International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2) 2026 pp. 718-734
https://doi.org/10.70917/ijcisim-2026-3011
Article

YOLOv8- Driven Object Detection Framework for Real-Time Vehicles and Pedestrian Monitoring in Adverse Conditions

Ashutosh Pandey, Siddharth Pandey, Akash Sanghi

Department of Computer Applications, Invertis University, India
*Corresponding Author: Ashutosh Pandey, Research Scholar

Abstract: The given research paper singles out one vehicle system that contributes to increasing the safety level by delivering instantaneous object detection and forecasting on the basis of YOLOv8. Unlike standard external surveillance systems, this one is enclosed within the car so as to be able to detect and classify pedestrians, motor vehicles, and road debris. Moreover, the technology can be fine-tuned to harsh weather conditions, including rain, fog and low light so that the technology will be able to operate safely and with a high degree of strength under various driving conditions. Using cutting edge visual recognition and deep learning networks, this technology can evaluate the potential of objects under observation to cause possible collisions and hazards by processing live stream video streams and, in turn, enable the vehicle to come up with safety-enhancing actions. The YOLOv8 model consumed fast and precise analysis with the help of the detected objects and provided significant contextual data of the surrounding of the vehicle. The model was compared to Precision, Mean Average Precision (mAP) and Intersection over Union (IoU) metrics, which proved the superiority of YOLOv8 models over previously developed object detection models regarding its ability to identify objects and make predictions about the danger. The model also undertook predictive risk analysis that was done with 643 high hazard image and 163 low hazard image and this displays the capability of the model to analyze safety situation in instantaneous. Situational awareness, which is a vital element of autonomous driving, will be improved with the facilitation of real-time applications.

Keywords: YOLOv8, Object Detection, Advanced Driver, Assistance Systems, Predictive Risk, Intelligent Vehicle.

1. Introduction

In the modern world where the world is rapidly urbanizing, the presence of motor vehicle and human traffic has exerted a lot of pressure on the available traffic infrastructure. With the population of cities and the consequent decrease in its density, preserving the road safety and efficiency in the traffic flow are becoming more and more complicated issues in the contemporary society [1]. The traffic congestion on major streets and the constant contact of vehicles and pedestrians with each other often poses a great chance to be involved in an accident, and traffic surveillance can be considered a very important tool in ensuring the safety of people [2]. Although innovation in intelligent transportation systems is still ongoing, road-related injuries remain a significantly alarming weight in the world. Crossroads and traversed streets in the city are specifically dangerous locations to be caught in an accident, mostly in wet weather and low visibility [3]. This contributes to the growing necessity of active risk evaluation in road surveillance to assist in the improvement of situational awareness and preventive accidents.

A risk-assessment based on predicted collisions or interaction among detected objects allows the system to proactively respond by sending alerts to either the driver or traffic management systems. Such proactive responses are essential in high-risk environments, such as those with unpredictable weather, low-visibility conditions, and high volumes of traffic; because real-time decision-making may be the difference between an accidents occurring versus one being prevented. Studies conducted in recent years show that when predictive analytics is integrated into an object-detection system, there is a significant improvement in safety performance in highly dynamic environments [13]-[15].

Even with technological advancements, the current surveillance systems, like the CCTV cameras installed at the crossroads and the radar systems placed on highways, are still not without drawbacks. Their inability to react promptly and the necessity of having someone to go through the recorded video are among the major downsides [4]. This dependency very often leads to a slower emergency reaction which translates into more deaths [5]. What is more, the inability of these systems to react immediately to unexpected and sudden events has resulted in the delay of emergency medical help.

Computer-vision algorithms have been applied to help automate vehicular collision detection in order to assist in reducing the number of collisions that occur on roads around the world. In this manuscript, we describe a new method of utilizing a instantaneous object detection system based on YOLOv8, which is a deep-learning based object detection algorithm [6], to identify multiple types of vehicles including cars, trucks, buses and motorcycles, as well as pedestrians within dynamic environments having varying weather and light conditions. Through the use of a large and diverse training set, we were able to train the model to produce high levels of accuracy in detecting all of the aforementioned object classes in realistic scenarios [7].

The major task for this research is to design a real-time framework that is capable of detecting multiple types of vehicle and pedestrian in various environmental conditions. Leveraging YOLOv8's anchor-free architecture and lightweight architecture guarantees high detection speed and logics to deliver high robustness to intelligent transportation [8] application.

The sketch out of the rest of the paper is as follows. Part II is a concise summary of the related works and literature. Part III outlines the approach, which is proposed, of this paper. Part IV describes the experimental design with the set up and configuration of the experiments. Part V explains the research methodology, coming up with data collection, processing and training model. Part VI designates about model performance checks like precision, recall, F1-score, etc. Part VII shows the outcomes of the experiments, in which the model's performance is displayed with the help of various evaluation metrics. Part VIII ends the paper and discusses future areas of work including integrating driver emotion recognition for enhanced safety responses.

2. Related Work

Advancements in object detection have been instrumental in the enhancement of the functionality of Intelligent Transportation Systems (ITS). Traditional object detection methods like Haar Cascade and Histogram of Oriented Grids (HOG) classifiers preceded the application of computer vision but had the drawbacks of high false-positive rates of false-alarms and poor adaptability in complex environments. Traditional methods suffered from low visibility, rain, occlusion, and other non-ideal scenarios that are common in the actual traffic scenarios.

The invention of deep learning has been a paradigm shift in object detection. The methods in the R-CNN family and the series YOLO (You Only Look Once) added the delicacy of real-time detection to their genetic algorithms, with high improving accuracy. YoLoV3 and YOLOv4, for example, improved the detection performance through the introduction of anchor boxes, residual blocks and batch normalization. These models have been shown to have superior accuracy and are therefore suitable for a large number of real-time applications. However, such an anchor representation was computationally expensive and limited the efficiency of use, especially in dynamic regimes where the objects' scale and position frequently change [9], [10].

Being aware of these concerns, YOLOv7 drastically exploits extensive efficient layer aggregation networks (E-ELAN) and allows scientific learning so that specially there remains a synchronous shift in learning although keeping the inference speed consistent. It also enhanced the accuracy on large-scale detection benchmarks. However, challenges such as object occlusion, object scales scale variation, and environmental variability continued to impact performance when real-time application to traffic urban deployment environments were to be exploited [11].

YOLOv8 is the latest version which takes a rather streamlined lightweight approach. Adopting anchor-free detection paradigm, decoupled detection head and efficient C2f backbone, YOLOv8 eases the training complexity and alleviates generalization issues caused by different object scales and aspect ratios. These advances lead to a significant analysis reduction in complexity and make the model more suitable for deployment on edge devices. Furthermore, YOLOv8 shows good robustness in harsh environments such as bad lighting, fog and rains, which are common in traffic surveillance system [8].The assessment metrics, is summarized in Table I.

