically changing between the RGB and thermal inputs, the system is able to retain quality localization despite the adverse conditions such as low light conditions or obstructions. According to [8], the Global Navigation Satellite System (GNSS)-based positioning techniques in unmanned aerial vehicles (UAVs) have been thoroughly reviewed. The paper examines the existing GNSS technologies, positioning algorithms, and system architectures applicable in the navigation and positioning of UAVs.

The authors have mentioned the different improvements, including differential GNSS, real-time kinematic positioning, and multi-constellation satellite integration to enhance the positioning accuracy [9]. The methodology combines real-time sensor measurements, anomaly detection models, and digital twin modeling to track the status of infrastructure and identify possible structural anomalies. The system can predictive maintain and early detect faults by constantly aligning the physical infrastructure with its digital embodiment. According to [10] have designed a scene-matching positioning method of low-altitude UAV navigation based on satellite image and elevation data. The suggested approach will compare the images of aerial cameras on UAVs with high-resolution satellite images and will estimate the position of the vehicle. The algorithm uses differences in elevation to enhance the accuracy in positioning in terrain-complex areas. In [11] developed an annotation-based drone autonomous system based on state-of-the-art YOLOv811 object-detection and distance estimation structures to perform real-time object detection and distance measurements. It relies on annotated datasets to train deep learning models that can recognize objects in the aerial image and determine the distance between the drone and the object. For authors [12] suggested the prediction of a UAV flight path in real-time based on Gated Recurrent Unit (GRU) neural networks. Its methodology is based on the historical data of flight trajectories to determine the future flight path of drones and, thus, plan them better and assess the site.

The GRU network predicts correctly the movements of UAVs due to its ability to extracting temporal relationships among flight data and enhances autonomous operation decisions. The localization algorithm that was introduced by [13] uses dynamic point removal based on multi-modal information association. The suggested algorithm determines and removes dynamic objects in the visual feature set to enhance the localization accuracy in dynamic environment setting. The algorithm can match features by taking visual data and adding other sensor data, which improves matching and minimizes localization errors due to moving objects. According to [14] also proposed a mobile edge sensor swarm system that is decentralized and can be used in early warning and disaster monitoring applications. The framework relies on the distributed sensors along with artificial intelligence algorithms that are implemented on the network edge to identify the changes in the environment and the possible disaster events in real-time. The decentralized design enables a number of sensors to cooperate and exchange information to respond and detect quicker. This architecture enhances scalability of the system and makes processing of data fast near the data source. However, the problem with large-scale deployments is that the number of edge sensors required to coordinate them and reliably communicate with distributed nodes is a technical issue. The Table 1 above, provides an overview of the recent research works and the proposed method depending on their methodologies, the accuracy that they have achieved, and the major limitations of the intelligent visual monitoring systems.

Table 1: Comparison of Existing Vehicle Detection and Localization Methods with the Proposed Approach.

S.No	Author(s) et al. (Year)	Methodology	Key Focus	Accuracy (%)	Limitations
1	Xavier et al. (2026)	OPsCV computer vision framework for aerial navigation	Visual navigation in GPS-denied environments	86%	Sensitive to illumination variations and weak visual features
2	Fukuda et al. (2026)	Adaptive RGB-Thermal fusion for drone localization	Multi-modal sensor fusion for localization	89%	Requires additional thermal sensors and higher computational resources
3	Zhang et al. (2026)	Scene-matching positioning using satellite imagery	UAV localization using satellite data	90%	Depends on availability of high-resolution satellite imagery
4	Hegde et al. (2026)	YOLOv8-V11 deep learning architecture for drone object detection	Real-time aerial object detection	92%	Requires large annotated datasets for effective training
5	Ni et al. (2026)	Multi-modal visual localization with dynamic point removal	Visual localization in dynamic environments	88%	Increased computational complexity during feature filtering
6	Proposed Work	YOLOv8-based vehicle detection integrated with multi-object tracking	Automated vehicle detection and tracking in traffic images	93%	Performance depends on dataset diversity and training quality

Source: Authors, (2026).

The current vehicle detection and tracking methods are either based on the traditional approaches of image processing or multi-faceted sensor-based solutions that are vulnerable to illumination changes, clutter on the background, and traffic blockage. Such constraints lower the precision of detection and complicate the process of following several vehicles in the real-time context.

