REAL-TIME OBJECT DETECTION USING DEEP LEARNING: A BRIEF REVIEW

Mrs. Sangeeta M. Borde, Assistant Professor
Department of Computer Science
MAEER'S MIT Arts, Commerce & Science College,
Maharashtra, Pune
smborde@mitacsc.ac.in

Dr. Harsh Lohiya, Assistant Professor
Department of Computer Application
SSSUTMS, Sehore, M.P

ABSTRACT
Object detection is the most common application of computer vision for the last 20 years. It has been widely used for quick & accurate identification and locating of a large number of objects from predefined image categories, real-time video frames, etc. The computer vision field requires several algorithms for detecting objects such as Single-shot detection (SSD), Faster region-based convolutional neural networks (faster R-CNN) & You Only Look Once (YOLO) with its variations using deep learning, etc. Based on parameters such as accuracy, precision & F-score performance of these algorithms is analyzed. In this paper mentioned comments are based on studied literature & key issues related to the topic are also identified which are relevant to the object detection area with accuracy and performance. The paper review begins with an introduction to deep learning and its techniques. (Allamki & Sateesha,2019) Further, CNN architecture and object detection are described along with some modifications to improve performance. Image classification is come into existence for decreasing the gap between human vision and computer vision by training the computer with data. At the end of the article; some promising guidelines & directions are provided for further work in deep learning & object detection techniques.
Keywords: Image detection, Computer Vision, Deep Learning, CNN

Introduction

In recent years, the development of deep learning for performing complex tasks such as image classification, image recognition, image segmentation & image detection has been rapidly increasing. (Gu, Xu, Wang & Shi, July 2019). For decreasing the gap between human vision and computer vision field image classification has come into existence by training the computer with data. Based on the contents of the vision; the image classification is achieved by differentiating the category of the image. A deep learning technique has been divided into three main categories namely Convolutional Neural Networks (CNN), Restricted Boltzmann Machines (RBM), and Autoencoders. The convolutional Neural Network has inspired by the visual system's structures. (Tonge, Chandak, Khiste ,Khan, & Bewoor,2020).One of the important techniques of deep learning is described below:

Techniques in Convolutional Neural Networks

CNN (ConvNet) is made up of neurons that have learnable multiple weights and biases. Each neuron takes some inputs, does a dot product, and then executes a non-linearity if desired. From raw picture pixels on one end to class scores on the other, the entire network is expressed by a single differentiable score function. (Iamudomchai, Seelaso, Pattanasak & Wibool,2020). The following layers are used to build ConvNet architecture:
i.Convolutional Layers.

ii. Pooling Layers.
i. Fully connected Layers.
The first layer is a convolutional layer, which is used to perform a convolutional operation on any image; so, the name is given as a convolutional Neural network. Usually, to sample down the input a pooling layer is used. In the end, a fully connected layer of the Convolutional architecture of CNN is used to handle the complete classification process. In comparison with a typical neural network, the Convolutional layer is better. It is made up of neurons that have been grouped into a three-dimensional convolution matrix with width, height, and color channels.

Object detection

Object detection is a process of "Localizing" the object and classifying the object as per the category to which it belongs. Commonly used applications in object detection image retrieval, security, surveillance, advanced driver assistance systems, etc. are assistance systems used for self, safety, and care. A common approach for the object

 Copyright © The Online Journal of Distance Education and e-Learning

detection framework is classified using the CNN feature which includes the creation of a large set of candidate windows. Object detection can be done in multiple ways; some of the important ways are as follows. (Liang & Juang,2015):

1. Feature-level Object detection
2. Jones Viola-based Object Detection.
3. SVM classifications with HOG Features.
4. Object Detection using Deep Learning.

Objectives

1. To understand the methodology & algorithms used in the recent research article for further research work.
2. To provide research insights, existing gaps, and future research directions.

Literature Review

A hybrid descriptor for feature extraction of an image, and it is based on Local Binary Pattern (LBP), the Hough transformation descriptors & Histograms of oriented gradients. Road Traffic images captured by CCTV were used as input. For better accuracy, the random forest algorithm has been chosen for classification. Object detection has two steps: Vehicle Segmentation &Vehicle classification. (Rumuere, Kelson & Thiago,2013; Wen, Yuan, Liu, & Zhao, 2007). In the first step, for the segmentation of vehicles; background detection and moving object segmentation are needed. In the background, detection process a video camera has used to capture the images & the determination of an image will be used for detecting a moving object. Also, in the moving object segmentation approach, it is necessary to define a line and the line must be marked by the user. The process of moving object segmentation begins when a vehicle crosses the line (CL), and the image frame is collected. (Rumuere, Kelson & Thiago, 2013; Wen, Yuan, Liu, & Zhao, 2007). A vehicle counting tracking algorithm (TA) is required to count each vehicle once.

