Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-10126-x

Detection of the triple riding and speed violation on two-wheelers using deep learning algorithms

Nikhil Chakravarthy Mallela[1] · **Rohit Volety**[2] · **Srinivasa Perumal R.**[1] · **Nadesh R. K.**[1]

Received: 15 February 2020 / Revised: 14 August 2020 / Accepted: 19 October 2020 /
Published online: 31 October 2020

Abstract

To curb the accident rate and traffic levels, strict implementation of the rules and continuous monitoring of the traffic is mandatory. Traffic Rule Violation Monitoring System ensures that the rules are followed strictly and it reduces the human effort. The main objective of this work is to identify the Triple Riding. To detect the triple riders, the deep learning framework darknet is used, which in turn uses a type of convolutional neural networks i.e. Deconvolutional neural network-based YOLO (You Only Look Once) algorithm for detection of the number of persons riding a bike, the system classifies the vehicle as to the rule-breach vehicle or not. The junctions acting as the data collections center, collects the data. The image of the vehicle classified as the rule-breach is stored along with the data such as vehicle manufacturing ID and vehicle speed transferred at the particular frame. The transfer of the data is facilitated using the GSM module and the NodeMCU deployed on the vehicle. The vehicle number will be verified with the transport office. To survive the lack of internet connectivity or low internet connectivity, the system is being equipped with the GSM module; else, the data related to the vehicle can be pulled by the development boards deployed at the junctions, acting them as the central part of the public internetwork deployed. This public internetwork acting the medium to pull the data from the vehicle to the central system. This is carried out using the concept of dynamic network configuration in NodeMCU. The use of Node MCU and the public network system makes the system much more viable, available and reliable. Thereby making the riders follow the rules properly and reducing irresponsible driving.

Keywords Triple riding · Object detection · YOLO · Speed violation · Deep learning

✉ Nikhil Chakravarthy Mallela
nikhilchakravarthy.mallela@gmail.com

Srinivasa Perumal R.
r.srinivasaperumal@vit.ac.in

1 School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

2 School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

1 Introduction

Road safety is the most important aspect of this automobile driven technological world. Considering the number of people taking road transport as the means to reach their destination, the number of people reaching the heavens instead of their safe home, increasing day-to-day. As per Indian government data, in 2017 alone, 1,47,913 people were killed in road accidents across India. One lakh forty-seven thousand nine hundred and thirteen dead bodies on Indian roads in just one year. This figure is 37.54% more than the total number of people killed in floods and heavy rains in the last 65 years in India. In the 17 years between 2001 and 2017, (the latest year for which official data on road accidents are available), a total of 20.42 lakh people lost their lives in road accidents and 82.30 lakh were injured. In total, India witnessed 79.10 lakh road accidents at an average of 9 crashes every 10 min. On account of the recent stats, the irresponsible driving of the two-wheelers or the heavy speeding of the four-wheelers is the major reason for the occurring accidents. These irresponsible drivers are making it hard for the drivers those follow the traffic rules. Just as the saying goes, the ueue has no part to play in "Queue", the prior rules do not affect the irresponsible drivers. The current increase in the fine/challan system might control these irresponsible drivers to an extent, but this is not a permanent solution that we can rely on. Monitoring the roads continuously for these kinds of irresponsible drivers is human effort consuming. The currently existing CCTV surveillance system can come to aid to some extent where the continuous monitoring of the system again makes it difficult. Hence, this CCTV surveillance system needs automation thereby, reduces the human effort in monitoring the traffic rule violators. This paper aims to provide the automation to the CCTV surveillance system, which will help in finding the riders, triple riding the vehicles.

