



Detection of traffic rule violation in University campus using deep learning model

Pooja Chaturvedi¹ · Kruti Lavingia¹ · Gaurang Raval¹

Received: 29 May 2023 / Revised: 3 August 2023 / Accepted: 11 August 2023 / Published online: 29 August 2023

© The Author(s) under exclusive licence to The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2023

Abstract Implementing and monitoring traffic rules are essential for reducing accident rates and traffic rule violations. The automated traffic rule monitoring system ensures strict adherence to the traffic rules with low human effort. The system proposed identifies and acknowledges the two-wheelers violating the traffic rule regarding triple riders. The intersection points in the University Campus act as the data collection centres and collect the data through live recordings captured by surveillance cameras. The proposed system consists of detection and automatic number plate recognition. To implement this system, the Darknet framework is used, which is based on You Only Look Once (YOLO v8) for identifying two-wheelers with triple riders. The Depth estimation algorithm is used to detect vehicles and persons, which can accurately detect near and far objects. The vehicles are classified as Violator or No Violator. The Connectionist Temporal classification algorithm is used to classify the vehicles as violators or no violators. The two-wheeler classified as a Violator is stored in the database along with the vehicle license plate number, which can be penalised by traffic monitoring authorities. The implementation results show that the system is viable, efficient and reliable. Thus, make the two-wheeler follow the traffic rules properly, reducing the chance of irresponsible driving. The self-generated dataset detects the traffic rule violation and

license plate number extraction. The system is trained on the considered dataset, and the best weights are used to implement the model. The proposed model achieves 94%, 96% and 97% performance for detecting triple rider, license plate and motorcycle detection.

Keywords Traffic rule monitoring · Deep learning · Triple rider · Automatic number plate recognition · YOLO v8

1 Introduction

The population in India has increased significantly in recent years. The increasing population has increased the number of vehicles registered and, subsequently, the road accident rates. The enforcement of traffic rules in the Indian context is essential because India recorded 155,622 road accident deaths in 2022. As per the latest report, India has over 21 crore two-wheelers registered as of August 2022. The two-wheelers have the highest share, approx. 74.7% in the composition of the vehicular population, followed by cars, jeeps and taxis at around 13.4% and other vehicles share is 6.9%, goods vehicle share is 4.4%, and buses has 0.7%. The fatalities have increased by 16.8%. This caused deaths per thousand vehicles from 0.45% to 0.53% in 2022. The analysis of the causes of deaths shows that road accident is mainly caused by overspeed, and the number of death persons is 2,40,828 from 4,03,116, which is 59.7% and has caused 87,050 deaths and 2,28,274 injuries. The impact of the other causes in road accidents is as follows: driving dangerously or carelessly and surpassing other vehicles has caused 1,03,629 road accidents (25.7% of the total) and has 42,853 casualties and injuries up to 91,893. Poor weather conditions were the reason for merely 2.8% of deaths (11,110 out of a total

✉ Kruti Lavingia
kruti.lavingia@nirmauni.ac.in

Pooja Chaturvedi
pooja.chaturvedi@nirmauni.ac.in

Gaurang Raval
gaurang.raval@nirmauni.ac.in

¹ Institute of Technology, Nirma University, Ahmedabad, India

of 4,03,116), whereas driving after drug/alcohol has caused 1.9% of the injuries, i.e., 7,235 and 2,935 casualties (<https://www.financialexpress.com/express-mobility/as-road-accidents-continue-to-rise-a-look-at-the-key-safety-norms-mandatory-in-india-now/2932079/>, https://morth.nic.in/sites/default/files/RA_2021_Compressed.pdf, <https://www.thehindubusinessline.com/data-stories/data-focus/overspeeding-accou-nts-for-60-of-the-road-accidents-in-india/article65852932.ece>).

The day-to-day increases in traffic rule violations by two-wheeler riders make it difficult for the manual traffic monitoring system to ensure safety and rule obedience. Hence, there is a need to automate the continuous monitoring and surveillance system (Kaffash et al. 2021; Chen et al. 2021; Sun and Boukerche 2021). The intelligent transportation system can aid in this problem. The innovative transportation system is an autonomous surveillance system combining several methodologies, such as edge-enabled cameras, sensors, communication, electronics, etc. The ITS system provides a fast, intelligent and efficient traffic monitoring solution based on the surveillance cameras mounted on the roads to detect traffic rule violations such as traffic signal violations, helmet violations, triple riding, etc.

The paper proposes an automated deep learning-based system to detect two-wheeler riders who violate the traffic rules regarding triple riders (Goyal et al. 2022). The YOLOv8 model is used to classify the two-wheeler rider in the violator or no violator classes. The self-generated data set through the surveillance system is considered to evaluate the proposed model's efficiency. The results show that the proposed system achieves the accuracy of 94% for the triple rider detection.

The main contributions of the work are summarised as follows:

- The automated system is designed to detect two-wheeler traffic violations in case of triple rider. The edge-enabled surveillance cameras were deployed at essential cross-points inside the university campus.
- YOLO v8 model is used to detect the two-wheeler and person. The results show that the YOLO v8 model effectively detects multiple objects and closer objects in a single image, and it can also detect overlapped objects with greater accuracy and in less time.
- The proposed system can be implemented in real-time surveillance, too.

The paper is structured as follows: object detection and recognition approaches in Sect. 2, recent work in Sect. 3, the proposed classification system in Sect. 4, performance evaluation in Sect. 5, limitations of the proposed method in Sect. 6 and Sect. 7 concludes the paper with a conclusion and future scope.

2 Object detection and recognition approaches

Object detection is the sub-task of computer vision, which involves identifying and locating the objects in the video or images. It can be applied to numerous applications such as robotics, surveillance and autonomous cars. Object class detection is usually based on a set of features, i.e., every object of a particular class has certain features, based on which the thing is classified into different classes. For example, if we want to classify animals with birds, we can distinguish both by the presence or absence of the wings. Object detection methods are generally categorised into two categories: deep learning-based and machine learning-based. The machine learning approach takes the defining features as input. Then, the different classification techniques (SVM, etc.) can be applied to classify the objects into other classes. In contrast, the deep learning-based methods provide a single-step solution for object detection and categorisation using convolutional neural networks. The Deep learning-based object detection algorithm can be broadly categorised into two types based on the number of times the input image is passed through a network: single-stage and two-stage detectors (Lohia et al. 2021). The different methods under these approaches are in the Fig. 1.

The stage detection approach uses two passes to process an input image to predict the presence and localization of the objects. In the first pass, the predictions are made regarding the object's existence and localization; in the second pass, the predictions are improved and final predictions are made. Examples of two-stage detection models are RCCN, Fast RCNN, Faster RCNN, RFCN and Mask RCNN (Girshick et al. 2014; Dai et al. 2016; Ren et al. 2015; Girshick 2015; Liu et al. 2016). These algorithms are also known as the region proposal-based approach because these algorithms attempt to build a boundary box around the object of interest to identify its location inside the image.

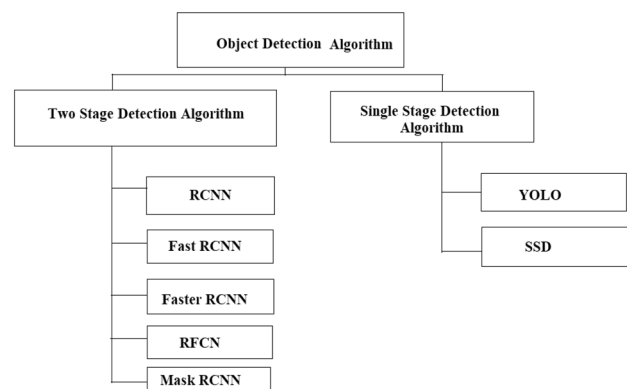


Fig. 1 Classification of object detection algorithms

Single-stage detection approaches pass the input image into the network only once to make predictions about the presence and localization of the objects. These methods process the input image in a single pass, which is computationally efficient. The limitation of the single shot detection algorithm is that they are less accurate in predicting the objects and is mainly suitable for real-time surveillance in a limited resource environment. YOLO is a single-stage detection system that processes an input image using a fully Convolutional Neural Network (CNN). A stage object detection strategy is the Single Shot Multi-box Detector approach (He et al. 2017). This method employs a single forward pass for object localisation and classification tasks. A multi-box approach is utilized for boundary box regression. The detector is a network that is divided into two phases: feature map extraction and object detection using convolutional filters. SSD extracts features using VGG 16. So, the object detection method selection depends on the application type for which it is implemented. Many optimization techniques using machine learning are proposed in the literature (Xiao et al. 2021a; Xing et al. 2022; Xiao et al. 2021b). In (Xiao et al. 2021a), authors have proposed a federated learning-based model for recognising human activity. In (Xing et al. 2022; Xiao et al. 2021b), authors have proposed the classification approaches for the time series data.

2.1 You look only once (YOLO) models for object detection and recognition

The proposed model is implemented for real-time surveillance of two-wheeler traffic, so we will use the YOLO-based detection and prediction model. You Look Only Once is a neural network-based model that can simultaneously predict the boundary boxes and class probabilities. YOLO is different from the previous classification approaches for the

detection of objects. The faster RCNN approach works by identifying the region of interest through the Region Proposal Network and reiterating these regions for the detection of the objects. In contrast, YOLO identifies and detect objects in a single step through a fully connected Convolutional Neural Network.