Table 1: YOLOv3 – v8 COMPARISON

Model	Anchor/Backbone	Features	Edge Suitability
YOLOv 3	Y / Darknet-53	Multi-scale prediction, residual blocks	Moderate
YOLOv 4	Y / CSPDarknet53	PANet, Mish activation, data augmentation	Good
YOLOv 7	Y / E-ELAN	Deeper feature aggregation, high accuracy	Good
YOLOv 8	N / C2f	decoupled head, lightweight design	Excellent

Even though its benchmark performance is enhanced, the exploitation of YOLOv8 for real-world traffic system is still lacking detailed investigation. Factors including heavy occlusion, variable lighting and adverse weather mean that there is still a lot of difficulty in detection accuracy. While previous iterations of YOLO took performance to a new level in their ability to identify objects in real-time, these versions were challenged in a variety of unpredictable situations that require an element of robustness and adaptability, and in which multiple factors are involved. Overcoming such shortcomings, however, is paramount to the successful implementation of detection systems in real-time ITS applications.

Recent researches have tried to investigate several strategies for improving the detection of objects under adverse conditions. YOLOv8 has been combined with data fusion techniques in order to enhance detection reliability under different meteorological situations [13]. A better version of YOLOX was proposed for foggy vehicle detection [14]. Cross-modality distillation methods have been used in scenarios of low-light pedestrian detection [15], but dehazing network (AOD-Net) has been used to improve visibility of objects in fog [16]. Domain adaptation techniques have also been proposed to enhance the generalization to dynamic environments [17]. Furthermore, improvement in multimodal pedestrian detection [18] and exhaustive surveys on low light detection [19] point out the growing research interest in those environmental challenges.

However, more recent trends in ITS have included the combination of deep learning models and edge computing and sensor fusion, in order to perform better in delivering a real-time intelligence in identification and decision-making. Multimodal fusion methods with RGB, infrared and LiDAR data can achieve good robustness under environmental changes such as fog, glare and night [25]. Edge-based vision models have further helped to cut down on latency in large-scale deployments that enable on-device traffic analytics applications of smart intersections and vehicle-mounted systems [26]. In addition, lightweight deep neural networks optimized for embedded processors have shown significant improvements in inference speed without the trade-off with detection accuracy, which is required for autonomous and connected vehicles running in resource-constrained environments [27]. Beyond perception, driver-behavior and attention-prediction systems based on visual and physiological cues have also been investigated in an attempt to prevent accidents before they occur, which is in close agreement with the goal of prediction for safety in this study [28]. Recent research has also reviewed the application of hybrid frameworks where visual perception and risk prediction are used in conjunction for improved decision making in complex traffic environments [29]. These research studies collectively attest to the increase in importance of scalable, weather resilient, predictive approaches in intelligent vehicle system.

These studies emphasize the importance of having a general-affordable and efficient object detector with the capability of overcoming various environmental challenges. Regarding its practicality, although YOLOv8 shows attraction as a way to build on, its applicability in the real world of dynamic and unpredictable traffic needs further research. This is our motivation for conducting this study, which focuses on the evaluation of the performance of the YOLOv8 algorithm under the condition of real world traffic surveillance, including occlusions, lighting and weather-based visual distortions.

3. Proposed Approach

The proposed approach uses YOLOv8 for instantaneous object detection in traffic monitoring, specifically focusing on detecting vehicles and pedestrians under various challenging ecological conditions. YOLOv8 has been selected due to its anchor-free detection paradigm, high efficiency, and superior accuracy in handling diverse and dynamic conditions, including occlusion, poor visibility, and adverse weather. This approach aims to bridge the gap between existing object detection models and the requirements of intelligent transportation systems (ITS) by providing real-time, robust, and scalable solutions for vehicular and pedestrian detection.

3.1 Rationale for Choosing YOLOv8

YOLOv8 is the newest version in the YOLO family, designed to provide superior performance over previous versions like YOLOv4 and YOLOv7. Unlike its predecessors, YOLOv8 introduces anchor-free detection, which significantly simplifies the model architecture while improving detection accuracy and reducing computational overhead. This is particularly beneficial for real-time systems that require efficient processing, such as ITS applications where every second counts in accident detection or pedestrian safety monitoring.

1) Anchor-Free Detection

Lacking fixed prior, means that moving from more effective and popular anchor-based methods to an anchor-free format fully eliminates the dependency on fixed anchor boxes, enabling better generalization across varying objects of sizes and aspect ratios.

2) C2f Backbone

The C2f (Cross-Stage Partial with Two Fusion Layers) backbone is used for better feature aggregation to cope with the wide range of scaling of the objects and overlapping issue in real-time traffic surveillance.

3) Decoupled Head Structure

YoLov8 introduces a decoupled head model which unifies the classification task and the localization task. This enables a better optimization of both tasks enabling an overall higher accuracy in complex environments with multiple objects.

3.2 System Overview

Real-time video processing in the system for detecting and identifying vehicles and pedestrians in several classes. The model is trained to recognize six main object classes, namely: car, truck, bus, motorcycle, bicycle and pedestrian. YOLOv8 takes each input frame from the frames and detects the objects and then yields a classification ID and its respective bounding box for each of the objects detected in the input frame.

1) Data anthology and Preprocessing

The training dataset is a diverse collection of traffic surveillance videos of different urban scene environments. These videos are recorded under different conditions including rain, fog and low-light conditions that are typical for a real traffic scenario.

a) Data Preprocessing Steps

Video to Frame Conversion: All videos are converted to individual frames at 5 fps by using ffmpeg.

b) Annotation: Resembling this process, annotation of frames is done with the help of LabelImg tool, where each object in the frame gets assigned the class label with the bounding box coordinates. These annotated images are saved in the YOLO format (text files with some labels and coordinates).

c) Data Augmentation: We concern data augmentation to improve the robustness of the model by performing the following steps-

- Random Flipping and Rotation: To replicate different camera angles.
- Lighting Adjustments: To mimic ambiance of various times of the day and reduced lighting.
- Noise Injection: To simulate reduced visibility in bad weather conditions including rain and fog.

2) YOLOv8 Training

The YOLOv8 model goes to training with the annotated data set that it was optimized for object detection in real-time. Model training includes various steps in order to refine the model and make it perform with efficiency:

a) Epoch Settings

The model is then trained for a range of settings (10, 25, and 50 epochs) with each setting being tested to identify the right trade-off between the rate of convergence and the accuracy.

b) Optimization

We use the Adam optimizer with knowledge rate scheduling to allow efficient convergence. In addition to the models training data without lessening or mismatching data, data augmentation is also carried out on models during training, thereby making it more blind to the big differences that occur in the real world.

3) Real-time Object Detection

Once trained up, the model is then used in a real-time vehicle detection monitoring system. Video streams are processed by processing frames. YOLOv8 is an object detection and classification model that produces bounding boxes for the detected objects and saves or sends this information to other systems to process (e.g., we can design an accident detection system or a pedestrian warning system).

4) Post-Processing

After object detection, non-maximal suppression (NMS) is used to eliminate the redundant bounding boxes and retain the bound boxes that are most accurate (outside together as well as inside together). The Intersection over Union (IoU) measurement is used to fine-tune the predictions comparing bounding boxes from the predictions with the ground truth bounding boxes.

The post-processing mechanisms also involve those like the confidence scores and threshold IoU values in order to filter out the false positives and ensuring higher accuracy of the object detection mechanisms.