2 Related work

In the last 10 years, researchers are trying to explore new ways to address the problems related to the traffic i.e. Bartlomiej Placzek et al. proposed a vision-based algorithm to detect the vehicle in the detection zones is important for the traffic control system [8]. Having this system to identify the high traffic prone zones to use more human effort to control the traffic and reducing the traffic rule violators is the objective of the author. The author used the linguistic variables and fuzzy sets to classify the input through several occurrences of a frame that background or vehicle in greyscale. The results verified through extensive testing in various conditions as well. Hu et al. used Histogram of oriented gradient features to detect the vehicles in recorded videos and classify the vehicles. Here in this paper, the concept of the people traveling on the vehicle has not been addressed [5]. In the case of detection of the vehicles moving on that particular junction, this algorithm can be implemented in the real world. Maharsh Desai et al proposed helmet detection system, alcohol detection system, fall out detection system to reduce the number of accidents. The authors used subtraction; Hough transform descriptors to detect the helmet. As it is always preferable to reduce the causalities before the detection of fall out, the triple riding which can lead to the fall out has not been addressed [2]. Kunal Dahiya et al. also proposed a method to detect the bike riders from video using subtraction, object segmentation which determines whether the bike rider is using the helmet or not [1]. This system acquired better accuracy of 93.80% with a processing time of 11.58 ms per frame in a surveillance video.

Vishnu et al. designed a framework for automatic rider detection without a helmet in the same surveillance videos. The authors [15] have used the adaptive background subtraction method on video frames to get moving objects, CNN to select the rider and to determine who is not wearing a helmet at a good rate. This best in being invariant to the illumination, poor video quality, etc. Hai Wang et al. formulated a model vehicle detection algorithm namely transfer learning method developed based on deep learning model. The authors [16] used a set of rules based on a couple of subspace characteristic distribution deep version with online switch getting to know. Sergio Montazzoli et al. proposed a complete deep learning algorithm ALPR system for unconstrained capturing situations. The authors [14] main idea is to introduce novel CNN, to make them capable of detecting, rectifying many distorted license plates. In this current world, people are smart enough to cover their license plates and sometimes to come on the road without license plates, this particular issue needs to be addressed. Dharma Raj KC et al. used image processing, a deep convolutional neural network for detecting the vehicle rider who are violating the rules such as helmet law [10]. It comprises vehicle detection, helmet recognition.

Using 2 stage models have always been very expensive to use and cannot be used in the embedded platform without having to sacrifice performance by compressing the model. When compared with the 2 stage target detection method, the 1 stage target detection methods which include the most popularly used YOLO, SSD, DSSD and RetinaNet. Detection has been increased in the newer series of smaller targets by the use of ResNET for extracting features and predicting multiscale.But this sacrifices the ability to detect medium targets. Yolov2 uses many things like high resolution classifiers,convolutional with anchor boxes,direct location prediction etc which has made it better and faster than yolov1. SSD has been considered to be slightly better than YOLOV1 method for detection of the small targets but very less than 2 stage method methods for detection of targets. In comparison with the SSD method, DSSD method uses a better basic network, ResNET, and deconvolution layers. We get a much better representational expression of the feature maps in the shallow layers because outfits skip connection. Also RetinaNet is a feature pyramid network based detection network. to solve the category imbalance which is actually being caused by excessive background, FPN relies on the focal loss at the end. RetinaNet has become one of the best detection methods because of its simplicity and powerful architecture [17].

The existing methods are not addressing the issue of triple riding, which our proposed work deals with the triple riding along with the vehicle tracking system. The summary of the related work is represented in Table 1.

3 Proposed model

The system is divided into 2 main subsystems setting up the environment, training, testing the model and getting the accurate coordinates of the vehicle comes along with the system, and to pull the data GSM module or the public network. The Proposed system of triple riding is illustrated in Fig. 1. The imposing challans or fines with an automated push of the data to the respective user account comes as an extension to the system.