2.2 Working of YOLO model

The YOLO model (Xing et al. 2022) is a real-time object identification model that recognizes objects in live movies or photos. In 2016, Joseph Redmon and Ali Farhadi created the first version of the YOLO model. In 2018, an updated version of the YOLO model, version 3, was released. The YOLO model is made up of a fully Convolutional Neural Network and a neural network post-processing technique. The model's operation begins with the extraction of a single frame from live video. The image is then scaled to 416*416 in the following phase. The scaled image is then sent into a deep convolutional neural network, which detects the objects. Figure 2 depicts the YOLO model's backbone design. The detection network has 24 convolutional layers, followed by two fully connected layers. An alternating sequence of 1*1 convolutional layers is used to decrease the feature space from the preceding layer. ImageNet trains the first 20 convolutional layers, including temporary average pooling and fully connected layers. This augmented model is used for detection, and the findings show that the augmentation improves detection accuracy. The YOLO model's last fully connected layer predicts the probabilities of an object belonging to a specific class as well as the coordinates of the boundary boxes. The YOLO model uses the S*S grid to partition the entire image. The grid in which an object's center is located is in charge of detecting the object. The number of boundary boxes and the confidence ratings of the

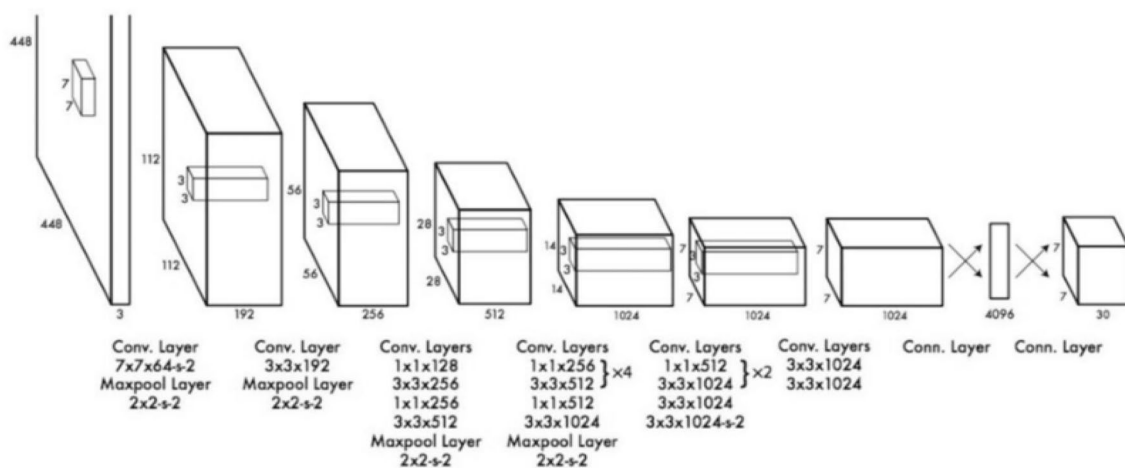


Fig. 2 Backbone architecture of YOLO Model (Xing et al. 2022)

boundary boxes are predicted for each object. Each prediction is a $C + B * 5$ tensor, where C is the number of classes of things and B is a set of border boxes. It is multiplied by five because each prediction of a boundary box consists of the following information:

[Box centre offset (B_x), Box centre offset (B_y), Box width (B_w), Box height (B_h), Object Score (S)].

Since the grid size is $S * S$, the predicted tensor will be of size $S * S * (C + B * 5)$.

The boundary box with the highest Intersection of Union concerning the ground truth is selected as the predictor of that object. The YOLO model uses Non-Maximum Suppression as a post-processing technique to discard the redundant or incorrect boundary box and gives a single boundary box as output.

Several versions of the YOLO model have been released since its inception in 2015. Each new version outperforms the previous model in some way (Xing et al. 2022; Xiao et al. 2021b; <https://www.v7labs.com>, blog, yolo-object-detection. xxxx; Redmon et al. 2016; Redmon and Farhadi 2016; Redmon and Farhadi 2018; Bochkovskiy et al. 2020). The timeline of the introduction of different versions of the YOLO model is shown in Fig. 3:

Yolo v2 model (Xiao et al. 2021b) improves the accuracy of the base YOLO model by incorporating the concept of anchor boxes. The anchor boxes are pre-defined boundary boxes of varying shapes and contrasts. The use of anchor boxes improves the detection efficiency. The YOLO v2 model introduced another stack of 53 convolutional neural network layers, resulting in 106 layers. Hence, it provides the better accuracy in the object detection tasks.

YOLO v3 (<https://www.v7labs.com>, blog, yolo-object-detection. xxxx) model improves the computational cost for multi-label classification tasks by replacing the softmax function with the independent logistic classifier for determining the label of the specific object. This version updates the loss function to binary cross entropy instead of a squared

mean error. Another improvement in this version was including a new backbone network, Darknet 53, for object detection, consisting of 53 layers. It also used the concept of a feature pyramid network, which determines the pyramid of feature maps to detect the objects at different scales.

YOLO v4 model (Redmon et al. 2016) was based on using new CNN architecture as Cross Stage Partial Network (CSPNET), a variation of RestNet architecture and comprises fewer layers up to 53. CSPNet is a specially enhanced version of ResNet explicitly designed for object detection tasks.

The next version of the YOLO model is YOLO v5 (Redmon and Farhadi 2016), which comprises a more dense architecture as a backbone network and is termed Efficient-Det. The training for the YOLO v5 is different from previous versions because it can now be trained for the objects belonging to many classes. Another improvement in this model was the introduction of Spatial Pyramid pooling, which can be used more accurately to detect objects of different shapes.

The next version of the YOLO model is YOLO v6, based on the new CNN architecture as EfficientNet-L2. The results show that this version is computationally more efficient and accurate than previous versions.

The next model in the timeline of YOLO models is the YOLO v7 (Redmon and Farhadi 2018). In this model, the seven anchor boxes were used to detect objects of different shapes and aspect ratios. The second improvement in this model uses a new loss function as focal loss, which is used to detect the smaller objects efficiently. This model is most efficient compared to the earlier versions regarding image accuracy, speed and resolution.

The latest version in this category is YOLO v8 (Bochkovskiy et al. 2020). The reason why this version outperforms its predecessors is as follows:

- It provides greater accuracy and flexibility in object detection tasks as measured on the well-known COCO Dataset.
- It has several developer-friendly features, including a command line interface for model training, detecting and predicting, and a structured Python library framework.
- It is an anchor-free model, which implies it can directly predict the centre of an object instead of predicting the offset relative to the anchor boxes. So, the detection using this model is more efficient.

3 Related work

Decision-making, pose detection, autonomous driving, path detection, and remote sensing are all applications of digital image processing and computer vision. (Jocher

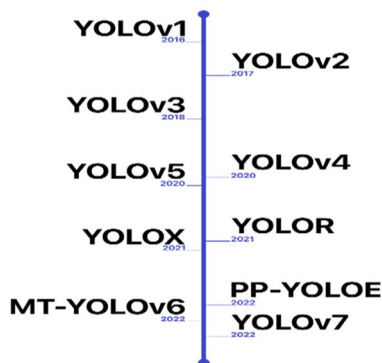


Fig. 3 Timeline of the inception of YOLO models (Xiao et al. 2021a)

2020; Wang et al. 2022). The present research in this subject focuses on using deep learning models, which have demonstrated considerable success in various application domains (<https://github.com>, ultralytics.com, ultralytics.com). Several deep learning-based algorithms for detecting helmet violations on bike riders have been proposed in the literature.

In (Ul-Haq et al. 2022), authors have proposed a CNN-based helmet violation detection system. The training of this model has been done on the dataset of 493 images. The detection system used 4 CNN-based models: MobileNet, GoogleNet, VGG 16 and VGG 19. The results demonstrate the supremacy of GoogleNet with an accuracy of 85%.

In (Mehmood et al. 2022), authors have proposed a helmet rule violation detection, which can also capture the license plate of the two-wheelers. The proposed system used HOG metrics along with the CNN models. The simulation results show that the proposed approach achieves 95% accuracy.

In (Boonsirisumpun et al. 2018), authors have used two YOLO-based models, YOLOv3 and YOLO-dense, for the helmet detection violation. The performance of the proposed system is evaluated using the self-generated and freely available datasets. The simulation results show that the proposed models achieve the mAP of 95% and 98%, respectively.

In (Yang 2022), authors have used the deep learning model RetinaNet for helmet violation detection. The proposed model was trained on the self-generated dataset and achieved an accuracy of approx.—73%.

In (Siebert and Lin 2020), authors have used the deep learning model based on Faster R-CNN to detect bikers without helmets. The proposed model achieves an accuracy of 97%.

In (Raj et al. 2018a), YOLOv2 has been used to identify riders without helmets.

In (Afzal et al. 2021), authors have used the YOLO v3 model to detect the motorcycle and number plates. The proposed system consists of two deep learning models for the above tasks and achieves 89% and 92% accuracy for motorcycle and number plate detection.

In (Wu et al. 2019), authors have proposed a CNN-based model to perform multiple tasks like vehicle classification, helmet detection and mask detection.

In (Kathane et al. 2022), a deep transfer learning-based algorithm for vehicle detection is proposed for the KITTI dataset. The results show that the proposed system achieves better efficiency than the state-of-the-art approaches.

In (Rajalakshmi and Saravanan 2022), different deep learning-based approaches were proposed for helmet/no helmet detection. The results show that the proposed approach outperforms the existing approaches and achieves an accuracy of up to 90%.

In (Sridhar et al. 2022), authors have proposed license plate detection and recognition in a constrained environment where multiple license plates are detected.

In (Raj et al. 2018b), authors have proposed a Raspberry Pi and webcam-based Open Automatic License Plate Recognition system, which takes images of the license plates from the front and the back. The system can process almost ten frames per second. The OpenCV package of Python is used for detecting the license plate, and the Tesseract Optical Character Recognition method is used for reading the characters of the license plate. The limitation of this system is that it can detect one license plate at a time, and the model is trained for the license plate of Indian vehicles.

In (Wang et al. 2018), a comprehensive survey about vehicle and pedestrian detection is presented. The survey discussed the different approaches used in the literature for detecting vehicles and pedestrians. The survey concludes that deep learning has been proven to be most accurate in predicting and detecting traffic objects. In (Desai and Bartakke 2019), authors have used the deep learning models RetinaNet and ResNet 50 to detect motorcycle users without helmets. However, the limitation of the proposed model was the effect of environmental conditions and camera angles on the frame collection and image annotations. The authors of Silva and Jung (2018) proposed an InceptionV3 model for recognizing bike riders without helmets. According to the simulation findings, the suggested method achieves an accuracy of 81% for head helmet classification. For the validation set, the findings reveal an accuracy of 74%. The authors of Siebert and Lin (2019) suggested a curriculum-based learning system for identifying, detecting, and counting triple riders and helmet violations on unconstrained road scenarios. An amodal regressor approach is proposed for generating border boxes even for occluded riders in the data preprocessing phase. The suggested model has an accuracy of approximately 86% for motorbike and rider detection and 90% for helmet/no helmet detection. In (Rohith et al. 2019), authors have proposed a low-cost IOT-based smart parking solution to address the parking problem in extremely crowded cities. Implementing the proposed system effectively reduces the time spent in finding and reserving parking slots.