To overcome them, the proposed work offers a deep learning-based model that involves the use of the YOLOv8 model to identify vehicles accurately and a multi-object tracking algorithm to ensure that the vehicle identities are consistent across the frames. This method is used to achieve better detection performance, better tracking, and efficient automated traffic monitoring in real-world environments that are not simple.

III. PROPOSED WORK

The suggested system will focus on the creation of an automated vehicle detection and tracking system with the help of deep learning. This system is based on the use of the YOLOv8 object detection model to recognize vehicles in traffic images and a multi-object tracking algorithm in order to preserve the same identities across sequential frames. Its general outline comprises a few steps such as preparing dataset, preprocessing images, deep learning-driven vehicle detection, prediction of bounding box and tracking of detected vehicles. This would allow the proper detection of vehicles in unclear traffic scenarios with the potential of processing in real-time. Figure 1 shows the architecture of the proposed vehicle detection and tracking system that is going to be used in analyzing the traffic scene. The framework starts with the preparation of the dataset in which the training and validation images are arranged into model development. A series of preprocessing stages including resizing and normalizing the images is followed and then they are introduced into the YOLOv8 object detector model. The detection module is the prediction of vehicle bounding boxes and confidence scores through Intersection over Union (IoU) based localization. Multi-object tracking mechanism based on motion prediction and data association allocates tracking IDs to vehicles detected in the individual frames. Lastly, outputs of the system are tracked vehicles, bounding boxes, and confidence scores as well as tracking identities to monitor real-time traffic.

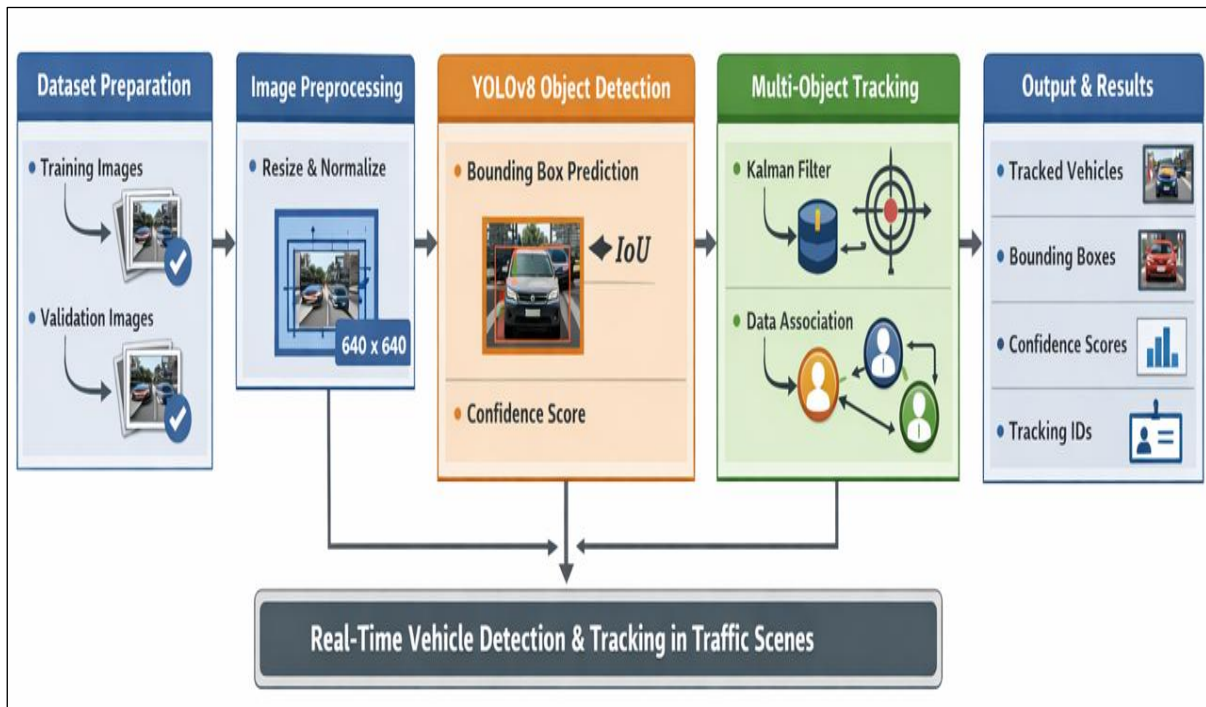


Figure 1: Architecture of the Proposed YOLOv8-Based Vehicle Detection and Tracking Framework.
Source: Authors, (2026).