1. Traffic Rule Violation Recognition System
2. Vehicle Tracking System

Table 1 Summary of related works

Year	Authors	Methodologies	Advantages	Disadvantages
2011	Placzek B	Linguistic Variables and Fuzzy Sets, Occupancy dependent actualization, Vision Sensors	It can be used to get accurate information in ambient lighting conditions, Weather, Camera Vibrations, Speed. Cost Effective	When the vehicle stops in the detection zone for a long period, Color Images.
2013	Youpan et al.	Haar Features, HOG, AdaBoost.	Accuracy is high for a large number of data Samples and Faster. No need to scan the whole image for Classifying. All Vehicles are detected and Classified. The error rate is minimized.	Need to identify a strong Classifier. Cost is too high.
2016	Desai et al.	Hough transform, Optical character Recognition, Background Subtraction	Reduce the number of deaths by intimating nearby hospitals. Detect helmet on public roads	Need clear images. The accuracy rate is low.
2016	Kunal et al.	Background Subtraction, Object Detection Method - HOG , SIFT, RBF and SVM	Good Classification Accuracy. It allows for low environmental Changes. It requires an alarming rate of less than 1%. 11.58 ms per frame.	Sudden Environmental Changes. Need Moving Objects.
2017	Vishnu et al.	Deep Learning, Convolutional Neural Networks	Efficiency is high Success rate of 92.87%. Low false alarm rate of provides good accuracy rate.	Very Close Vehicles pose a little difficult to get accurate results.
2018	Wang et al.	Subspace characteristic distribution technique, DBM	It provides accurate information while there is high variability in the video	Processing time is long. The number of subspaces are hard to determine.
2018	Silva et al.	Convolutional Neural Networks, Optical Character Recognition	The achieved success rate of the recognition of license plate is high. It can get the results from oblique images. Heuristic methods not required.	It was not defined for Motor Vehicles.

Table 1 (continued)

Year	Authors	Methodologies	Advantages	Disadvantages
2018	Raj et al.	Deep Learning, Convolutional Neural Networks,	Good accuracy Provided when rider won't wear anything. (97.52%)	If the rider wears a hat then the accuracy rate will reduce. If we won't use a perfect segmentation method it introduced errors that cannot be solved by the classifier.

3.1 Traffic rule violation recognition system

The system mainly uses the deconvolutional approach of the deep learning along with the deep learning framework Darknet and the object detection algorithm YOLOv3 performs the various functionalities like detection, recognition, and Identification, followed by classification.

The subsystem having the support of the image, video and live feed as the input, enables the system to process all kinds of data. Having the image and video input assessment as an extension, the live feed of the camera modules deployed at the junctions processes every single frame. Every single frame is subjected to the detection of the various objects. In our prototype model, the various objects that were subjected to the process of the training are the bike and the person. The two objects bike, and person are the most important aspects of the focus.

The detection process of the YOLO [12] model makes use of the target regression as a regression problem for the spatially separate target box and confidence.

1. The YOLO first divides the image into convolutions of size NxN like 13x13, and the size of each of the NxN cells depends on the size of the input.

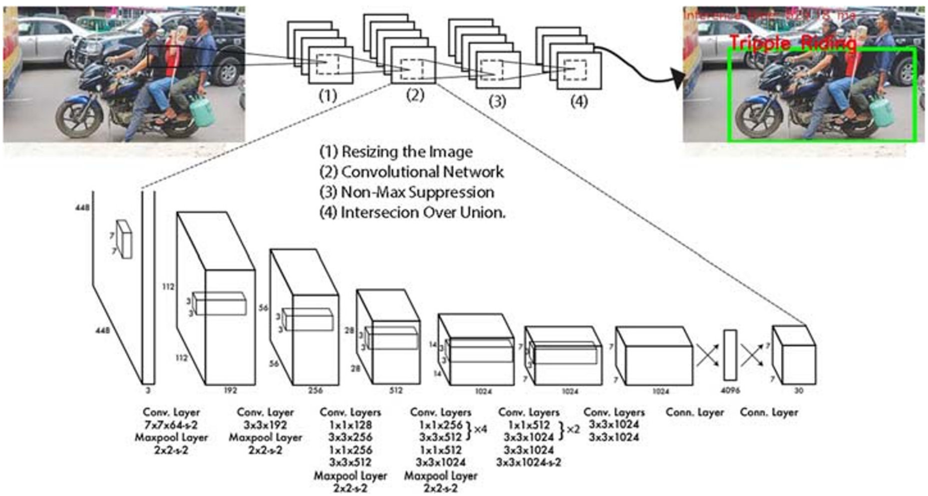

Fig. 1 Proposed model

Fig. 2 Test images for the triple riding

2. Each cell of these NxN cells is responsible for predicting the number of bounding boxes in the input image.
3. For every box in the input, the deconvolutional network predicts the confidence that the bounding box contains the object and the probability of the enclosed object being from one of the classes mentioned in the configuration file.
4. After that, it applies the Non-Max Suppression to remove the bounding boxes with less confidence value.
5. The processed images are subject to the Intersection over union function developed to find out the relation between the persons detected in the frame along with the motorbike. Thereby, marking the rectangular bounding box with triple riding as the label.