In (Yang et al. 2018), authors have proposed blockchain-enabled secure federated learning in-vehicle network systems to predict traffic flow in urban areas. The extensive simulation results of the MNIST dataset show that the proposed approach achieves 93% accuracy.

The authors of suggested a deep dual neural network with phase structure and attention mechanism for sentiment analysis from Chinese text in Goyal et al. (2022). First, in this approach, the Chinese short financial text corpus is built. Several ablation tests were carried out in the second phase, employing five techniques: Pinyin, segmentation, lexical

Table 1 Comparison of related work

S. No	References	Year	Methodology used	Dataset	Conclusion	Limitation
1	Ul Haq et al. (2022)	2018	CNN-based four models, Mobile Net, Google Net, VGG 16 and VGG 19, were used to detect helmets	Self-generated 493 images	Achieves an accuracy of 85%	Multiple objects are not detected
2	Mehmood et al. (2022)	2018	The proposed method used a Histogram of Oriented Gradient (HOG) and the CNN models for helmet and license plate detection	Self-generated	Achieved an accuracy of 95%	The traffic rule violation is not detected
3	Boonsirisumpun et al. (2018)	2019	Used YOLO v2 and YOLO v3 for the detection of helmets	Self-generated and open-source	Achieves the mAP of 95% and 98% for the helmet detection	Smaller objects are not detected
4	Yang (2022)	2020	Deep learning model Retina Net for helmet detection	Self-generated	Achieved an accuracy of 73%	Overlapped objects can not be detected
5	Siebert and Lin (2020)	2021	Deep Learning model for the detection of helmet rule violation using the Faster RCNN	Self-generated	Achieves an accuracy of 97%	Multiple objects are not detected in an image
6	Raj et al. (2018a)	2022	YOLO v2-based helmet rule violation detection	Self-generated and open-source	Achieves an accuracy of 95%	The speed of detection is low
7	Afzal et al. (2021)	2022	Three deep learning-based models were proposed for detecting motorcycles and license plates	Self-generated	Achieves the mAP as 89% and 92% for the motorcycle and the license plate detection	Traffic rule violation is not detected
8	Wu et al. (2019)	2022	The performance of multiple deep learning-based models for vehicle recognition and classification of image input is evaluated. To boost efficiency, the transfer learning method is used	Self-generated	Transfer learning improves the accuracy of the considered deep learning-based techniques	Overlapped objects cannot be detected
9	Kathane et al. (2022)	2018	Deep transfer learning-based method for vehicle detection	KITTI	The suggested approach's performance is compared to state-of-the-art transfer learning approaches	The processing speed is low
10	Rajalakshmi and Saravanan (2022)	2018	Different deep learning-based models were implemented to detect helmet rule violation/no violation	Self-generated	Achieves an accuracy of 90% for the real-time traffic	The efficiency is low
11	Sridhar et al. (2022)	2018	The performance of the proposed approach for license plate detection is compared for the different datasets	Open source dataset	Achieves higher accuracy for all the datasets	Multiple objects are not detected
12	Desai and Bartakke (2019)	2019	The HAAR and LBP feature extraction-based license plate recognition system is implemented	Self-generated	Achieves an accuracy of 95% for the HAAR method and 88% for the based method	The traffic rule violation is not detected

Table 1 (continued)

S. No	References	Year	Methodology used	Dataset	Conclusion	Limitation
13	Silva and Jung (2018)	2019	The automated solutions for helmet detection have been compared with the manual observation of the different images captured at different sites	Self-generated	The results achieve better performance than manual observation	The efficiency is low
14	Sieberta and Linb (2019)	2019	The cafe model-based helmet detection system is proposed. The inception v3 model is used for vehicle classification	Self-generated	Achieves the accuracy of 86% for the detection and 74% accuracy for the validation	The overlapped objects are not considered

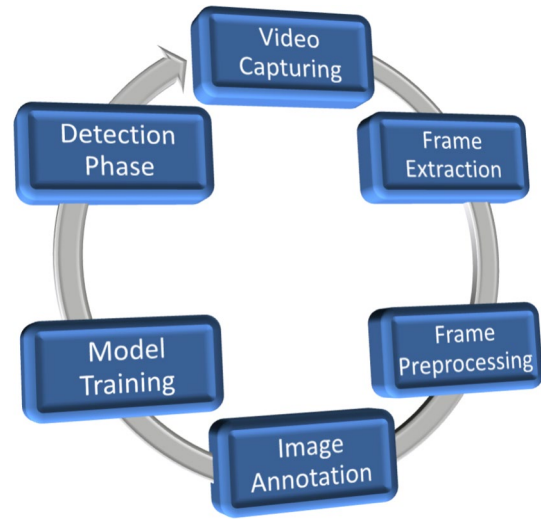


Fig. 4 Flow Chart of the Proposed Approach

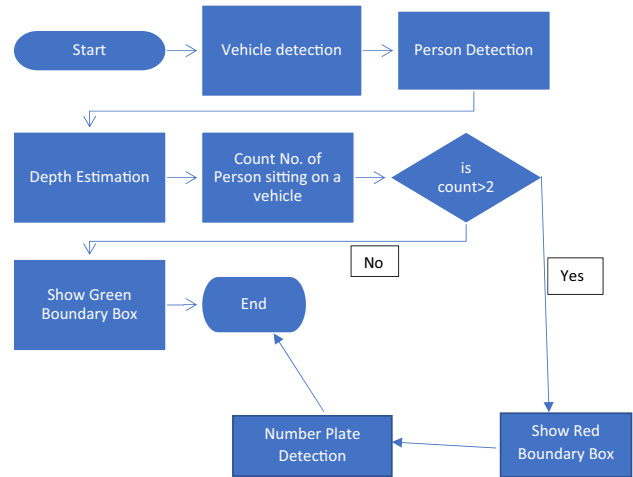


Fig. 5 Detection Phase

analysis, phrase structure, and attention mechanism. The sentence form and attention mechanism produce the best effects. The findings suggest that the proposed method outperforms the existing methods.

In (Gopal et al. 2019), authors have compared the different machine learning algorithms for big data analytics. The proposed approach consists of three phases: In the first phase, the data is collected from different social networking sites such as Twitter, Facebook, etc., using APIs. In the second phase, the different machine learning algorithms, supervised, unsupervised and reinforcement learning, are implemented and compared.

In (Thirunnavukkarasan et al. 2023), authors have proposed an intelligent anomaly detection framework for detecting anomalous activities in cyber-physical systems. The system consists of two processes; the first process is

the preprocessing of data through transformation and filtering operations. In the second process, the Gaussian Mixture Model (GMM) uses the Kalman filter-based Deep CNN model to detect anomalous activities in cyber-physical systems.

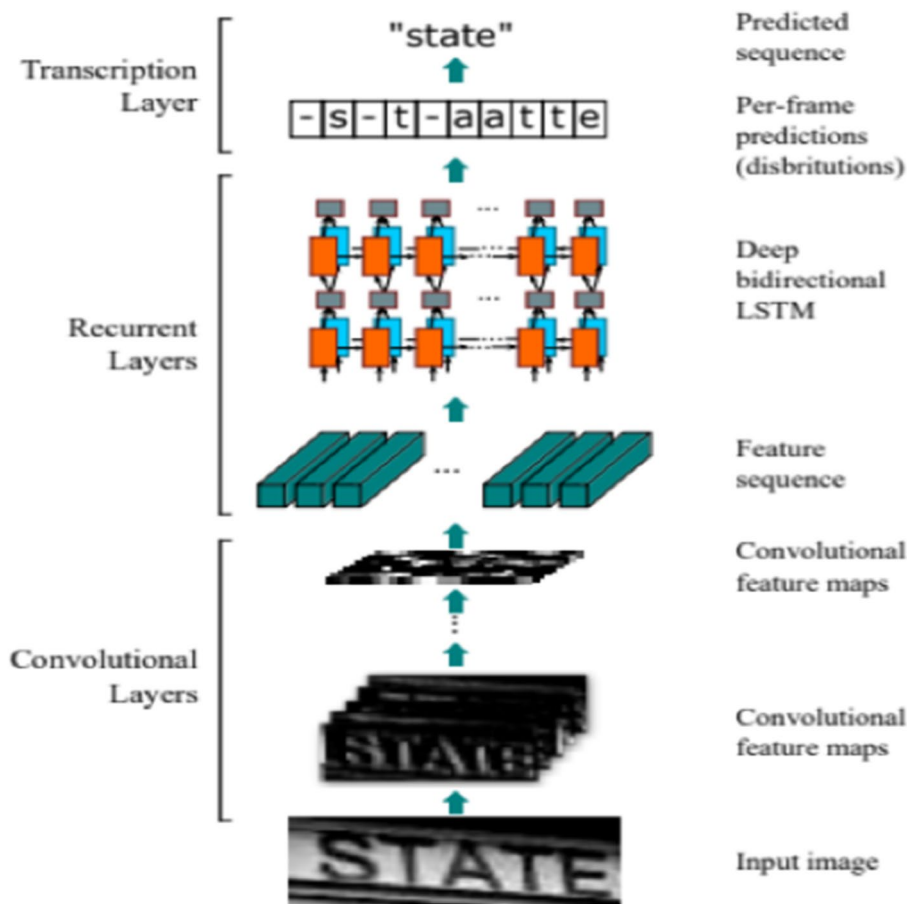
The authors of Rao et al. (2021) presented an ambient intelligent technique for intrusion detection in a smart house setting. The suggested approach is divided into two phases: the first is the learning phase, which employs the reinforcement-based learning algorithm. The Deep Q Networks are used in the second phase to identify and classify. The suggested approach’s performance is examined for four publicly available datasets, and the findings reveal that it outperforms previous approaches in terms of accuracy, precision, and recall.