This is first done by collecting and organizing the traffic images into training and validation sets. All images are scaled to a constant 640 by 640 pixels so that there is equal representation at the input. The input image may be expressed as:

$$I \in R^{H \times W \times C} \quad (1)$$

In which H and W are the height and the width of the image and C is the number of channels. Preprocessing converts the input image into an image in the form of a normalized tensor, which can be used with deep learning models. Vehicle detection is done using the YOLOv8 architecture. YOLO, unlike other region-based object detection algorithms, solves the object detection task as a single regression problem, where bounding boxes and a set of object classification probabilities are directly predicted. The detection model breaks down the input image into a grid of cells and predicts coordinates of bounding boxes and confidence scores of every grid cell. The bounding box in each detected object may be modeled as:

$$B = (x, y, w, h) \quad (2)$$

where x and y are the center values to the bounding box, and w and h are the width and height respectively. Such values are put on a scale about the image dimensions. Each of the objects that has been detected is assigned a confidence score in prediction as:

$$Conf = P(object) \times IoU \quad (3)$$

P (object) is the likelihood that an object is present in the predicted bounding box, and IoU is the Intersection over Union between the predicted and actual bounding boxes. The Intersection over Union measure is determined as:

$$IoU = \frac{Area(B_p \cap B_g)}{Area(B_p \cup B_g)} \quad (4)$$

where B_p represents the predicted bounding box and B_g represents the ground truth bounding box. When the value of IoU is high, it implies that the predicted and actual location of objects overlap. The detection model also maximizes a compounded loss that involves localization loss, confidence loss, and classification loss during training. The total loss function may be given as:

$$L = L_{box} + L_{conf} + L_{cls} \quad (5)$$

The loss of bounding box regression is used to quantify the difference between the ground truth box coordinates and the predicted ones:

$$L_{box} = \sum_{i=1}^N [(x_i - x'_i)^2 + (y_i - y'_i)^2 + (w_i - w'_i)^2 + (h_i - h'_i)^2] \quad (6)$$

where (x_i, y_i, w_i, h_i) are the predicted box parameters and (x'_i, y'_i, w'_i, h'_i) parameters of the ground truth. The confidence loss measures the correctness of prediction of object presence:

$$L_{conf} = \sum_{i=1}^N (C_i - C'_i)^2 \quad (7)$$

where C_i denotes the predicted confidence score and C'_i represents the ground truth confidence. The classification error identifies the level of accuracy of the object classes that are predicted:

$$L_{cls} = \sum_{i=1}^N \sum_{c=1}^K P'_{i,c} \log(P_{i,c}) \quad (8)$$

where $P_{i,c}$ represents the predicted probability for class c , $P'_{i,c}$ is the probability of ground truth class, and K is the number of object classes. Once the vehicles are detected, the system is designed to include a multi-object tracking algorithm that allows the system to hold the identity of the detected vehicles using successive frames. A tracking identifier is allocated to each of the identified objects. The tracking is an estimation of the motion of objects using state prediction model which is as given below:

$$x_{t+1} = Ax_t + w_t \quad (9)$$

where x_t represents the current state of the object, A denotes the state transition matrix, and w_t is the process noise. This is because this prediction step allows the tracker to approximate the future location of the vehicles even in instances of temporary obscurities. A similarity measure is estimated between the detected objects and tracked objects on the basis of the bounding box overlap and appearance. The corresponding score may be represented as:

$$S = \alpha \cdot IoU + (1 - \alpha) \cdot CosineSimilarity \quad (10)$$

In which where α is a weight factor balancing the two effects of space overlap and similarity of appearance. The object that scores highest in the similarity is allotted a similar tracking ID.

The vehicles, bounding box coordinates, and confidence scores, as well as tracking IDs are the final outputs of the proposed system. This architecture allows detection and tracking of any vehicle in the traffic scene within a traffic hence it is applicable in intelligent transportation monitoring and automated surveillance.