3.1.1 YOLOv3 (You Only Look Once)

YOLO is an incremental approach, in this paper the third version of the YOLO [12] is used. This comes with an incremental improvement from the initial YOLO. The YOLO algorithm forwards the image as a whole only once through the network. On the other hand, The Single Shot Detector (SSD) also takes the entire image at the same time, but YOLOv3 performs much faster, while achieving a better accuracy [13].

The YOLOv3 system has been trained on the COCO dataset [6] which was available on the internet, capable of detecting various classes. The following figures depict the working of the model, Figs. 2 and 3 are the test images for the triple riding whereas Fig. 4 is the test image for the non-triple riding.

Fig. 3 Test images for the triple riding

Fig. 4 Test images for the non-triple riding

3.1.2 Intersection over union

In general, Intersection over union (IOU) is an evaluation metric used to measure the accuracy of an object detector on a particular dataset. This is IOU is one of the major components even in the incremental approach, YOLO to eliminate the unwanted bounding boxes formed during the forward cycle of the YOLO algorithm on an image. The same IOU concept which was modified a bit to the customer use, to identify the Triple riding case in the input image. After the image is processed through all the network layers, the image is subjected to the calculation of the IOU over the classes motorbike and person. The detected classes motorbike and the person after the calculation of the IOU, the triple riding class can be identified. Using this evaluation metric IOU, the triple riding identification is carried forward [9]. The IOU is calculated by (1).

$$IOU = Areaofoverlap/AreaofUnion \tag{1}$$

To determine the (x, y)-coordinates of the intersection rectangle

$$\begin{aligned} x_A &= \max(boxA[0], boxB[0]) \\ y_A &= \max(boxA[1], boxB[1]) \\ x_B &= \min(boxA[2], boxB[2]) \\ y_B &= \min(boxA[3], boxB[3]) \end{aligned}$$

The above equations represent the calculation of the coordinates of the intersection rectangle using the bounding boxes coordinates as the arguments. boxA references the person detected in the frame wherein the boxB subjects to the bike detected. The area of intersection rectangle is calculated using (2)

$$interArea = \max(0, x_B - x_A + 1) * \max(0, y_B - y_A + 1) \tag{2}$$

The area of both the prediction and ground-truth rectangles is calculated using the subsequent equations represented below

$$\text{boxAArea} = (\text{boxA}[2] - \text{boxA}[0] + 1) * (\text{boxA}[3] - \text{boxA}[1] + 1)$$
$$\text{boxBArea} = (\text{boxB}[2] - \text{boxB}[0] + 1) * (\text{boxB}[3] - \text{boxB}[1] + 1)$$
$$\text{IOU} = \text{interArea}/\text{float}(\text{boxAArea} + \text{boxBArea} - \text{interArea})$$

3.1.3 COCO dataset

Microsoft Common Objects in context (MS COCO) dataset [6] has 91 common object groups. This dataset has 2,500,000 labeled instances in 328,000 images. This dataset is very extensively used for training because it has images of everyday scenes in their natural environment. Objects have been labeled using per-instance segmentation to help in achieving accurate object localization. This dataset contains images of 91 objects types. It can be noticed from the Fig. 5, that while MS COCO has fewer categories than ImageNet and SUN, it has more instances per category.