The summary of the literature review and related works in the field of traffic surveillance have been summarized in Table 1.

Based on the literature review, the following research gaps are identified:

- i. There is tremendous work done for the detection of vehicles and riders, detection of riders with/without helmets. But the other rule violations should also be addressed, for example, the detection of riders with triple rider and overspeed vehicles.
- ii. There needs to be more suitable datasets that can be applied to determine vehicles violating the traffic rules.
- iii. There needs to be more work in real-time video surveillance.
- iv. The dataset’s quality also hinders the detection of traffic rule violations.
- v. The objects in the images may be occluded so that the detection accuracy may be low.

To address these issues, we propose automatic surveillance, which can automatically detect the two-wheeler with



CONNECTIONIST TEMPORAL CLASSIFICATION

Fig. 6 CTC RNN Algorithm



Fig. 7 Output of the ANPR

three passengers. Due to the lack of a dataset for the task, the self-generated dataset is used to implement the proposed system. The current YOLO v8 model is utilized to train the suggested model, and riders who violate traffic rules are detected. The results demonstrate the effectiveness of the proposed method.

4 Proposed system

The complete model of the proposed system is shown in Fig. 4, which consists of three sub-systems: 1. Identification of vehicle and rider and 2. Classification of two-wheelers as violator or no violator. 3. Automatic Number Plate Recognition.

The traffic rule violation detection system comprises the first six steps, and the next two steps are used in the license plate detection and recognition system.

Steps

1. *Video Capturing* The first step in the proposed system is to capture real-time video. For this purpose, the edge-enabled surveillance cameras are placed at the intersection points at the main gate and basement parking.
2. *Frame extraction* The second step in the proposed system is extracting the frames from the recorded video, which will be provided as input to the deep learning model. The frames from the recorded videos have been captured at 10 frames per second using an open CV package from Python.
3. *Frame Preprocessing* After the frames have been extracted from the video streaming, some preprocessing is done to remove the redundant and irrelevant frames. Out of 650 images, 622 images were considered relevant.
4. *Image annotation* The next step in the proposed model is to annotate the image into different classes that we want to detect or track from the image. The labelling package is used to annotate the image, which labels dif-

Table 2 Implementation details

Model	YOLOv8.0.34
Python	3.8.10
Pytorch	1.13.1+cu116
CPU	2 (Tesla T4, 15110MiB)
RAM	12.7GB
Hard disk	78.2GB

Table 3 Description of dataset

Dataset	Images	Triple rider	Non triple-rider
Training	540	120	420
Test	31	10	21
Validation	51	20	31

ferent objects in an image. In our system, we are interested in three classes of objects: motorcycles, triple riders and license plates. After the image annotation, the information about the image is stored in the YOLO format, which is a text document form having the following fields: [Box centre offset (B_x), Box centre offset (B_y), Box width (B_w), Box height (B_h), Object Score (S), Probability of class A (P_a), Probability of class B (P_b), Probability of class N (P_n)].

5. *Model Training*- After the data set has been prepared, the classification model is developed using the YOLOv8 to classify the two-wheeler as either violator or non-violator.
6. *Detection and Classification Phase* The test set images are used to determine the model's efficiency in this phase. The best weights obtained by the training process are used for detection purposes. The trained model is used to detect objects in three classes: motorcycle, triple rider and license plate. If any two-wheeler has triple riders, it will be shown as a red boundary box surrounding

them. Otherwise, the green boundary box is shown to detect the rider and the motorcycle. The detailed process of the detection phase is shown in Fig. 5

a. Non-Max suppression for Multiple object detection
Non-max suppression algorithm is used to detect individual objects of the same kind.

b. Depth Estimation algorithm

The depth estimation algorithm is used to distinguish the objects that are near or far from the camera. The application of the depth estimation algorithm generates the disparity map and disparity values for the pixels to determine the objects' depth in the image. The disparity value for a pixel is calculated using Eq. 1

$$d = w * s \quad (1)$$

where d , w and s represent the disparity values, width and scaling factor, respectively.

The depth of the object is calculated using the Eq. 2.

$$\text{Depth} = (\text{focal} * \text{baseline})/d \quad (2)$$

where $focal$ represents the distance to the image and the $baseline$ represents the difference in the camera positions.

7. Automatic number plate recognition

The two-wheeler's license plate should be captured and analyzed for the details of the traffic law violation. The license plate image is separated from the input image. The connectionist temporal classification technique is then used to detect text from a cropped license plate image. Figure 6 depicts the process of the connectionist temporal classification algorithm. The CTC RNN model has three layers: the convolutional layer, which extracts relevant features from the input image; the recurrent layer, which consists of multiple Deep bidirectional LSTM and is used to learn the characters of the license



Fig. 8 Sample image **a** Triple Rider **b** Non Triple Rider



Fig. 9 Results of the detection **a** Detection of vehicle and **b** Detection of person

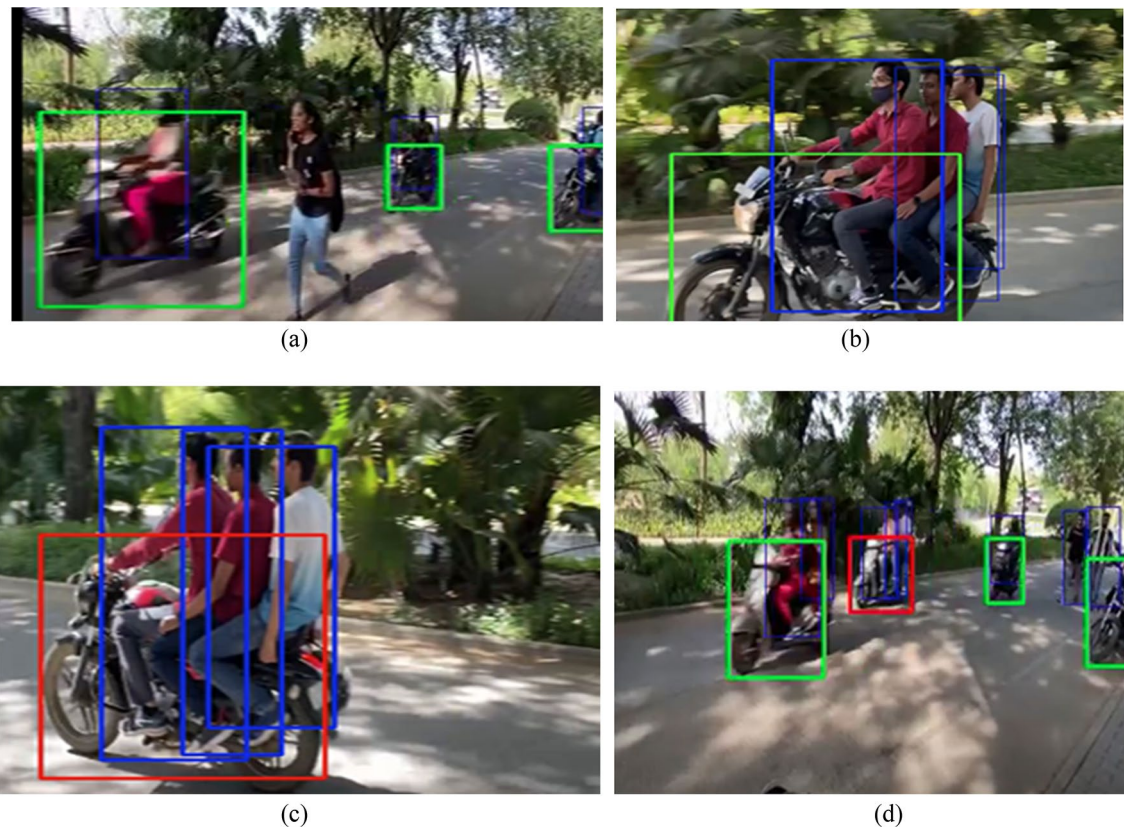


Fig. 10 a TP: Rider is Safe and Model also Predicted Safe/No Violation b. FP: Rider Violated the Rule, but Model predicted Safe c. TN: The Rider Violated the Rule, and the Model also detected Violation d. FN: The rider is Safe, but Model predicted a Violation

plate; and the transcription layer, which predicts the sequence of the characters of the license plate.

The output of the license plate recognition is shown in the Fig. 7.

5 Performance evaluation

The efficacy of the object detection model is determined using the parameters as follows:

- i. *Intersection over Union (IOU)*- It is the most common performance measure to determine the localization accuracy and the localization errors in the object detection models. The boundary box may overlap in the case of multiple object detection of the same class, so this metric is useful in such cases.

First, the intersection area of the two boxes for the same object is determined to determine the IOU between the ground truth boundary boxes and the predicted boundary boxes. Then, the total area covered by the boundary boxes is determined, known as the Union and the overlap area is defined as the intersection. The ratio of Intersection

divided by the Union provides the ratio of overlap to the total area, which gives a good measure of how accurately the predicted boundary box detects the actual object.

- ii. *Average Precision (AP)*- The weighted average of precision at each threshold level is average. Mean Average Precision (MAP) is the average of the Average Precision for each class. The following are the steps for calculating the MAP:
 - a. Determine the prediction level of an object belonging to a class using the object detection model.
 - b. Determine the class labels from the predictions.
 - c. Determine the confusion matrix for the model depending on the parameters True Positive, False Positive, True Negative and False Negative.
 - d. Determine the Precision and Recall measures for the model.
 - e. Determine the area under the curve of Precision vs. Recall.
 - f. Determine the average precision.