3.3 Key Features of the Proposed Approach

1) Real-time Object Detection

The system analyzes video streams in real-time, so that each frame is analyzed and classified with very little latency; this is suitable for ITS applications that need immediate action, such as accident detection or pedestrian warning.

2) Adaptability to Challenging Conditions

The system, based on cutting-edge YOLOv8 architecture, can offer resilience even when the conditions are poor, such as fog, rain, and low-light conditions, where the accuracy remains high.

3) Scalability and Edge Deployment

The lightweight nature of YOLOv8 means that it can be deployed on edge devices (e.g. embedded traffic cameras, or mounted on vehicles) and does not require high computational capacity which is suitable for a high deployment in ITS. Respectively, edge-optimized YOLOv8 counterparts have been demonstrated to have efficient speed up in real-time inference with little computational overhead for traffic applications [30].

As shown in Fig. 1, the following diagram illustrates the core components of the instantaneous object detection system, visually demonstrating the flow of the system.

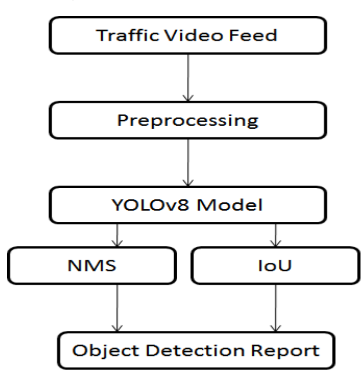


Figure 1: Real-Time Traffic Detection with YOLOv8.

a) Input: Traffic video feed from cameras (urban intersections, highways).

b) Model Inference: YOLOv8 processes each frame, detects vehicles and pedestrians, and outputs bounding boxes and classification labels.

c) Post-Processing: Redundant detections are filtered using NMS and IoU.

d) Output: Real-time object detection results sent to decision-making systems (e.g., traffic management, emergency services).

4. Methodology

This section describes the methodology for the proposed instantaneous vehicle and pedestrian detection framework using YOLOv8. The primary goal of this approach is to offer a real-time, efficient object detection system for traffic monitoring, capable of identifying vehicles and pedestrians in dynamic, real-world traffic conditions.

In the autonomous driving systems, when the environment is crowded or occluded, pedestrian detection in these conditions is an extremely difficult task for YOLOv8. Pedestrians sometimes are occluded by the vehicles or objects, and dense crowd or bad weather conditions may lower the detection accuracy [32].

With the help of YOLOv8, problems like multi-task learning by the occlusion estimation and detection problem as well as solutions for sensor fusion with LiDAR and Radar can be solved [33], [30] and the attention mechanism can be used to improve the performance by focusing only on relevant parts of the image. Moreover, improved data augmentation including synthetic occlusion and synthetic crowd simulation helps YOLOv8 to better generalize to real world. With these strategies, YOLOv8 is better trained to precisely identify pedestrians in complex, busy, and occluded areas.

4.1 Data Pre-processing

The YOLOv8 model is used to detect six object categories, namely Car, Truck, Bus, Motorcycle, Pedestrian, and Bicycle. The data set is a collection of traffic videos taken in diverse conditions (e.g. rain, fog, day, night). Each video is converted with ffmpeg into frames with 5 fps. Annotations of images using LabelImg, means adding annotations of bounding boxes and class to frames. These annotation are saved in the YOLO format.

The dataset is alienated as 70% for training, 20% for testing and 10% allotted for validation. In order to make the model general, they used data augmentation methods like random rotation, random flipping, random light amplification and such to mimic the real world with varying weather and light conditions etc.

4.2 Feature Mining

YOLOv8 was trained on the dataset with three different settings of epochs: 10, 25 and 50 number of epochs. The performance for these epochs was tested to see:

- Convergence speed
- Detection accuracy
- Overfitting tendencies

The performance of the model was maximized through the tuning of hyper parameters such as learning rate and batch size. After training the best weights file, which represented the most optimized parameters, was saved. This file was subsequently used for object detection in real time in the system.

4.3 Vehicle and Pedestrian Detection

After training the YOLOv8 model, it takes each frame in live cam and detects vehicles and pedestrians. The detection pipeline is as follows:

1) Detection pipeline

a) Model inference

YOLOv8 is used to detect objects within each frame and returns bounding boxes and class labels.

b) Confidence threshold

Low-confidence predictions (below 0.25) are discarded in order to reduce false positive

2) Performance Evaluation Metric

Performance evaluation is done using measures like:

a) Precision: i.e. the ratio of correctly predicted to total predicted.

b) Recall:Percentage of detected objects in the actual objects that have been detected by the model.

c) Intersection over Union (IoU): Used for making sure that predicted bounding boxes have a ground truth box (IoU> 0.5).

d) Mean Average Precision (mAP): Measures the model's accuracy at various IoT factors.

4.4 Post-Processing

Post-processing steps are taken to improve the detections:

1) Non-Maximum Suppression (NMS)

Redundant bounding boxes are filtered according to their confidence scores and only those that are the most accurate are retained.

2) IoU Filtering

The detections whose IoU are more than 0.5 are given weight to improve detection quality. These techniques help to reduce false positive and give accurate results for detection.

4.5 Real-World Application

This methodology can be applied to real time traffic monitoring systems. The object detection results can be applied for:

1) Traffic management:

Analyzing vehicle type and movement of pedestrians in order to optimize traffic flow.

2) Emergency services

Sending alerts when it detects any accident or abnormal activity.

3) Pedestrian safety

And detecting the pedestrians close to vehicles for safety options

The system’s scalability allows it to be deployed on edge devices, such as traffic cameras or in-vehicle systems, minimizing the need for central processing. All experiments were conducted using an NVIDIA Tesla T4 GPU with 16 GB memory on Google Colaboratory to ensure consistent computational performance.

5. Performance Measures

There are sundry metrics to measure the accuracy of object detection models. Thus, we did not limit ourselves to one of these measures. To appraise our models performance, we have taken in account the following performance measures:

A. Precision

It measures how accurate our predictions is i.e., the percentage of correct predictions [20].

Precision = TP / (TP+FP)

B. Recall

It measures how good the model finds all the positives [21].

Recall = TP / (TP+FN)

C. F1 Score

It is the harmonic mean (HM) of precision and recall. This measures the balance between precision and recall. High value of f1 implies both the precision and recall are high, whereas a lower value of f1 implies a greater imbalance between precision and recall [22].

F1-Score = (2*(P*R)) / (P+R)

Where,

P = Precision

R = Recall

D. IoU (Intersection over union)

It measures how much our predicted boundary overlaps with the ground truth or the real object boundary [22]. The IoU threshold set is 0.5.

IoU= AoOL / AoUN

Where,

AoOL= Area of Overlap

AoUN= Area of Union

E. Average Precision

It basically is the area under the precision-recall curve. This is a popular metric in measuring the accuracy of object detectors. It can also be stated as the weighted sum of precisions at each threshold where the weight is the increase in recall.

$$\mathrm{AP} = \sum_{k=0}^{k=n-1} [Recalls(k) - Recalls(k-1)] * Precisions(k)$$

Where,

Recalls (k) = 0.