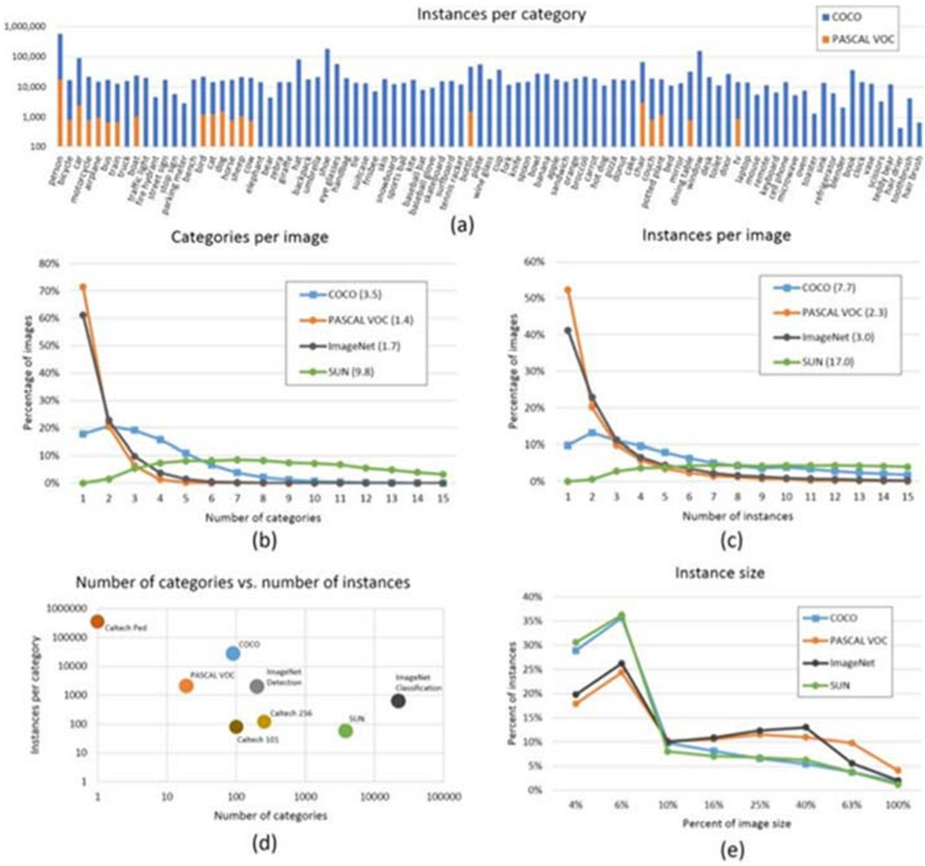

Fig. 5 Comparison With PASCAL VOC, COCO, ImageNet and SUN

3.1.4 Training and testing data

For training purposes, COCO [6] dataset is being used as it provides us 2,500,000 labeled instances in 3,28,000 images. The alternative would have been to collect data and label it manually, using COCO dataset helps us avoid this work. As it can be seen from Table 1, the COCO dataset provides better performance than the PASCAL dataset. DPMv5-P is the performance reported by [3] in VOC release 5. DPMv5-C uses the same implementation but is trained using MS COCO dataset. For testing purposes, 1000 unseen images were used. Figure 6 shows sample of detecting triple riding on which the model was tested on.

3.2 Vehicle tracking system

3.2.1 Vehicle monitoring system

This system mainly consists of a GPS module and a NodeMCU. The GPS module gives us the latitude, longitude and the velocity of the vehicle. This system makes use of the public network to connect with the NodeMCU module to extract the latitude, longitude, and velocity and send this data to the cloud for effective visualization of data. The platform used at the time is Firebase.

3.2.2 Velocity extraction system

The common traffic rule violated by the citizens is the speed limit. The GPS module also provides us with the velocity extraction of the vehicle. Every place has a speed limit which has to be followed by the citizens while driving in that area. Our system proposes importing

Fig. 6 Detection of triple riding

these speed limits of each place from the internet and then comparing it with the speed of the vehicle that is being driven in that area. If the speed of the vehicle is more than the speed limit of the area than the coordinates of the vehicle will be sent to the concerned authorities for processing of the data. Our system also takes into account some factors like frame per second which if used viably can reduce the cost of the system. In places that have schools, the speed limit is low and hence do not require very high-resolution cameras for classification. In places like highways, the speed limit is high and hence we need high-resolution cameras for effective capturing of images.