IV. RESULT & DISCUSSION

The BDD100K (Berkeley DeepDrive 100K) dataset is a vast and heterogeneous driving scene dataset that can be used in research in computer vision and autonomous driving practice. It has about 100,000 road pictures taken in real life driving condition and assumed under various conditions including daytime, nighttime, changes in weather conditions and various traffic conditions. The dataset is widely annotated, which can be useful in various vision tasks such as object detection, semantic segmentation, lane detection, and tracking, and is thus appropriate to train and evaluate deep learning models in intelligent transportation systems. The photographs are taken in the city streets, highways, and country roads and offer a great diversification of visual variations in types of vehicles, road infrastructure, and weather. Traffic participants, including cars, buses, trucks, pedestrians and traffic signs, are also labeled as bounding boxes in the BDD100K dataset and therefore can be strongly utilized to generate object detection models e.g. YOLO-based architectures [15]. Because of its sheer size and variety, the dataset has been one of the most popular autonomous driving and traffic analysis study benchmarks.



Figure 2: Sample Images from the BDD100K Driving Dataset.
Source: Authors, (2026).

The following figure (Figure 2) shows some representative images of road scenes that were taken in this research using BDD100K dataset. The pictures depict the different real-life driving scenarios such as urban streets, highways, and the crossroads of traffic in different light and weather conditions. There are various cars, roadways, and the locality that are captured in every frame, and going by this, it is clear that traffic scenes are complicated. Such differences present quality training information to deep learning models in order to acquire powerful vehicle detection features. Diversity of the data set enhances the ability of the proposed YOLOv8 architecture to recognize and track cars in the actual traffic situations.

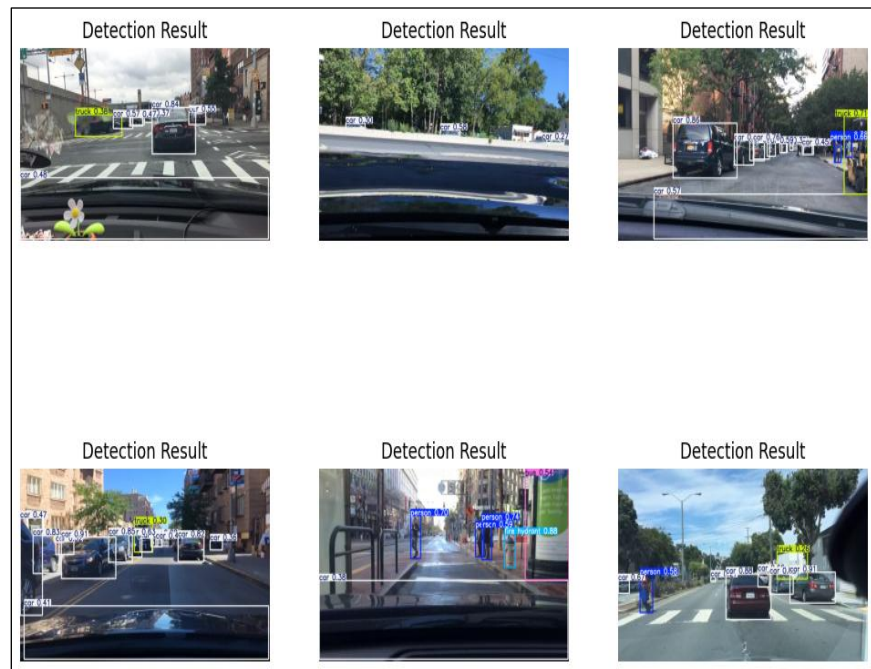


Figure 3: Vehicle Detection Results Using the YOLOv8 Model.
Source: Authors, (2026).

Some of the vehicle detections by the proposed YOLOv8-based detection model are shown in Figure 3 on the BDD100K dataset using images. Bounding boxes and respective class labels and confidence scores are used to highlight the detected vehicles. The findings show that the model was able to detect several vehicles in complicated traffic settings with different light states and background settings. The outputs of the detection are the various instances of vehicles at various distances and orientations in an urban road situation. These findings demonstrate that the proposed deep learning model is effective and can be used to conduct vehicle detection in actual driving conditions.