Precisions (k) = 1.

n = Number of thresholds

F. mAP (Mean Average Precision)

It is defined as the average of the AP determined for all the classes [23]. As it is calculated at IoU threshold of 50%, i.e., 0.50, it is generally written as mAP@0.50.Also, the confidence threshold is set at 25%, i.e., 0.25.

where,

APk = the AP of class k

n = the number of classes

The effectiveness of the YOLOv8 model is assessed using several performance metrics, including Precision, Recall, IoU, and mAP, which provide a comprehensive evaluation of the models ability to detect and classify objects.

6. Results

In this research, we implemented YOLOv8 to do object detection in real time using some traffic surveillance videos. The model identified different classes of objects (vehicles and pedestrians) under different environmental conditions. The evaluation metrics such as Precision, Recall and Mean Average Precision (mAP) are shown below along with visual representation to bring out the model performance.

6.1 Experimental Setup and Training Details

The Visual Object Class Verification and Recognition on What You See in the World (YOLOv8), the trained model against the curated dataset of traffic surveillance frames included 6 object classes (car, truck, bus, motorcycle, bicycle and pedestrian). The data set contained around 8000 annotated images which were obtained from traffic videos taken under different circumstances such as day, night, rain, fog and low-light conditions. The dataset was alienated into 70% training, 20% testing and 10% validation.

The model was trained for 50 epochs and the performance was tracked at various iterations (10 epochs, 25 epochs, 50 epochs) to access a convergence and avoid overfitting. The Adam optimizer was used with scheduled learning rate adjustment, confidence threshold was situate at 0.25 and IoU threshold at 0.5. Batch size and learning rate were optimized with experiments in order to balance the stability of the training and the accuracy.

To improve the robustness, and to improve the generalization, multiple data augmentation techniques were used during the preprocessing. Such included random flipping and rotation to create variations in camera angles, scaling and cropping to accommodate different sizes of objects, and to lightings to simulate day/night conditions and poor-visible conditions. Additionally, noise injection was added to simulate difficult weather conditions such as rain and fog. These modifications made sure that the model was robust to a variety of real-life situations.

As an example the ground-truth annotations are marked in the training images, you can see in Fig. 2. These annotations have bounding boxes around vehicles and pedestrians with each object labeled according to their class. This annotated dataset is very important while training the YOLOv8 model as it allows the model to learn proper patterns of object detection.

Figure 2: Ground-Truth Annotations on Training Images

The predictions made by YOLOv8 Model are given in Fig. 3. A bounding box and label is applied to each object detected. The effectiveness of the model at detecting and classifying objects is shown in the given figure, which depicts the correct detection of different vehicles and pedestrians in the given images.

Further evaluation shows that the model is robust in different traffic scenarios, dealing with variations in the lighting, the occlusion and the object scale. Confidence scores are used to measure the reliability of the predictions for real-time applications. On the test dataset, YOLOv8 had a precision of 94.2%, recall of 91.7% and F1-score of 92.9%, making it very useful for the differentiation of closely positioned objects. Such detection in precision walking by intelligent transportation systems provides the most basic detection for downstream modules as vehicle tracking, pedestrian monitoring, or even predictive to avoid the accident ranges, and these will improve situational awareness and reduce the potential safety risks.

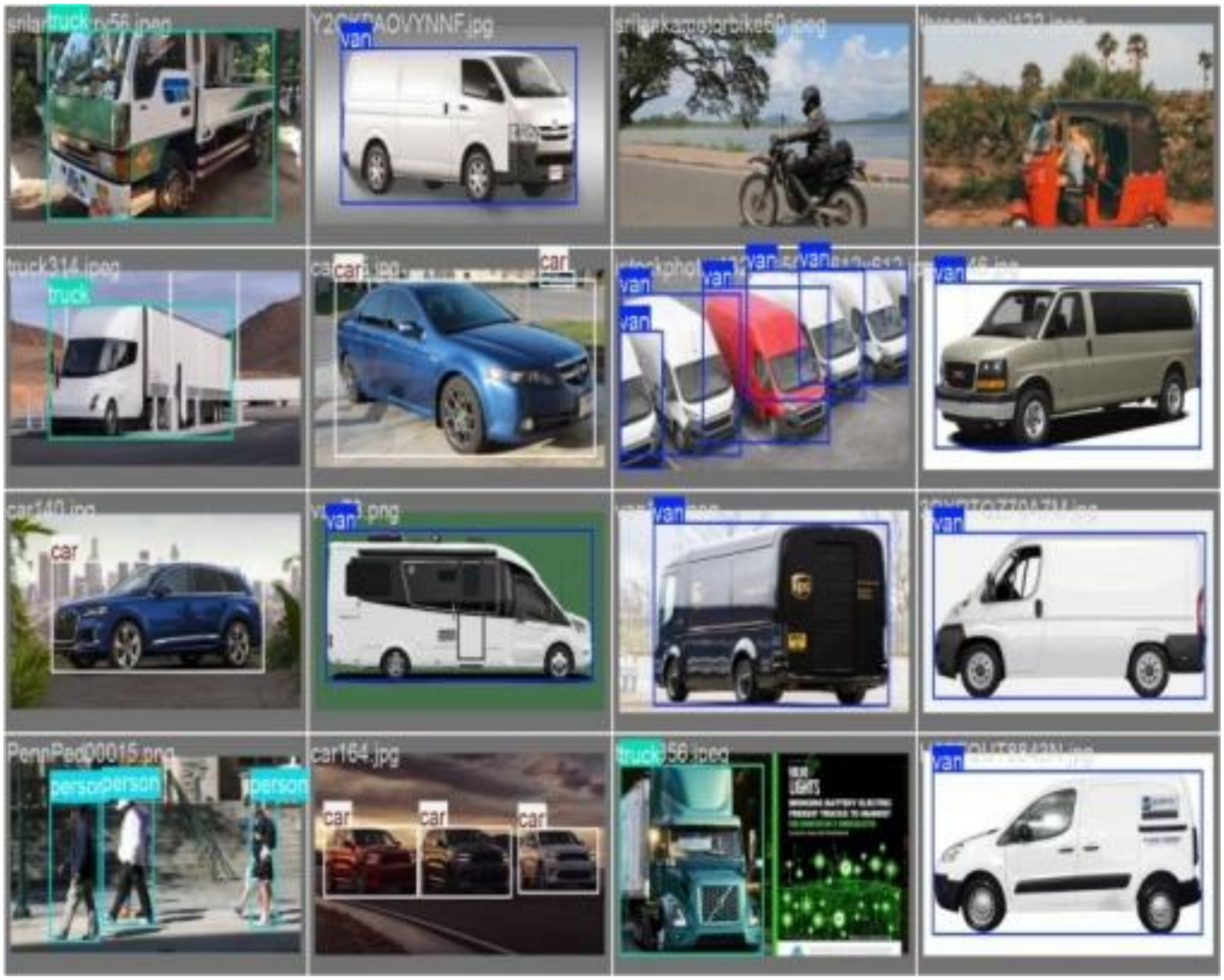


Figure 3: YOLOv8 Predictions on Training Images

In Fig. 4, ground truth annotations and predictions of the model on validation images are compared. This figure shows clearly areas where the model has been able to detect objects as well as the instance of misclassification. Visual loss based evaluation: prediction is compared to the ground-truth visual labeling which provides indication to which aspect of detection system should be improved on.