3.2.3 Coordinates extraction system

This system consists of the nodeMCU, GPS module and the real-time database of Firebase.

- GPS MODULE NEO6M - The GPS module is used to get the latitude, longitude and the velocity of the vehicle. This module is interfaced with the nodeMCU to get values and store them in the real-time database of the firebase [4].
- Node-MCU - The nodeMCU is the best choice of board because of the WIFI module present on it which can be accessed easily with the help of Arduino IDE which is the environment we code in. It is cheaper than other boards like raspberry pi which makes it the perfect choice of the board for this project. It also has a deep sleep mode which reduces the power usage of the board [4].
- Firebase - Firebase is being used here as a platform where we can update our real-time values latitude, longitude and velocity. Firebase provides a Realtime database. on Firebase's cloud. The Firebase Realtime Database is a cloud-hosted NoSQL database that is letting us store the values of latitude,longitude and velocity of the vehicle [4].

Three GPS modules have been interfaced with NodeMCU and are transmitting data to the real-time database Firebase.

4 Experimental results

The experimental results were conducted on a machine with INTEL CORE i5 with a memory of 7.7GiB with an Intel core i5-8250U CPU@1.60GHZ X 8 PROCESSOR WITH An INTEL UHD Graphics 620(Kabylake GT2). with GNOME 3.28.2 AND OS TYPE AS 64 BIT with a disk of 187.5 GB. The programs for the triple riding detection are written in Python3.7.3 with the help of various libraries like OpenCV, NumPy and Darknet [11] which is a neural network framework.

Table 2 Quantitative measures

Measure	Value
Accuracy	0.917
Precision	0.9
Recall	1
Error Rate	0.08
Sensitivity	0.9
Matthews correlation coefficient	0.77
F1 Score	0.947

The results of Table 2 were carried out using 7000 test samples, upon which the model carried out its detection process and these samples were collected in random from Google. The data set includes the frames generated from the videos. In order to, avoid redundancy the insignificant frames were removed in the final data set.

Experimental results from [7] on the PASCAL VOC, COCO, and ILSVRC datasets have confirmed that SSD has many methods for utilizing an additional object proposal step and is also faster. It gives a unified framework for the purpose of both training and inferences. For 300 $\times$ 300 input, SSD gives 74.3% mAP1 on VOC2007 test at 59 FPS on a Nvidia Titan X and for 512 $\times$ 512 input, SSD gives 76.9% mAP, which has outperformed state-of-the-art Faster R-CNN model. Even with a smaller input size SSD has a better accuracy when it is compared to other single stage methods.It shows comparison between SSD,Faster RCNN and YOLO. The SSD300 and SSD512 has beaten Faster RCNN in terms of speed and accuracy. Fast YOLO can be used at 155FPS but it has lower accuracy by 22% mAP. It concluded that if a faster base network is used,it could improve the speed and also make the SSD512 model real time.

5 Results and discussion

The results of the proposed system will be discussed here. The system is tested on images that have triple riding as well as images that do not have triple riding to check how accurate our system is. The results are displayed in the table. As depicted from Fig. 6 the system detects triple riders and puts a bounding box on them and labels them as triple riders. The accuracy of our model is 91.7% with an F1 score of 0.947 and also the precision value of our system is 0.9. The system without IOU layer resulted in lower accuracy of 90.4%. Because of the GPS module and the Node-MCU provides us with the exact coordinates of the vehicle and sent to the real-time database of Firebase for storing and constant update of data. Table 3 lists various models used in different papers and their accuracies. Most of the systems that are already proposed by other people concentrate on traffic rules like wearing helmets and breaking the speed limit. But our system has put together an algorithm that can distinguish triple riding bikers which is another violation of a traffic rule. Implementation of this kind of system will increase general awareness and hence reduce accidents.