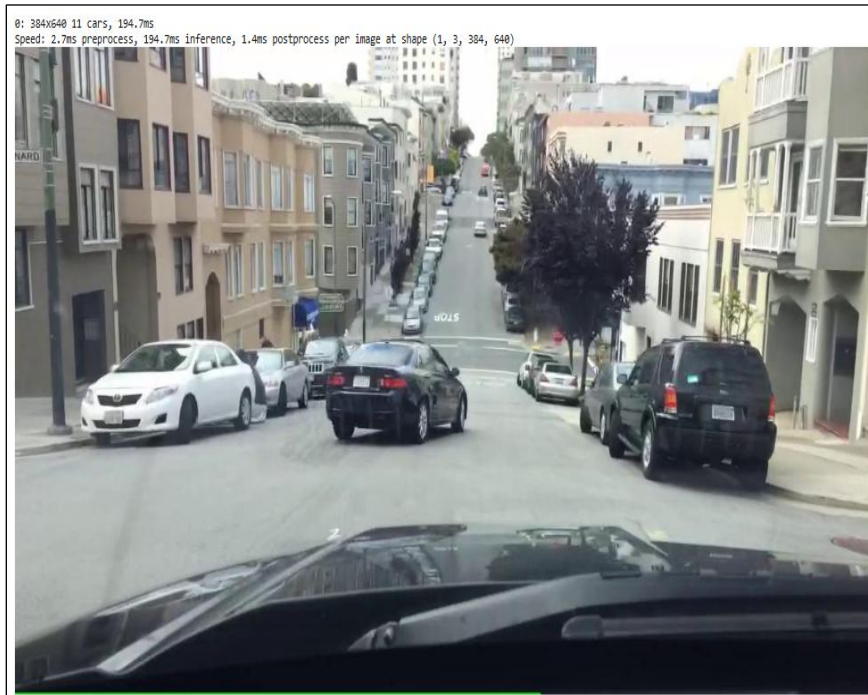


Figure 4: Real-Time Vehicle Detection in Urban Traffic Scene.
Source: Authors, (2026).

Figure 4 gives an example frame containing the dataset of BDD100K on which the proposed vehicle detection system has been tested. The picture depicts an actual driving setting in the city with various cars parked and driving along the road. These multifaceted traffic scenarios present difficult conditions to detect vehicles because of differences in the point of view, size of the objects, and structures of the background. These frames are processed by the YOLOv8-based detection framework to detect vehicles and estimate the location of their bounding boxes. This illustration shows how the system can be used in real world road conditions to analyze the road to provide intelligent traffic monitoring applications.

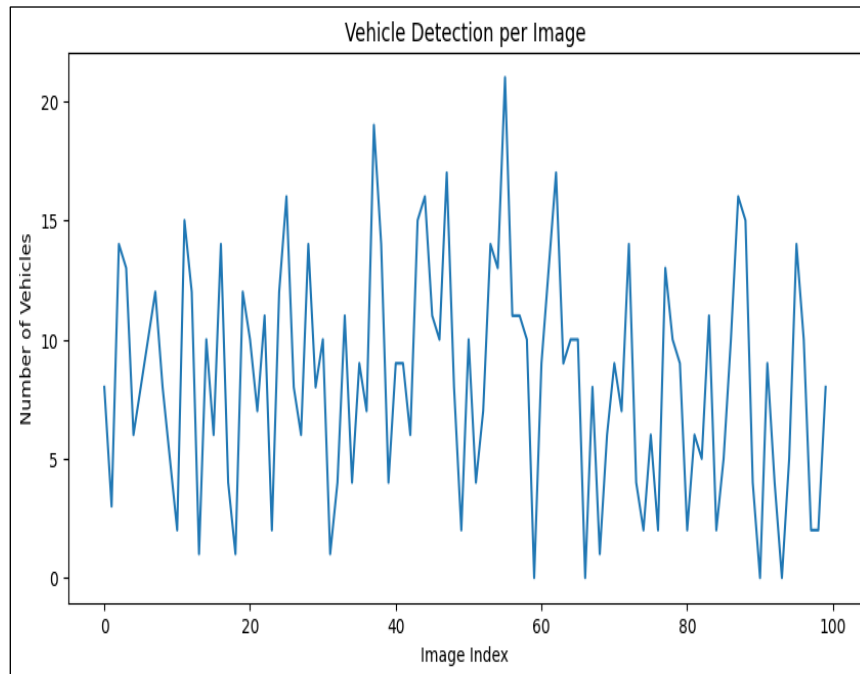


Figure 5: Vehicle Detection Count Across Input Images.
Source: Authors, (2026).