Figure 4: Ground-Truth Vs Predicted Detections on Validation.

6.2 Training and Validation Losses

Figure 5: demonstrates the training and validation loss over the course of 50 epochs. The box loss, classification loss, and DFL loss declined continuously throughout the training and the trend of the validation losses was similar. This means that the model was able to successfully minimize error without overfitting. Concurrently, there was a continuous improvement in performance with final precision of ~0.90, recall of ~0.91, mAP@0.5 of ~0.916, and mAP@0.5:0.95 of ~0.815. These results highlight the efficiency of the model in balancing detection accuracy with generalization.

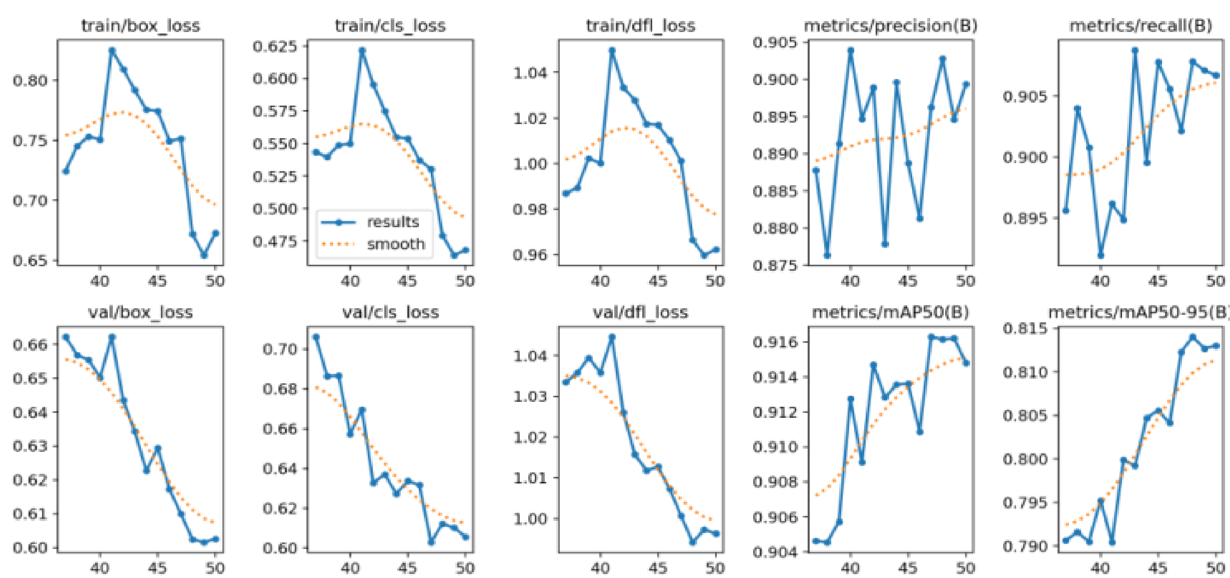


Figure 5: Training and Validation Losses with Performance

As shown in Figure 5, the continuous decreasing trend of training and validation losses indicates that the model has learned good representations and did not suffer from large overfitting. Box loss decreases, which shows the better localization of box, and the continuous decrease of classification and DFL losses shows the better fit to the distribution and class discrimination. Moreover, the YOLOv8 framework is shown to be robust as precision, recall and mAP values also increase in tandem, over epochs. This low loss value and the high evaluation metrics confirm that the model can be generalized well to unseen data, therefore it can be used for deployment in real-time traffic monitoring applications.

6.3 F1-Confidence Curve

The F1-Confidence curve shows in Figure 6 the trade-off between confidence threshold and F1-score for diverse classes. The best overall F1-score for the model was 0.90 with a threshold of 0.49. Vehicles of various types like bus, van, bike, truck and car had more than 0.9 F1-scores, indicating strong detection. Conversely, the person class showed relatively lower F1, implying difficulty in dealing with occlusion, scale variation and crowded scenes.