Table 3 Comparative analysis with existing algorithm and proposed method

Algorithm	Accuracy
Proposed Model	91.70
YOLO(Without IOU)	90.40
SubCNN	79.48
VGG16	78.09
VGG19	79.11
InceptionV3	84.58
MobileNets	85.19
SSD	78.3

6 Conclusion

This paper proposed an automatic approach for detecting the cases of the triple riding and the speed limit violation system. The system was used the advanced version of the YOLO and is being trained on the COCO dataset. The IOU approach over the YOLOv3 using Darknet framework has been tested on the various set of images from the internet. Experimental results and the quantitative measures on different scenarios prove that the proposed system is reliable at the proper implementation of the system. With such high levels of accuracy and precision, this system has the potential to convert one of the most tedious tasks of the organization into an automated process. Systems like these are very important and should be implemented as these make sure that the general citizens are aware of traffic rules and follow them with complete devotion. In addition to making the general public aware, the safety of the public is also important. When the citizens start following the rules, the roads will become safer and less congested which is the ultimate goal behind the implementation of the traffic rules. This system can further be developed, like integrating the Face Recognition (FR) technologies to form an integrated hybrid algorithm to classify the images. Many FR systems were already been deployed, one such existing FR can be used as an extension to our system. By which the redundancy in identifying the triple riding persons can evade to a certain extent. The frame classified as a triple riding can be given as an input to the FR system which later can detect the persons in the frame and Aaadhar linked database can fetch the details of the persons to subject them for the necessary judicial procedures. As the currently proposed paper deals only with the triple riding and speed limit violation of the traffic rules, it is expected that many other functionality will be added to the proposed model. This can in turn completely wipe out the human effort in monitoring the traffic system.

References

1. Dahiya K, Singh D, Mohan CK (2016) Automatic detection of bike-riders without helmet using surveillance videos in real-time. In: 2016 International joint conference on neural networks (IJCNN). IEEE, pp 3046–3051
2. Desai M, Khandelwal S, Singh L, Gite S (2016) Automatic helmet detection on public roads. Int J Eng Trends Technol (IJETT) 35:185–188
3. 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
4. Gokhale M, Wagh R, Chaudhari P, Khairnar S, Jadhav S (2018) Iot based e-tracking system for waste management. In: 2018 Fourth international conference on computing communication control and automation (ICCUBEA). IEEE, pp 1–6
5. Hu Y, He Q, Zhuang X, Wang H, Li B, Wen Z, Leng B, Guan G, Chen D (2013) Algorithm for vision-based vehicle detection and classification. In: 2013 IEEE International conference on robotics and biomimetics. IEEE, ROBIO, pp 568–572
6. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft coco: Common objects in context. In: European conference on computer vision. Springer, pp 740–755
7. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) Ssd: single shot multibox detector. In: European conference on computer vision. Springer, pp 21–37
8. Płaczek B (2010) A real time vehicle detection algorithm for vision-based sensors. In: International conference on computer vision and graphics. Springer, pp 211–218
9. Rahman MA, Wang Y (2016) Optimizing intersection-over-union in deep neural networks for image segmentation. In: International symposium on visual computing. Springer, pp 234–244
10. Raj KD, Chairat A, Timtong V, Dailey MN, Ekpanyapong M (2018) Helmet violation processing using deep learning. In: 2018 International workshop on advanced image technology. IEEE, IWAIT, pp 1–4
11. Redmon (2016) Darknet: open-source neural networks in c. http://pjreddie.com/darknet/

12. 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
13. Satya (2018) Deep learning based object detection using YOLOv3 with OpenCV (Python/C++). https://www.learnopencv.com/deep-learning-based-object-detection-using-yolov3-with-opencv-python-c/
14. Silva SM, Jung CR (2018) License plate detection and recognition in unconstrained scenarios. In: European conference on computer vision. Springer, pp 593–609
15. Vishnu C, Singh D, Mohan CK, Babu S (2017) Detection of motorcyclists without helmet in videos using convolutional neural network. In: 2017 International joint conference on neural networks. IEEE, IJCNN, pp 3036–3041
16. 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
17. Xu Z, Shi H, Li N, Xiang C, Zhou H (2018) Vehicle detection under uav based on optimal dense yolo method. In: 2018 5th International conference on systems and informatics. IEEE, ICSAI, pp 407–411