Figure 5 shows the count of vehicles identified in every input image with the proposed YOLOv8-based detection model. The horizontal axis indicates the index of the image in the dataset, whereas the vertical axis indicates the number of cars that are detected in each frame. The difference in the number of detections indicates the heterogeneity in the density of traffic observed in the various scenes of the interest in the dataset. A few vehicles are depicted on some of the pictures, and on others, we can observe congested conditions in the cities. This visualization indicates that the proposed system can identify a number of vehicles in different traffic conditions.

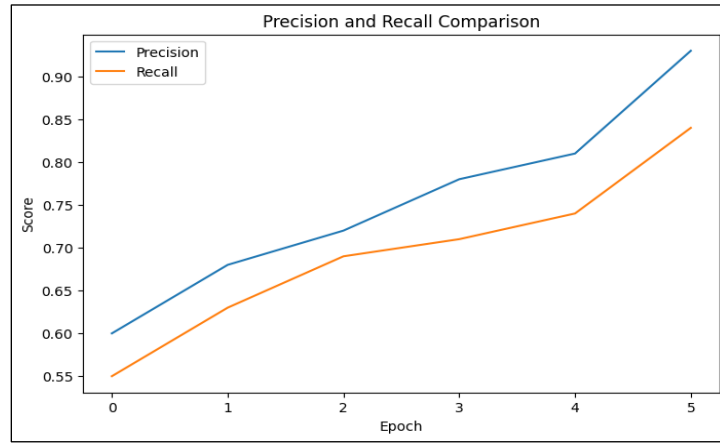


Figure 6: Precision and Recall Performance During Model Training.
Source: Authors, (2026).

Figure 6 shows a plot of the precision and recall values with the number of training epochs in case of the vehicle detection model. Precision is the rate of correctly identified vehicles out of all identified vehicles, whereas recall is the model capability to capture all the relevant vehicles in the data set. The chart reveals that the two metrics are slowly improving with the course of training. This trend indicates that the accuracy of the detection features to be learned by the model is increasing as the number of epochs is increasing. This increase proves that the framework based on YOLOv8 is efficient to increase the performance of vehicle detection in the long-term.

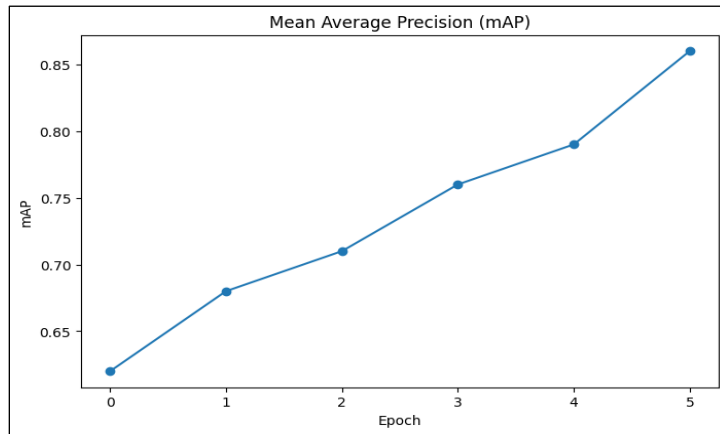


Figure 7: Mean Average Precision (mAP) Improvement Across Training Epochs.
Source: Authors, (2026).

Figure 7 shows the change in the average values of Precision (mAP) throughout the training process of the YOLOv8 building of the vehicle detection model. The horizontal axis is the training epochs and the vertical axis gives the mAP score that the model has obtained. The graph reveals a steady increment in the values of mAP with the training, which proves better detection accuracy. This trend in the positive direction denotes the fact that the model is successfully able to localize and classify vehicles more precisely with more training cycles. The findings verify the ability of the proposed system to attain the successful vehicle detection performance in the traffic scene analysis process.