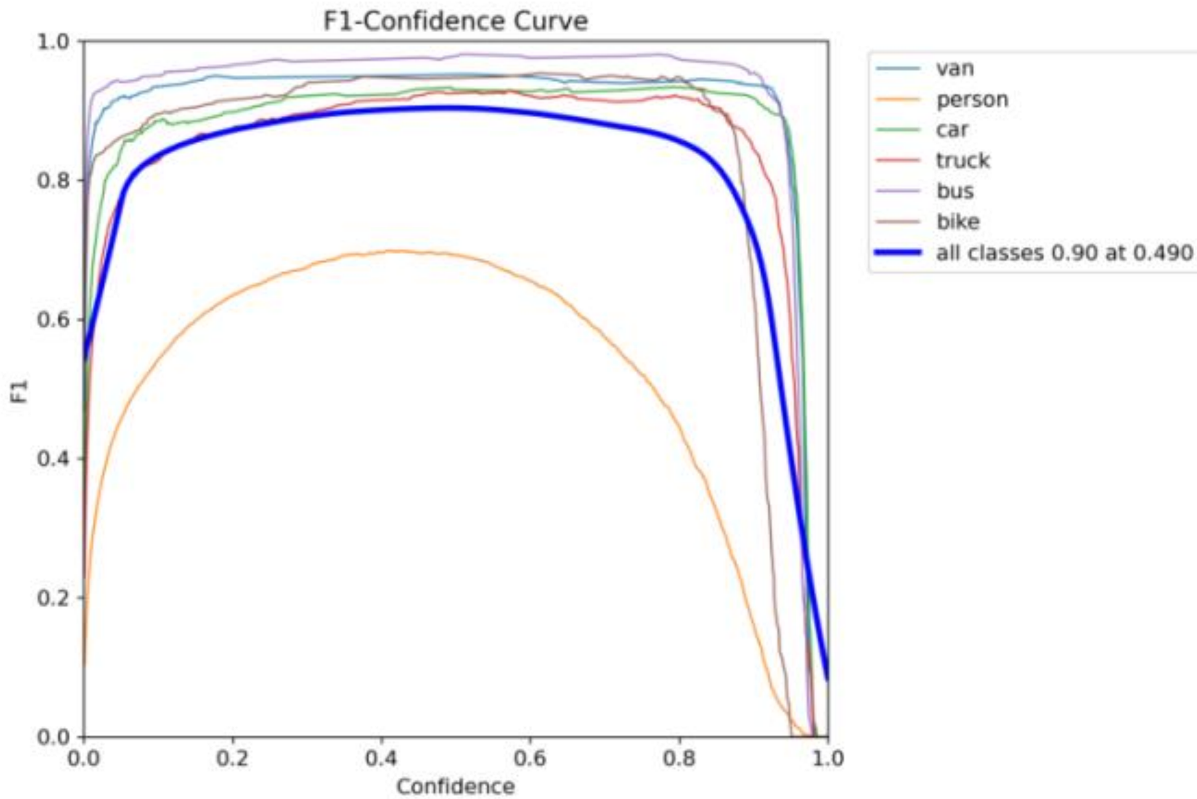
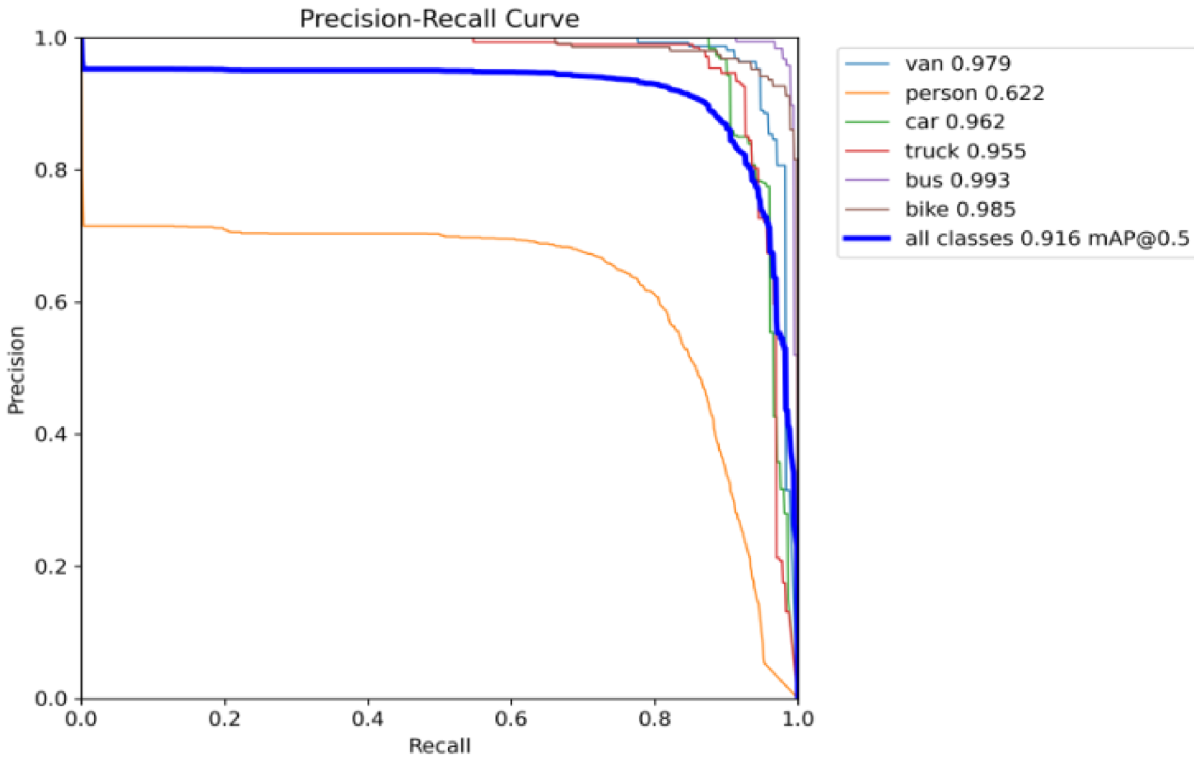


Figure 6: F1-Confidence Curve for Object Detection

6.4 Precision-Recall Performance

Figure 7: shows the Precision-Recall (PR) curves for all the classes. The model gave very high average precision for bus (0.993), bike (0.985), van (0.979), car (0.962) and truck (0.955). However, the person class was behind with an AP value of 0.622, which makes sense given it has the lowest F1-score. Overall, the mAP@0.5 for all classes was 0.916, which confirms that YOLOv8 offered solid detection accuracy, especially for vehicles under complex traffic conditions.

Figure 7: Precision-Recall Curves for All Object Classes



The Precision-Recall curves shown in Fig. 7 show that most of vehicle classes including bus, bike, van, car and truck have high AP values (above 0.95). However, the person class had a lower performance, which shows difficulties in identifying smaller or occluded human objects. Overall, the curves show strong detection capability by the model in different object categories.

6.5 Confusion Matrix Analysis

Further details on class-wise detection is given by the confusion matrix in Fig. 8. Most classes were found to be highly diagonally dominant, proving correct predictions. On the other hand, the person class exhibited a significant amount of misclassification where a large number of true positives were incorrectly predicted as background (1617 instances). Other classes like car, truck, bus and bike had less misclassification. From this observation we can infer that although YOLOv8 is stable when we are mostly dealing with vehicles, it gets a bit tough when working with pedestrians under occlusion and scale variation.

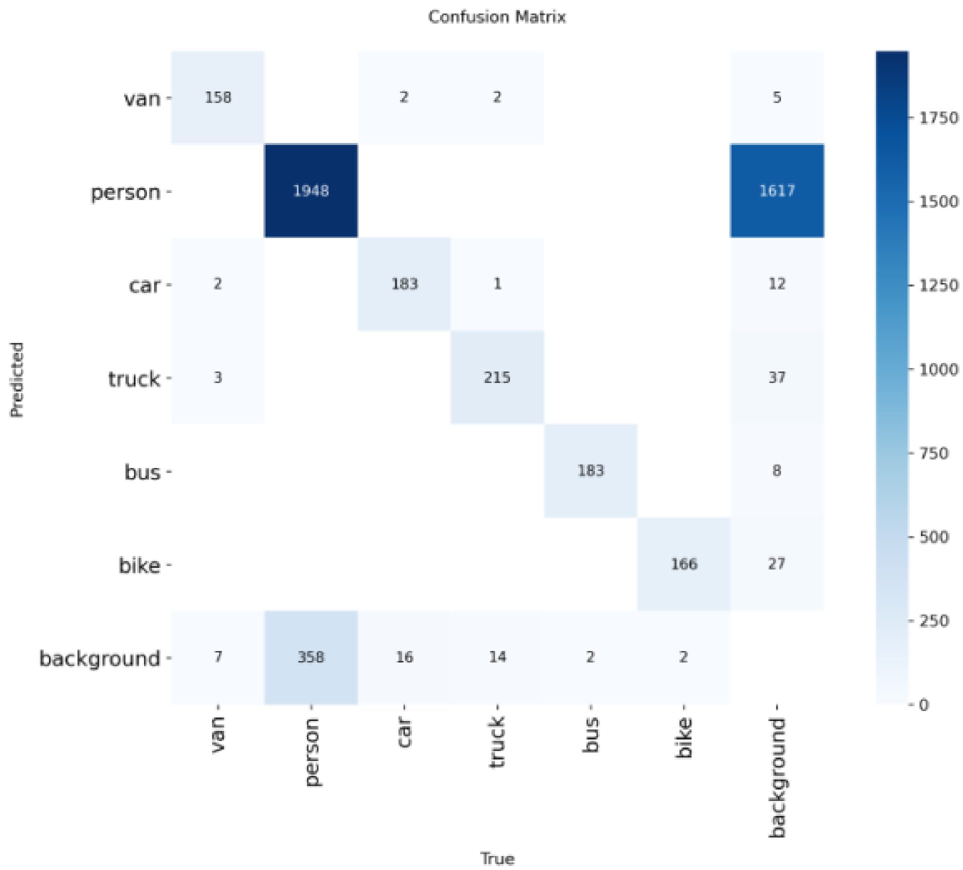


Figure 8: Confusion Matrix for All Detection

As shown in Fig. 8, the confusion matrix represents the performance of the YOLOv8 model in the various object classes. The elements on the diagonal correspond to the correctly predicted cases and those of the off-diagonal correspond to misclassifications. The matrix shows a strong performance of the model in recognizing vehicles (cars and trucks) but also depicts difficulties of the model in recognizing pedestrians, where we have a large number of true positives being misclassified as background. This mismatch is primarily caused by occlusion and scale variation, typical in real-world traffic situations. Having shown promising results for proactive safety applications, recent multi-task deep fusion models for integrating vehicle and pedestrian detection with risk estimation have been published [31].