After performing these steps, the MAP is determined as the average precision for each class and the average for the class count. The formula to calculate the MAP is as follows:

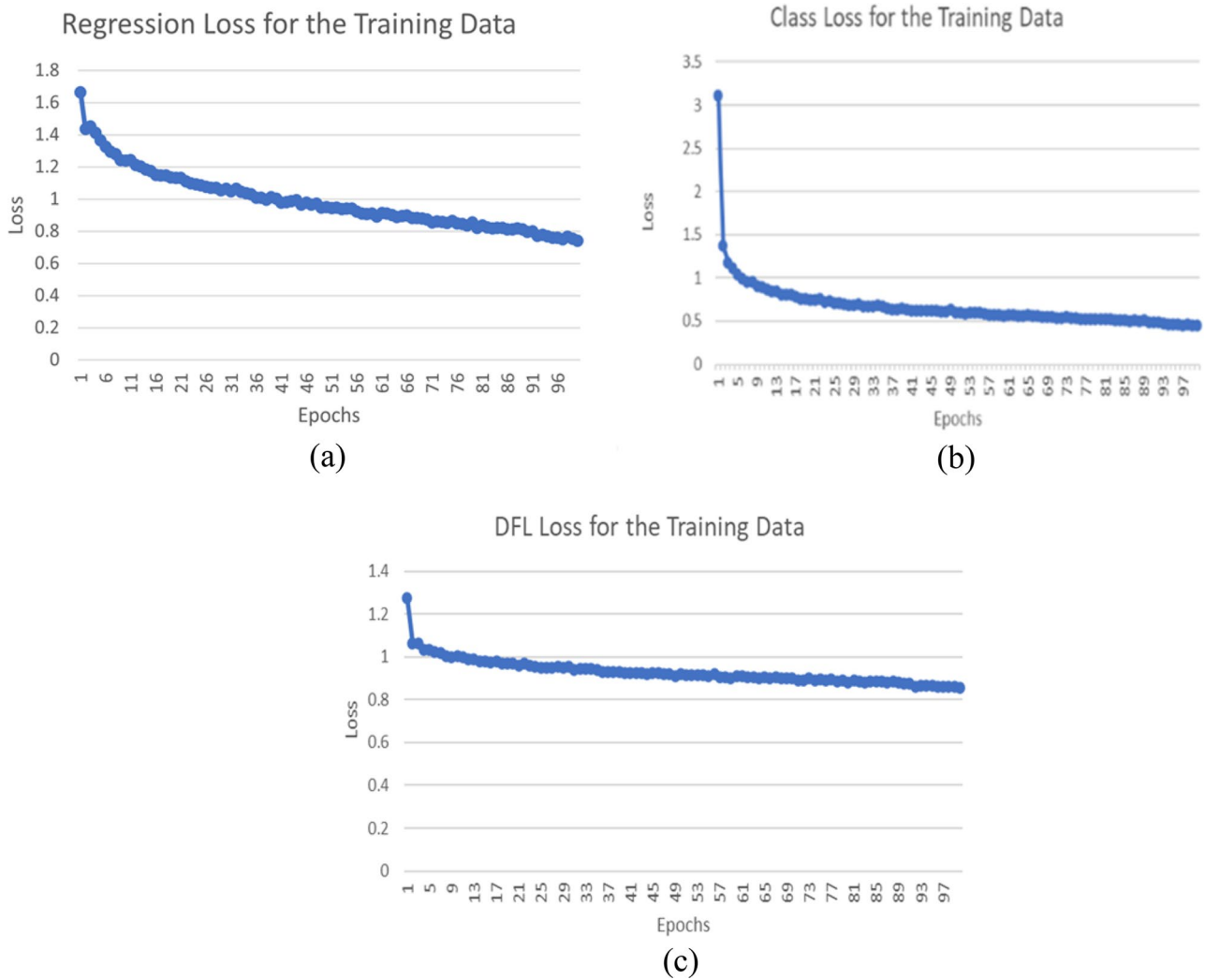


Fig. 11 Loss for the Training Data a. Regression Loss b. Class Loss c. DFL Loss

$$mAp = \frac{1}{n} \sum_{i=1}^n AP_i \tag{1}$$

5.1 Simulation settings

The implementation of the proposed model is done as per the details in Table 2.

5.2 Results and discussion

To develop the model, we considered a dataset of 622 photos separated into the following subsets: 540 images in the training subset, 31 in the test set, and 51 in the validation set. Table 3 contains a description of the dataset.

The three objects considered for the identification and detection tasks are motorcycles, license plates and triple riders. The sample image for the triple rider and triple rider are shown in the Fig. 8a and b.

The model is trained using the YOLOv8 model for object detection for batch size 16 and number of epochs as 100. The training was done in 1.082 h. After training the model, the best and last weights are determined. The result of the detection of the triple rider is shown in Fig. 9a and b.

The results of the traffic rule violation are shown in Fig. 10a–d.

Training results

The proposed model for the detection of triple riders is being developed and trained for 100 epochs

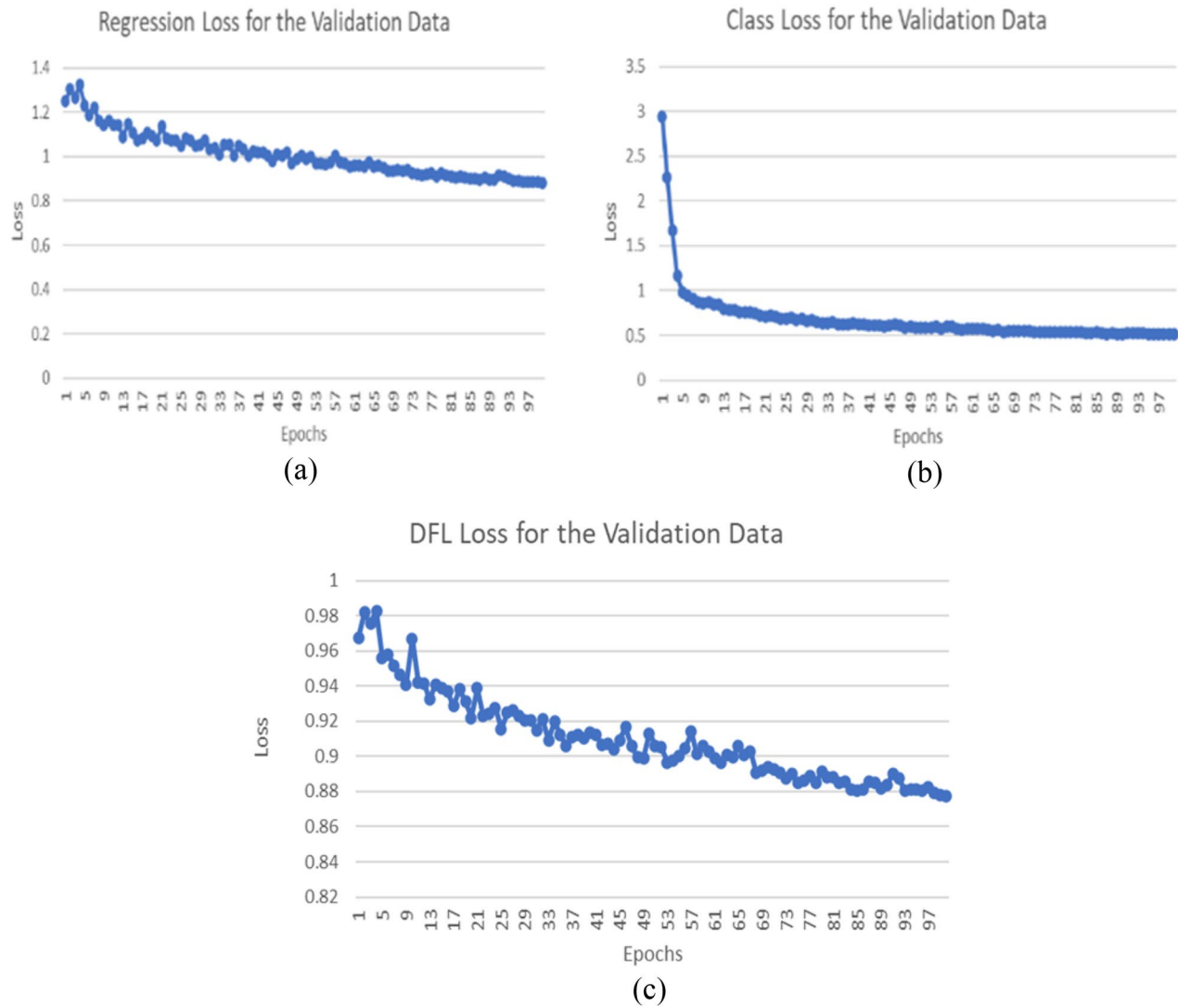


Fig. 12 Loss for the Validation Data a. Regression Loss b. Class Loss c. DFL Loss

1. *Loss*

The binary cross entropy function determines the loss in the training phase. The regression loss gives accuracy in detecting the boundary box concerning the ground truth. The class loss represents the accuracy of detecting the objects of different classes. The distributed focal loss represents the optimised distribution of boundary box boundaries. The results of the losses for the training dataset are shown in Fig. 11a–c. It can be observed that the loss value decreases with the increasing value of epochs.

The results of the losses for the training dataset are shown in Fig. 12a–c. It can be observed that the loss value is decreased with the increasing value of epochs.

The results shown in Fig. 13a–c represent the variation of the learning rate concerning the epochs for the three classes. The result shows that the learning rate is decreased concerning the increasing number of epochs.