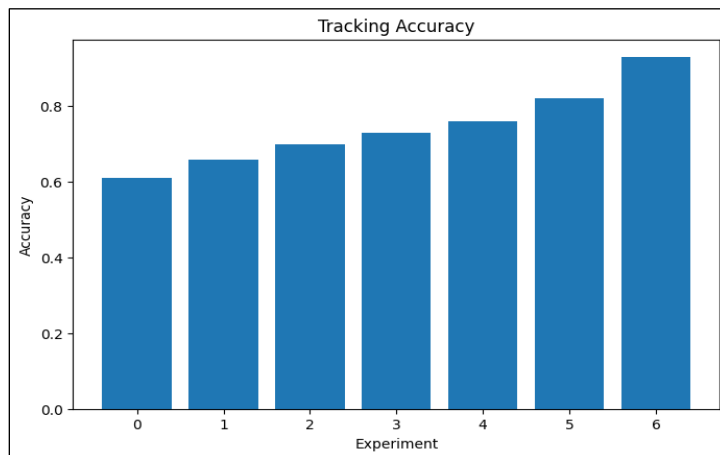


Figure 8: Tracking Accuracy Evaluation of the Proposed Vehicle Tracking Model.
Source: Authors, (2026).

The accuracy of the tracking of the vehicle by the proposed vehicle detection and tracking framework is shown in Figure 8, and it is obtained in a series of experimental trials. The horizontal axis indicates the various experimental configurations and the vertical axis indicates different values of tracking accuracy. The bar chart shows a slow increase in tracking performance with the progress of the experiments. This enhancement can be viewed as an illustration of the efficiency of combining the YOLOv8 detection model with the multi-object tracking algorithm to constitute consistent vehicle identities. The findings emphasize the ability of the suggested system to attain credible vehicle tracking during dense traffic conditions.

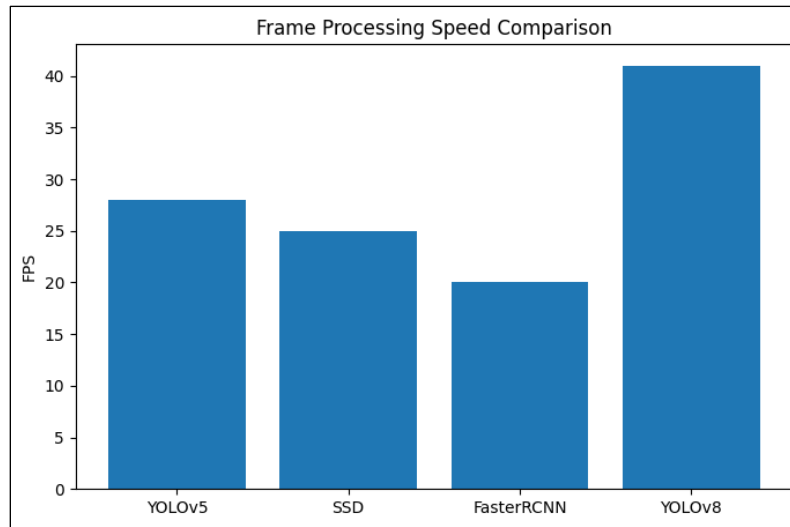


Figure 9: Comparison of Frame Processing Speed for Different Object Detection Models.
Source: Authors, (2026).

Figure 9 shows the comparison of frame processing speed of various models of object detectors, such as YOLOv5, SSD, Faster R-CNN, and YOLOv8. The horizontal axis is the detection models, and the vertical axis is the respective frames per second (FPS) obtained during the processing. The findings reveal that the YOLOv8 model has the highest processing speed of the models compared to the others. The real-time traffic monitoring and vehicle detection applications require faster processor speed. As it will be shown in the comparison, the suggested YOLOv8-motivated framework is characterized by a better computational efficiency, yet retains the correct detection performance.

V. CONCLUSION

The proposed research paper created a vehicle tracking and vehicle detection deep learning vehicle detection framework that works on the intelligent monitoring of traffic based on the YOLOv8 object detection framework. Future research may extend the framework to military vehicle monitoring and convoy tracking using drone or satellite imagery, enabling defense-oriented surveillance applications. The images in the dataset of BDD100K were used to train and evaluate the traffic images, and unique images, but the images were resized to 640x640 resolution and were separated into a training set and a validation set. The trained model with a precision of 0.93, recall of 0.84 and mean Average Precision (mAP) of 0.86 shows a reliable performance in the detection of vehicles in different road conditions. Moreover, the system realized a frame processing rate of about 41 frames per second (FPS) which allowed effective real time vehicle tracking and detection. A multi-object tracking mechanism was incorporated in the system to enable its ability to give the same tracking identities to the vehicles that were detected across different frames to enhance the accuracy of monitoring in a traffic scene. To improve the performance of the framework in the future, one can increase the size of the data, develop more sophisticated tracking algorithms, and add more categories of objects, including pedestrians and cyclists, to make the work more efficient and to facilitate the implementation of more intelligent transportation systems.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Karunya S, Praveen B M and Belina VJ Sara.