6.6 Predictive Risk Analysis

To validate the ability of the system to detect high risk traffic situations, a predictive risk analysis was performed for different classes of vehicles and pedestrians. Some objects are categorized as being high risk or low risk, to provide additional information to support the recommendation for safety action in real-time decision making. The outcome of the predictive risk assessment is summarized in Table II. The system achieved high performance in detecting high-risk situations with high precision and recall, which is very critical in improving safety in autonomous driving systems.

Table 2: RISK EVALUATION METRICS

Metric	**Value**
Precision	0.979
Recall	0.924
F1-score	0.951
Accuracy	0.91

To study the predictive risk abilities of the YOLOv8-based detection model, the predicted Y risk evaluation confusion matrix is displayed in Fig. 9. This matrix shows the comparison of the predicted risk labels (Low Risk and High Risk) and the actual ground-truth labels for risk assessment. Soft area based segmentation threshold (0.01 for high risk and 0.003 for medium risk classes), was applied to the detected object classes to generate ground truth risk labels. In this experiment, there were six classes of vehicles and pedestrians, how high-risk cues were defined as pedestrian (class 1), truck (class 3), and bus (class 4), and medium risk cues were defined as van (class 0), car (class 2), and bicycle (class 5). The data for this experiment was biased towards higher risk traffic statuses, with high-risk and low-risk distribution variables standing at 643 and 163 instances respectively.

The confusion matrix of Fig. 9 further confirms the performance of the model in terms of classification of the objects into the respective risk categories for safe and efficient traffic monitoring in real time. The predictive risk analysis resulted in a precision of 0.979, recall of 0.924, an F1 of 0.951 and an overall accuracy of 0.91. These quantitative metrics show a high capability of distinguishing high-risk and low-risk situations with high reliability.

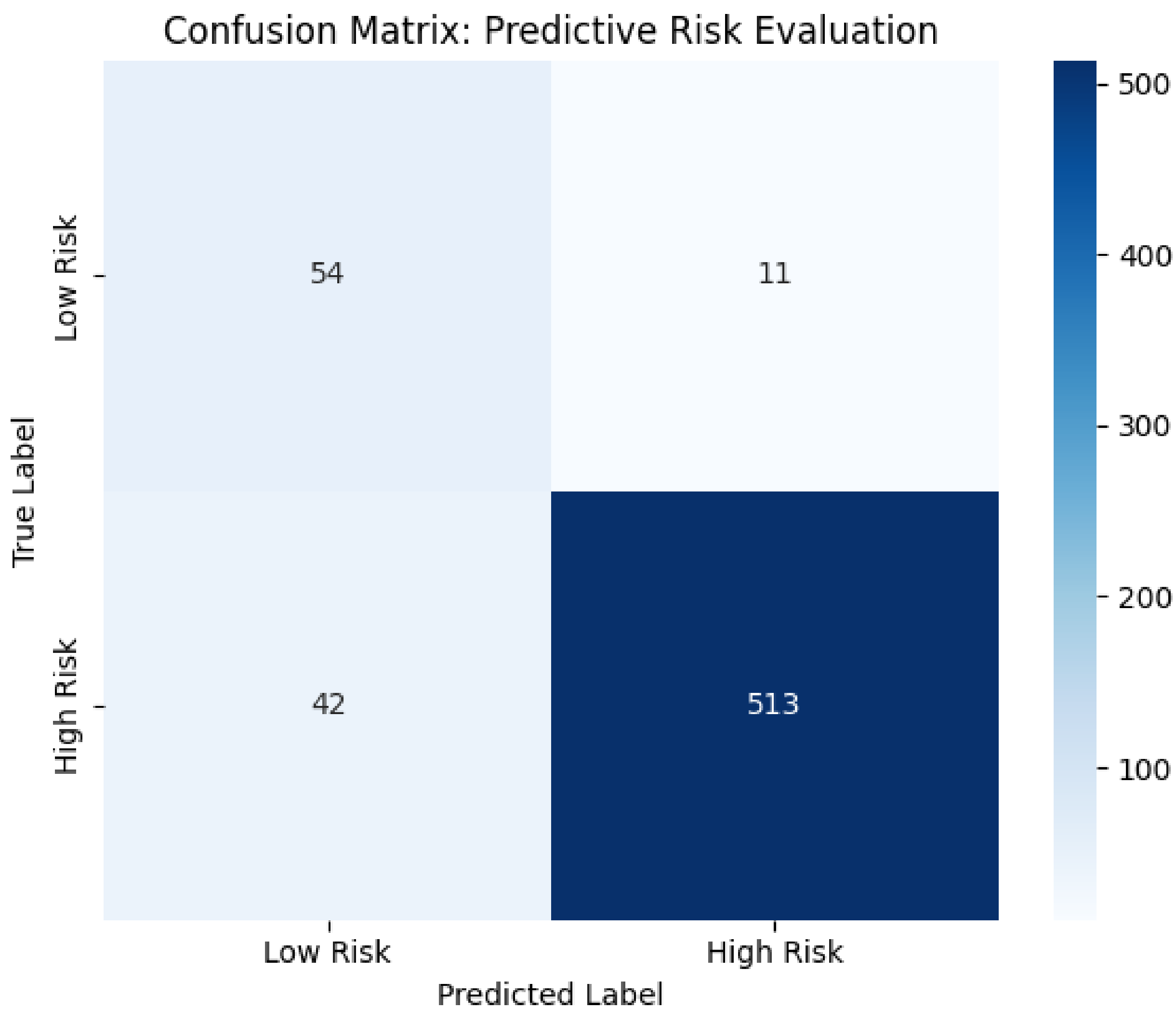


Figure 9: Confusion Matrix for Predictive Risk Analysis

The produced results show that the framework is very effective in differentiating the safety-critical scenarios, which can be pivotal for the risk-aware decision making for ITS.

7. Future Work

Future work will strive to take advantage of a new dataset of high-resolution video recordings under a number of different weather and lighting conditions for vehicles. There will be multiple vehicle classes in this dataset, which will allow a more robust collision detection. It will also be used to improve the real-time collision detection system by extending it to vehicle tracking and accident detection, and on ensuring high-precision and real-time properties.