2. *Confusion matrix*

The confusion matrix represents the performance of the proposed model. It consists of four fields, as shown in Table 4:

TP means true positive, FP means False Positive, FN means False Negative and TN means True Negative. TP represents the count of instances when the riders are correctly classified as non-violator, FP represents the count

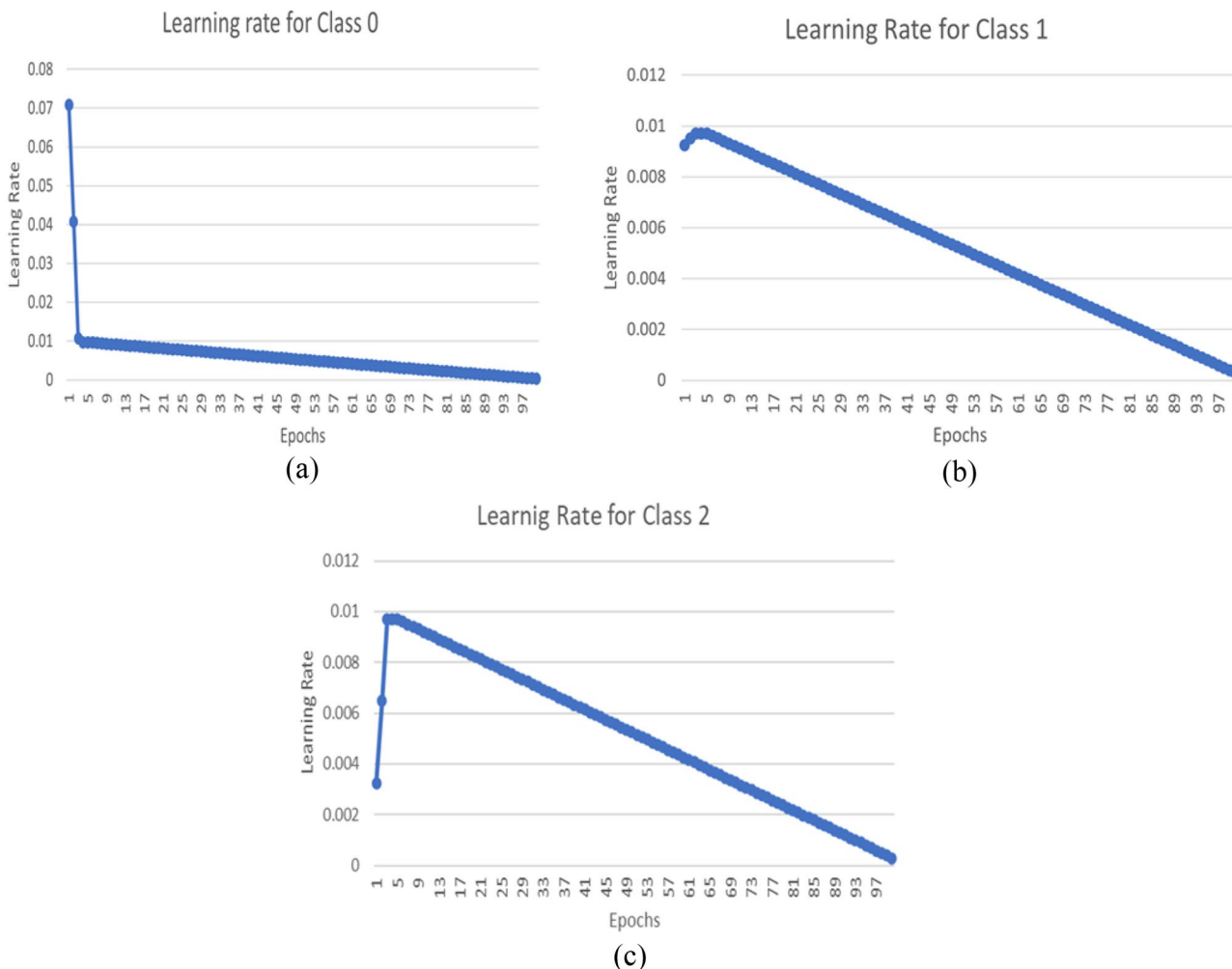


Fig. 13 Learning Rate a. Class 0 b. Class 1 c. Class 3

Table 4 Confusion matrix

Predicted values	Actual values	
	Positive-1	Negative-0
Positive-1	TP	FP
Negative- 0	FN	TN

of instances when the riders are wrongly classified as non-violator, FN represents the count of instances when the non-violator riders are wrongly classified as violators, and TN represents the count of instances when the violator riders are classified as violator. Figure 14 represents the confusion matrix. The count of instances correctly classified is 94%, 96% and 97% for the triple rider, license plate detection and motorcycle detection, respectively.

The results in Figs. 15, 16, 17, 18 show the performance metrics for the proposed model. It can be seen from the results shown in Fig. 15 that the precision of the model is increased up to 90% with the increasing number of epochs. The higher precision value indicates that our model is returning more relevant than irrelevant results. It can be seen from the results shown in Fig. 16 that the recall of the model is increased up to above 80% with the increasing number of epochs. The higher recall value indicates that most of the relevant results are being returned by our model. Figures 17 and 18 represent the Mean Average Precision for the proposed model.

3. Recall Confidence Curve

The results of recall and confidence are shown in Fig. 19. It can be concluded from the results that there is

Fig. 14 Confusion Matrix for the proposed approach

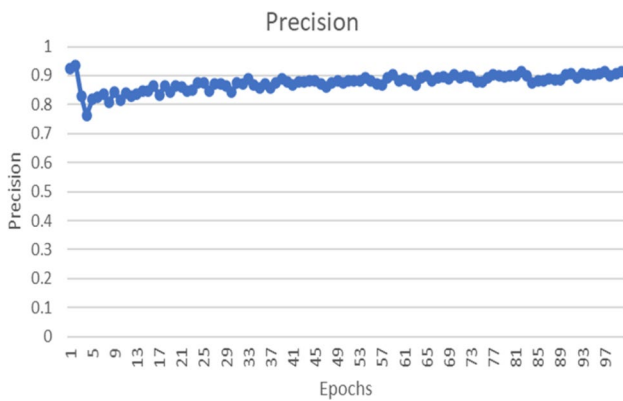
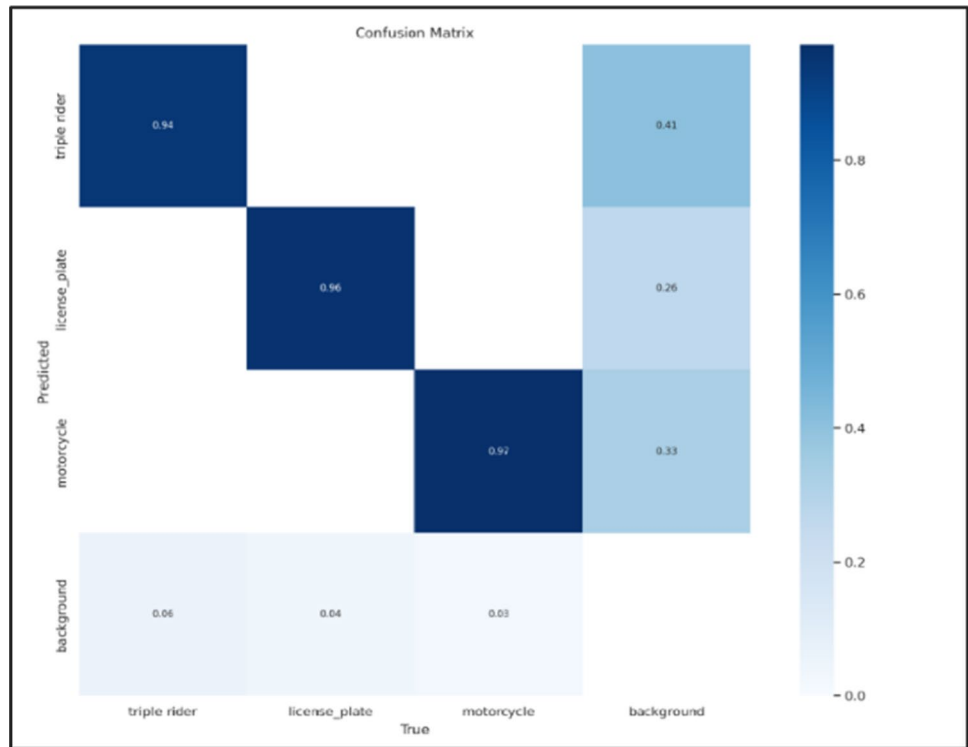


Fig. 15 Precision

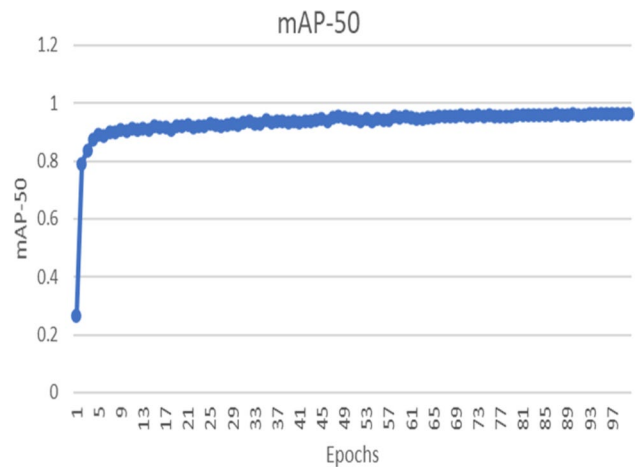


Fig. 17 mAP-50

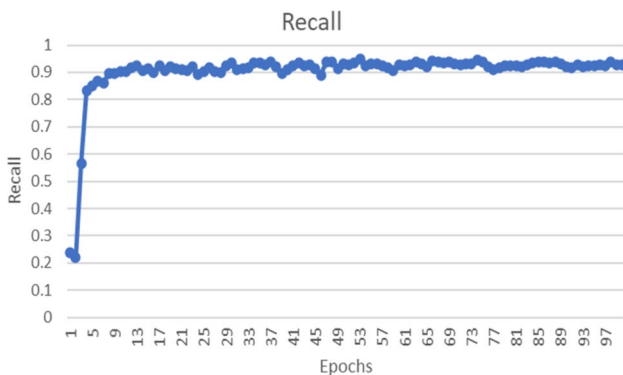


Fig. 16 Recall

a high area under the curve recall and confidence, which means the recall and confidence value is high. The high recall value indicates the low false negative rate. The increased confidence indicates high confidence in the importance of the results obtained for the proposed model.