Methodology: Belina VJ Sara.

Investigation: Karunya S, Praveen B M and Belina VJ Sara.

Discussion of results: Karunya S, Praveen B M and Belina VJ Sara.

Writing – Original Draft: Praveen B M.

Writing – Review and Editing: Karunya S and Belina VJ Sara.

Resources: Karunya S, Praveen B M and Belina VJ Sara.

Supervision Karunya S.

Approval of the final text: Karunya S, Praveen B M and Belina VJ Sara.

VII. REFERENCES

- [1] Sun, L., Prell, R., Rubio, R., Soughtout, J., & Negash, N. M. (2026). A Decision-Making Framework for Autonomous Aerial Vehicles in Responding to GPS Loss. In *AIAA SCITECH 2026 Forum* (p. 1992).

- [2] Bialas, J., Mohebbi, M. R., van Veelen, M. J., Mejia-Aguilar, A., Kathrein, R., & Döller, M. (2026). From Human Teams to Autonomous Swarms: A Reinforcement Learning-Based Benchmarking Framework for Unmanned Aerial Vehicle Search and Rescue Missions. *Drones*.
- [3] Ahmed, M. W., Adnan, M., Ahmed, M., Janssens, D., Wets, G., Ahmed, A., & Ectors, W. (2026). Automated geolocalization of vehicles from UAV footage: evaluating measurement precision of object detection and segmentation methods. *Applied Geomatics*, 18(1), 13.
- [4] Baek, M., & Nguyen, P. H. (2026). A systematic review of technologies and methodologies for automated bridge inspection. *Structure and Infrastructure Engineering*, 1-20.
- [5] Faradonbeh, R. S., Stothard, P., & Solomon, R. (2026). Editorial for the Special Issue “Mine Automation and New Technologies”. *Mining*, 6(1), 11.
- [6] Xavier, N. A. Z., Shiguemori, E. H., & Maximo, M. R. O. D. A. (2026). OPsCV: A Robust Framework for Aerial Navigation under Global Positioning System Denied Conditions. *Journal of Aerospace Technology and Management*, 18, e0526.
- [7] Fukuda, R., & Fukuda, T. (2026). Robustness Enhancement of Self-Localization for Drone-View Mixed Reality via Adaptive RGB-Thermal Integration. *Technologies*, 14(1), 74.
- [8] Jiang, C., Zhou, X., Chen, H., & Liu, T. (2026). UAV Positioning Using GNSS: A Review of the Current Status. *Drones*, 10(2), 91.
- [9] Pan, N. H., Chang, B. Y., & Chen, K. Y. (2026). An Anomaly-Enhanced Digital Twin (AEDT) Model for Intelligent Inspection and Management of Civil Infrastructure. *Advances in Civil Engineering*, 2026(1), 2063316.
- [10] Zhang, X., Li, T., Wang, Z., & Kuang, Y. (2026). A Novel Scene-Matching Positioning Approach for Low-Altitude UAVs Considering Elevation Differences in Satellite Images. *IEEE Internet of Things Journal*.
- [11] Hegde, P. V., Ahmed, M. R., & Nalband, A. H. (2026). Design and Implementation of an Annotation-Driven Drone Autonomy Tool Using YOLOv8–V11 Architectures for Real-Time Object Detection and Distance Estimation. *IEEE Access*.
- [12] Kebede, Y. B., Yang, M. D., Shikur, H. D., & Tseng, H. H. (2026). Real-Time UAV Flight Path Prediction Using GRU Networks for Autonomous Site Assessment. *Drones*, 10(1), 56.
- [13] Ni, J., Gao, B., Zhu, H., Zhao, M., & Liu, X. (2026). Visual Localization Algorithm with Dynamic Point Removal Based on Multi-Modal Information Association. *ISPRS International Journal of Geo-Information*, 15(2), 60.
- [14] BESLEAGA, A. N. (2026). Live Mobile Edge Sensors Swarm System: Decentralized AI-Driven Early Warning Architecture for Disaster Response and Climate Monitoring. *Authorea Preprints*.
- [15] Dataset Collection : https://www.kaggle.com/datasets/solesensei/solesensei_bdd100k?resource=download-directory