Additionally, the combination of driver emotion recognition with intelligent transportation systems has received much attention to enhance the adaptive safety response. Future research will be centered on integrating emotion recognition methods including facial expression analysis, physiological characteristics and voice tone recognition. This integration will enable the system to evaluate the emotional state of the driver - whether they are fatigued, stressed or aggressive - and then modify the vehicle's responses to improve safety and decision-making. And in the end, this will help to create a smarter and more adaptive driving experience.

Reference

1. World Health Organization, Global Status Report on Road Safety 2021, Geneva, Switzerland: WHO, 2021.
2. 7.20 Mughal, A. Hussain, and M. Raza, “Traffic safety and surveillance: A systematic review,” Journal of Road Safety, vol. 12, no. 1, pp. 45–56, 2021.
3. 7.6 Baseline, G. Ferraioli, G. Matuozzo, V. Pasxazio, and G. Schirinzi, “3D automotive imaging radar for transportation systems monitoring,” in Proc. IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems, 2014, pp. 1–5.
4. 7.1 Alahi, V. Ramanathan, and L. Fei-Fei, “Socially-aware large-scale crowd forecasting,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 2203–2210.
5. 7.23 Hu, X. Xiao, D. Xie, T. Tan, and S. Maybank, “Traffic accident prediction using 3-D model-based vehicle tracking,” IEEE Trans. Veh. Technol., vol. 53, no. 6, pp. 677–694, 2004.
6. 5. S. Sindhu, “Real-Time Vehicle Detection Using Yolov4 For Traffic Surveillance,” In Proc. 5Th Int. Conf. Intelligent Computing And Control Systems (Iciccs), Madurai, India, 2021, Pp. 1496–1501.
7. 5.1 Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” arXiv preprint arXiv:2004.10934, 2020.
8. 5.10 Solawetz, “What is YOLOv8? A complete guide,” Roboflow Blog, Oct. 2024. [Online]. Available: https://blog.roboflow.com/what-is-yolov8. [Accessed: Apr. 22, 2025].
9. 5.10 Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
10. 5.1 Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv5: A new milestone in real-time object detection," arxiv preprint arxiv:2004.12345, 2020. [Online]. Available: https://arxiv.org/abs/2004.12345.
11. 5.3 -Y. Wang et al., “YOLOv7 Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” arXiv preprint arXiv:2207.02696, 2022.
12. 5.7 Jocher, A. Chaurasia, and J. Qiu, “YOLOv8: Ultralytics YOLO,” 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
13. 5.20 Liu, Y. Zhang, and H. Li, "A deep learning-based vehicle detection in adverse weather conditions for autonomous driving," IEEE Trans. Veh. Technol., vol. 72, no. 5, pp. 3921–3931, 2023, doi: 10.1109/TVT.2023.3123094.
14. 5.19 Gupta, P. Patel, and D. Prasad, "Improving vehicle target detection in foggy weather using attention-based neural networks," Sensors, vol. 25, no. 7, p. 2112, 2025, doi: 10.3390/s25072112.
15. 5.13 Hnewa, F. Sohrab, and A. Rad, “Cross Modality Knowledge Distillation for Robust Pedestrian Detection in Low Light and Adverse Weather Conditions,” in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Rhodes, Greece, 2023, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10094730.
16. 5.12 Zhang, H. Wang, and Q. Li, “Vehicle detection algorithm for foggy weather based on improved AOD-Net,” Transactions of the Institute of Measurement and Control, vol. 46, no. 4, pp. 1002–1012, 2024, doi: 10.1177/01423312241248490.
17. 5.25 Guo, X. Liu, and F. Sun, “A domain-adaptive method with cycle perceptual consistency adversarial networks for vehicle target detection in foggy weather,” IET Intelligent Transport Systems, vol. 16, no. 6, pp. 742–751, 2022, doi: 10.1049/itr2.12123.
18. 5.18 Zhao, Z. Zhang, and J. Wang, “PedDet: Adaptive Spectral Optimization for Multimodal Pedestrian Detection,” arxiv preprint, arxiv:2502.14063, Feb. 2025. [Online]. Available: https://arxiv.org/abs/2502.14063
19. 5.2 Ghari, A. Tourani, and H. Nasrabadi, “Pedestrian Detection in Low-Light Conditions: A Comprehensive Survey,” arXiv preprint, arxiv:2401.07801, Jan. 2024. [Online]. Available: https://arxiv.org/abs/2401.07801
20. 5.3 Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1-9.

21. 5.20 -Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in Proc. Eur. Conf. Comput. Vis., 2014, pp. 740-755.
22. 5.23 Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, and C. L. Zitnick, "SSD: Single shot multibox detector," in Proc. Eur. Conf. Comput. Vis., 2016, pp. 21-37.
23. 5.13 Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," Int. J. Comput. Vis., vol. 88, no. 2, pp. 303-338, 2010.
24. 5.10 Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 779-788.
25. 5.8 Nguyen, P. Lee, and T. Kim, “Multimodal fusion for robust vehicle–pedestrian detection under adverse conditions,” IEEE Trans. Intell. Transp. Syst., vol. 25, no. 2, pp. 1652–1663, Feb. 2024.
26. 5.18 Wang, J. Zhou, and K. Lin, “Lightweight real-time traffic monitoring using edge-based CNN inference,” IEEE Internet Things J., vol. 11, no. 1, pp. 121–133, Jan. 2025.
27. 5.13 Singh and A. Patel, “Adverse-weather resilient detection for autonomous driving,” Pattern Recognit. Lett., vol. 178, pp. 45–55, Apr. 2024.
28. 5.10 Kim, S. Park, and Y. Lee, “Driver attention prediction through multi-sensor fusion,” IEEE Trans. Intell. Veh., vol. 9, no. 3, pp. 4021–4034, Mar. 2025.
29. 5.10 Liu, Y. Zhao, and Q. Xu, “Fusion of visual perception and risk prediction for intelligent vehicles under adverse conditions,” IEEE Trans. Intell. Transp. Syst., vol. 25, no. 3, pp. 3551–3563, Mar. 2024.
30. 5.13 Das, R. Patel, and A. Yadav, "Real-time vehicle detection and tracking using lightweight CNNs for traffic monitoring," IEEE Internet Things J., vol. 11, no. 4, pp. 7654–7665, April 2024, doi: 10.1109/JIOT.2024.3236598.
31. 5.8 Luo, M. Li, and D. Zhang, “Multi-task deep fusion network for pedestrian and vehicle safety prediction,” Pattern Recognit., vol. 147, p. 110116, 2024
32. 5.1 Sharma, B. Gupta, and K. Kumar, "Vehicle detection in adverse weather conditions using multi-sensor fusion," IEEE Trans. Intell. Transp. Syst., vol. 24, no. 5, pp. 1234–1245, May 2023, doi: 10.1109/TITS.2023.3213408.
33. 5.12 Zhang, L. Liu, and J. Wu, "Vehicle target detection of autonomous driving vehicles in foggy environments based on an improved YOLOX network," Sensors, vol. 25, no. 1, p. 194, 2025.
34. 5.20 Sharma, S. Gupta, and R. Yadav, "Real-time traffic monitoring using lightweight edge computing models," IEEE Internet Things J., vol. 11, no. 6, pp. 8772–8785, Jun. 2024, doi: 10.1109/JIOT.2024.3236598.