4. Precision-Recall Curve

The results of precision and recall are shown in Fig. 20. It can be concluded from the results that there is a high area under both curves, which means the model has higher

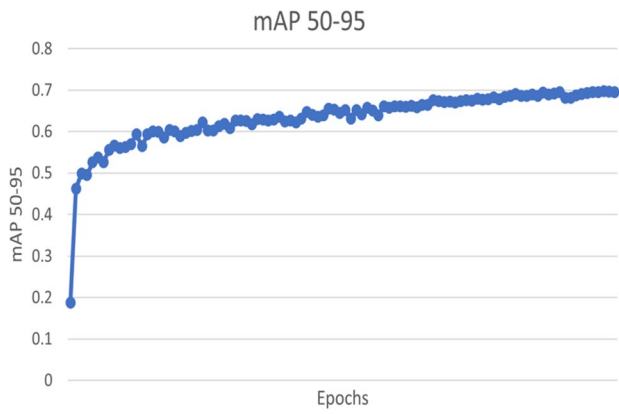


Fig. 18 Map 50–95

precision and recall values. The higher precision indicates that our model is returning more relevant than irrelevant results.

5. F1-Confidence Curve

The F1 score and confidence results are shown in Fig. 21. The F1 score is used to determine the optimal point where we can balance the precision and recall value. A higher F1 score above 80% can be achieved for a wider range of confidence values for all three tasks.

Fig. 19 Recall confidence curve

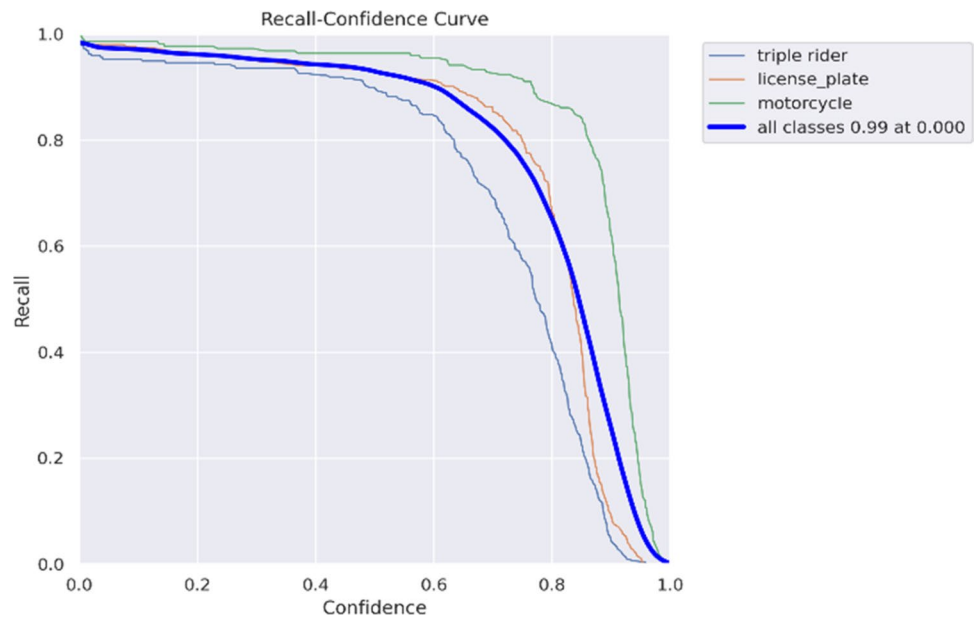


Fig. 20 Precision and Recall curve

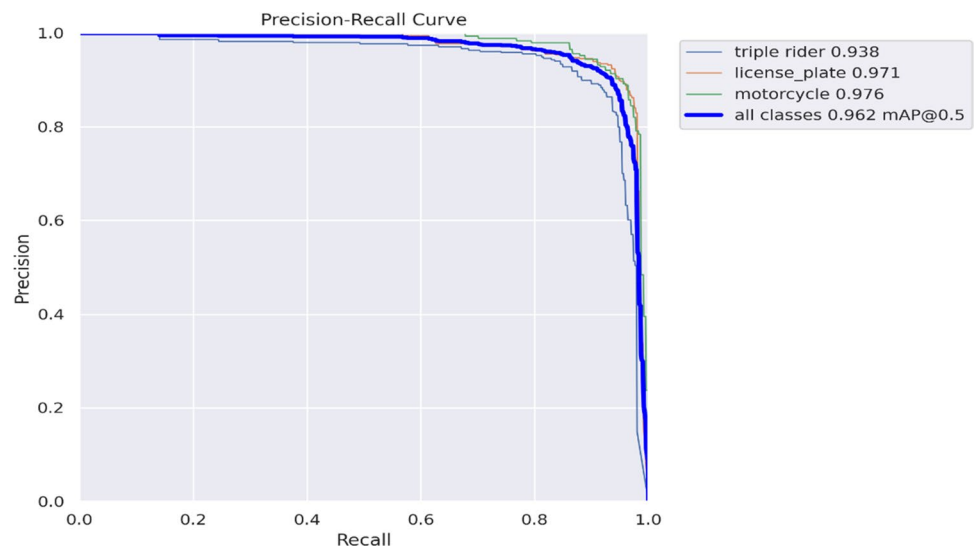


Fig. 21 F1 Score and confidence curve

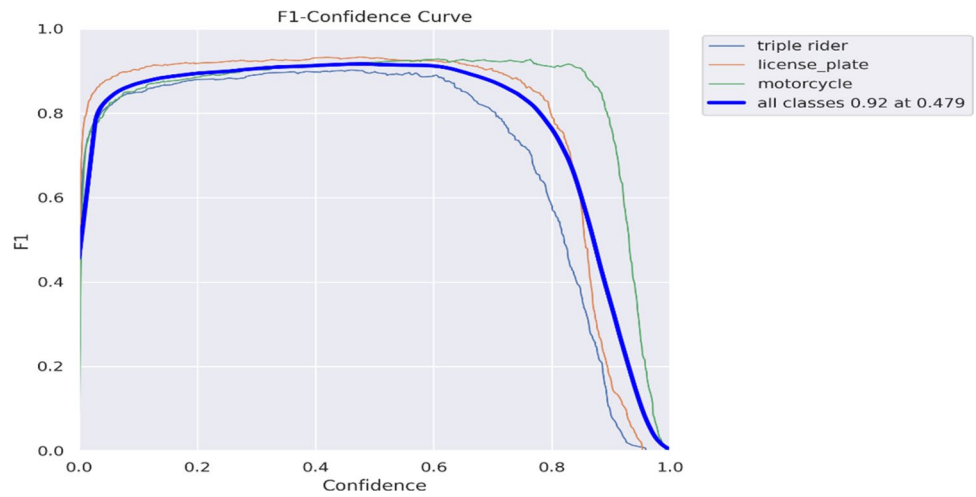


Fig. 22 Precision and confidence curve

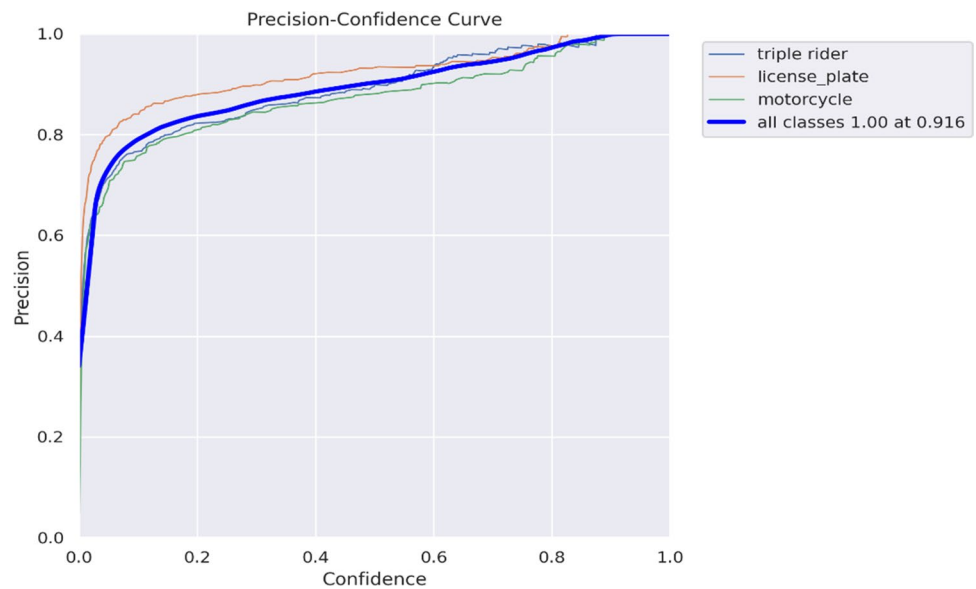


Table 5 Summary of the model

Parameter	Value
No. of layers	168
Parameters	3,006,233
Gradients	0
GFLOPs	8.1

Table 6 Simulation time in different phases

Task	Process	Time (ms)
Training	Pre-process	3.5
	Inference	2.8
	Post Process per image	2.7
Prediction	Pre-process	0.5
	Inference	13.5
	Post Process per image	1.9
Validation	Pre-process	3.9
	Inference	6.2
	Post Process per image	2.5

6. Precision-Confidence curve

The results of precision and confidence are shown in Fig. 22. It can be concluded from the results that a higher confidence level will result in higher precision for all three tasks.

After training the model and determining the best weights, the model is validated for the 51 validation test images. The model summary is in Table 5.

The time taken for the model in the prediction and validation phases is summarised in Table 6. The results show that the model takes a higher simulation time during the

Table 7 The performance measure for the validation dataset

Class	Box (Precision)	Recall	mAP50	mAP50-90
All	0.899	0.938	0.962	0.697
Triple rider	0.889	0.914	0.938	0.578
License plate	0.929	0.935	0.971	0.677
motorcycle	0.878	0.964	0.976	0.836

training and validation phases, but the predictions can be made in less time.

The result of the validation dataset is in Table 7

It is observed in the table that the approach obtained higher precision and recall values for the three tasks. The efficiency of the model is evaluated for the two cases:

- This* implies that the Intersection of Union (IOU) measure threshold is considered as 0.50.
- mAP50-90*- This implies that the average precision is considered for the IOU threshold from 50 to 90 with the increment of 5

The mean average precision is also higher for all the tasks. This shows the superiority of the proposed model.

6 Limitations

Although the proposed system achieves significant performance, there are some limitations of the proposed approach as well:

- The proposed model is effective for low-traffic areas.
- The proximity of the people with the vehicle may reduce the accuracy.
- The accuracy and precision of the proposed system depend on the angle of the placement of the camera and imaging quality.
- The system is implemented for the college premises under the above constraints.

7 Conclusion and future scope

The paper presented a classification model based on YOLOv8 for detecting triple riders on a two-wheeler. The model was trained to determine the best weights for the detection. Then, the model is tested for the validation test of 51 images with the accuracy level of 91%, 94% and 96% for triple rider detection, license plate detection and motorcycle detection. The system can help regularise the traffic rule implementation with more accuracy and less effort than manual monitoring methods.

Detecting overspeed through the proposed model is part of the future work. Integrating the proposed system for the automatic alert message in case of traffic rule violation is also part of the future scope.

Funding This work is funded by the Institute of Technology, Nirma University, Ahmedabad. The authors are grateful for the support throughout this work.

Declarations

Conflict of interest The authors declare that there is no competing interest.

Human or animal rights The authors declare that prior permission is taken.

References

- Afzal A, Draz HU, Khan MZ, and Khan MUG (2021) Automatic helmet violation detection of motorcyclists from surveillance videos using deep learning approaches of computer vision. In Proceedings of the 2021 international conference on artificial intelligence (ICAI), (pp 252–257)
- Bochkovskiy A, Wang C-Y, Mark Liao H-Y (2020) YOLOv4: optimal speed and accuracy of object detection. arXiv:2004.10934
- Boonsirisumpun N, Puarungroj W, and Wairotchanaphuttha P (2018) Automatic detector for bikers with no helmet using deep learning. In Proceedings of the 2018 22nd international computer science and engineering conference (ICSEC) (pp 1–4)
- Chen C, Liu B, Wan S, Qiao P, Pei Q (2021) An edge traffic flow detection scheme based on deep learning in an intelligent transportation system. *IEEE Trans Intell Transp Syst* 22(3):1840–1852
- Dai J, Li Y, He K, Sun J (2016) R-can: object detection via region-based fully convolutional networks. *Adv Neural Inf Process Syst* 29
- Dakwala A, Lavingia K (2016) A machine learning approach to improve the efficiency of fake websites detection techniques. *Int J Comput Sci Commun (IJCSC)* 7:236–243
- Girish G Desai, Prashant P Bartakke (2019) Real-time implementation of indian license plate recognition system
- Girshick R (2015) Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision (pp 1440–1448)
- Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (pp 580–587)
- Gopal DG, Jerlin MA, Abirami M (2019) A smart parking system using IoT. *World Rev Entrep Manag Sustain Dev* 15(3):335–345
- Goyal A, Agarwal D, Subramanian A, Jawahar CV, Sarvadevabhatla RK, Saluja R (2022) Detecting, tracking and counting motorcycle rider traffic violations on unconstrained roads. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp 4303–4312)
- He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision (pp 2961–2969)
- <https://github.com/ultralytics/ultralytics>
- https://morth.nic.in/sites/default/files/RA_2021_Compressed.pdf

- <https://www.financialexpress.com/express-mobility/as-road-accidents-continue-to-rise-a-look-at-the-key-safety-norms-mandatory-in-india-now/2932079/>
- <https://www.thehindubusinessline.com/data-stories/data-focus/overspeeding-accounts-for-60-of-the-road-accidents-in-india/article65852932.ece>.
- <https://www.v7labs.com/blog/yolo-object-detection>
- Jocher G (2020, June). YOLOv5. (GitHub)
- Kaffash S, Nguyen AT, Zhu J (2021) Big data algorithms and applications in intelligent transportation system: a review and bibliometric analysis. *Int J Prod Econ* 231:107868
- Kathane M, Abhang S, Jadhavar A, Joshi AD, Sawant ST (2022) Traffic rule violation detection system: deep learning approach *Advanced Machine Intelligence and Signal Processing*. Springer, New York, pp 191–201
- Lavingia K, Vaja M, Chaturvedi P and Lavingia A (2022) Machine learning based approach for traffic rule violation detection. In: 2022 IEEE 7th International conference on recent advances and innovations in engineering (ICRAIE-2022), Mangalore, India, (pp 244–249)
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) Ssd: single shot multibox detector. In: *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* (pp 21–37). Springer International Publishing.
- Lohia A, Kadam KD, Joshi RR, Bongale AM (2021) Bibliometric analysis of one-stage & two-stage object detection. *Libr Philos Pract* 4910:34
- Mehmood M, Shahzad A, Zafar B, Shabbir A, Ali N (2022) Remote sensing image classification: a comprehensive review and applications. *Math Prob Eng* 2022:1–24
- Nagarajan SM, Bashir AK, Mahapatra RP, Al-Numay MS (2022) IADF-CPS: intelligent anomaly detection framework towards cyber physical systems. *Comput Commun (Elsevier)* 188(15):81–89
- Pooya Sagharichi H and Shakeri M (2016) License plate automatic recognition based on edge detection. In: 2016 artificial intelligence and robotics (IRANOPEN) (pp 170–174)
- Punia SK, Kumar M, Stephan T, Patan R (2021) Performance analysis of machine learning algorithms for big data classification: ml and ai-based algorithms for big data analysis. *Int J E-Health Med Commun (IJEHMC-IGI)* 12(4):60–75
- Raj KD, Chairat A, Timtong V, Dailey MN, and Ekpanyapong M (2018a) Helmet violation processing using deep learning. In *Proceedings of the 2018a international workshop on advanced image technology*, (pp1–4)
- Raj KD, Chairat A, Timtong V, Dailey MN, Ekpanyapong M (2018b) Helmet violation processing using deep learning. In: 2018b International workshop on advanced image technology. *IEEE IWAIT*, (pp 1–4)
- Rajalakshmi N and Saravanan K (2022) Traffic violation investigation using transfer learning. In *Proceedings of the 2022 international conference on electronic systems and intelligent computing (ICESIC)*, (pp 286–292)
- Ramana TV, Thirunavukkarasan M, Mohammed AS, Nagarajan SM (2022) Ambient intelligence approach: internet of things based decision performance analysis for intrusion detection. *Comput Commun (Elsevier)* 195(1):315–322
- Rao D, Huang S, Jiang Z, Patan R (2021) A dual deep neural network with phrase structure and attention mechanism for sentiment analysis. *Neural Comput Appl (Springer)* 33:11297–11308
- Redmon J, Farhadi A (2016) YOLO9000: better, faster, stronger. *arXiv preprint [arXiv:1612.08242](https://arxiv.org/abs/1612.08242)*.
- Redmon, J., & Farhadi, A. (2018). Yolov3: an incremental improvement. *arXiv preprint [arXiv:1804.02767](https://arxiv.org/abs/1804.02767)*.
- Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp 779–788)
- Ren S, He K, Girshick R, Sun J (2015) Faster r-CNN: towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst* 91–99
- Rohith CA, SA Nair, PS Nair, S Alphonsa (2019) An efficient helmet detection for MVD using Deep learning. In: *Proceedings of the third international conference on trends in electronics and informatics*
- Siebert FW, Lin H (2020) Detecting motorcycle helmet use with deep learning. *Accid Anal Prev* 134:105319
- Sieberta FW, Linb H (2019) Detecting motorcycle helmet use with deep learning. *Accid Anal Prev* 134:105319
- Silva SM, Jung CR (2018) License plate detection and recognition in unconstrained scenarios. In: *European conference on computer vision*. Springer, pp 593–609.
- Sridhar P, Jagadeeswari M, Sri SH, Akshaya N, and Haritha J (2022) Helmet violation detection using YOLO v2 deep learning framework. In: *Proceedings of the 2022 6th international conference on trends in electronics and informatics (ICOEI)*, (pp 1207–1212)
- Sun P, Boukerche A (2021) AI-Assisted data dissemination methods for supporting intelligent transportation systems. *Internet Technol Lett* 4(1):e169
- Thirunavukkarasan M, Irfanullah Amanullah S, Vignesh T, Sivaraman A (2023) An integrated security approach for vehicular networks in smart cities. *Trans Emerg Telecommun Technol-Wiley*. <https://doi.org/10.1002/ett.4757>
- Ul Haq HB, Asif M, Ahmad MB, Ashraf R, Mahmood T (2022) An effective video summarization framework based on the object of interest using deep learning. *Math Prob Eng* 2022:1–25
- Wang H, Yu Y, Cai Y, Chen L, Chen X (2018) A vehicle recognition algorithm based on deep transfer learning with a multiple feature subspace distribution. *Sensors* 18(12):4109
- Wang CY, Bochkovskiy A, Liao HYM (2022) YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint [arXiv:2207.02696](https://arxiv.org/abs/2207.02696)*
- Wu F, Jin G, Gao M, Zhiwei H and Yang Y (2019) Helmet detection based on improved YOLO V3 deep model. In: *Proceedings of the 2019 IEEE 16th international conference on networking, sensing and control (ICNSC)*, (pp 363–368)
- Xiao Z, Xu X, Xing H, Song F, Wang X, Zhao B (2021a) A federated learning system with enhanced feature extraction for human activity recognition. *Knowl-Based Syst* 229:107338
- Xiao Z, Xu X, Xing H, Luo S, Dai P, Zhan D (2021b) RTFN: a robust temporal feature network for time series classification. *Inf Sci* 571:65–86
- Xing H, Xiao Z, Qu R, Zhu Z, Zhao B (2022) An efficient federated distillation learning system for multitask time series classification. *IEEE Trans Instrum Meas* 71:1–12
- Yang M (2022) Research on vehicle automatic driving target perception technology based on improved MSRPN algorithm. *J Comput Cognit Eng* 1:147–151
- Yang Z, Li J, Li H (2018) Pedestrian detection and vehicle detection for autonomous vehicles. *IEEE Intell Veh Sympos (IV)*. <https://doi.org/10.1109/IVS.2018.8500642>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.