The detection process in (1) technique for helmets has divided into three basic steps: I) The area of interest (ROI) II) Image categorization and III) Feature extraction. (Rumuere, Kelson & Thiago,2013; Wen, Yuan, Liu, & Zhao, 2007). To decrease the computational cost & search area ROI is important. For the feature extraction, the hybrid descriptor has been used by combining Circular Hough Transform (CHT), LBP & HOG descriptors. For Image classification, three different classification families were described Naïve Bayes, Random Forest, and Support Vector Machine (SVM). The main drawback of the proposed research work is that captured images are not in better resolution so the quality of the resolution can be improved to get an accurate result of classification.(f1) Because of the low resolution, several images are not included in the image detection. In the helmet detection step, the Random Forest (RF) algorithm obtained a best result than other algorithms. The accuracy rate is 0.9380.(Babu, Rathee, Kalita, & Deo, 2018) Reintroduced a circular arc detection method on (MHT) modified Hough transform. This kind of transformation is used to detect a helmet by an ATM Surveillance system (Babu, Rathee, Kalita, & Deo,2018). In the next stage, the SIFT (Scale Invariant feature transform) algorithm is used to detect moving objects; that is motorcycle conventional neural network results in the field of CV and neural language processing on small training data. For large training data, the choice has been a deep neural network. (Babu, Rathee, Kalita, & Deo,2018).

The proposed methodology has been classified into the Moving object detection process, vehicle classification process, Helmet detection process, and License plate extraction process (Babu, Rathee, Kalita & Deo,2018). Video Road Traffic video frames were taken as input for the image detection process. In the Moving object detection process; video frames in the moving objects are segmented from the background. The foreground detector needs a certain number of video frames but this detector is not perfect as it often includes undesirable noise. Bounding boxes will be connected to each required moving object for image detection. Vehicle classification has been done using a Machine learning algorithm. This will be performed accurately with limited data. For the helmet detection process, numerous ML classifiers are used to select the best one. ROI has been extracted from the cropped image by giving appropriate coordinates to extract the license plate of the vehicle. Several Machine learning classifiers are used such as Random Forest (RF), Gradient Boosted Trees (GBT), Support Vector Machine (SVM), DNN, etc. (Babu, Rathee, Kalita, & Deo,2018; Zamani, Mehdi,2019). The drawback of research work is due to a lack of data deep Neural networks are not performed better than Random Forests. So, the DNN algorithm shines when there is huge training data. (Babu, Rathee, Kalita, & Deo,2018) All the above four classifiers are trained on about 2000 input images only. Here classifier accuracy on test images using Random Forest in vehicle classification has 91% and Helmet detection result using the Random Forest has 92%. Therefore, for a large amount of data classifier accuracy decreases.

 Copyright © The Online Journal of Distance Education and e-Learning

(Narong, Wichai,2018) the method used is a Single Shot Multi-Box Detector (SSD) of deep learning that worked on Helmet detection problems (Rose, 2020; Narong, Wichai, 2018). Here, In the experiment, some CNN models were used to compare the result (VGG, GoogleNet & Mobile Nets). In the Helmet detection challenge, the author discovered that the network models Mobile Nets Model and SSD model combination produced accuracy are best than other network models. when compared to GoogleNet and VGG(Narong, Wichai,2018). Four different methods were used for image classification experiments on input datasets. The proposed system has been used to solve the helmet detection problem with the help of convolutional neural networks such as VGG16, VGG19, GoogleNet, or Inception V3 & Mobile Net(Narong & Wichai, Nov 2018). The experiment was done on a video segment of a vehicle that was passed through the gate. The research is carried out with the help of four different steps: steps are as follows: 1. Image & video gathering.2 Image classification experiment.3. Image detection Experiment and 4. Result from interpretation. (Rose, 2020; Narong, Wichai, 2018). In the last step, the performance from the image, video gathering & image classification steps is compared.

Table No.01 reveals the best CNN model is the MobileNet network model for recognizing motorcyclists wearing helmets and those who are not wearing helmets (Narong, Wichai,2018). The accuracy of classification is the best. The drawback of using this network model is; it gives the highest accuracy if the size of the training data is less in comparison with another network model. If the size of the data increases accuracy will be decreased (Narong, Wichai,2018).

Network Model	Model Size (KB)	Accuracy
VGG16	434,580	78.09%
VGG19	451,258	79.11%
Inception V3/GoogLeNet	85,447	84.58%
MobileNets	16,754	85.19%

Table No.01, Result of Network Model MobileNets. (Narong, Wichai, 2018).

(Siebert, Lin 2020) The most advanced object-detecting algorithm is a formula for automatically registering motorcycles. Helmet usage as determined by video footage using the DL method. The algorithm's effectiveness was evaluated by comparing it to the outcomes of a prior observational study on the use of helmets and by using an annotated test dataset (Siebert & Lin, 2020). Single-stage object detection algorithms like YOLO and Retina Net were applied to training sets of cities in Myanmar from numerous observation points at various times of the day. For the broad topic of bikes, the algorithm has great accuracy (Siebert &, 2020). It is capable of correctly counting the riders on the motorcycles and determining where they are seated. The algorithm for object detection is trained using the training dataset. A validation set is employed to identify the best-generalized model. (Allamki, Panchkshri & Sateesha 2019); working on the DNN-based model You Only Look Once (YOLO). The model is chosen to get more accuracy & speed for the object detection system. To train for the custom classes; the annotated images are given as input to the YOLOV3 model.
The flowchart in Fig No.01 Allamki, Panchkshri & Sateesha (2019) below for Helmet Detection and LP recognition shows the overall methodology.

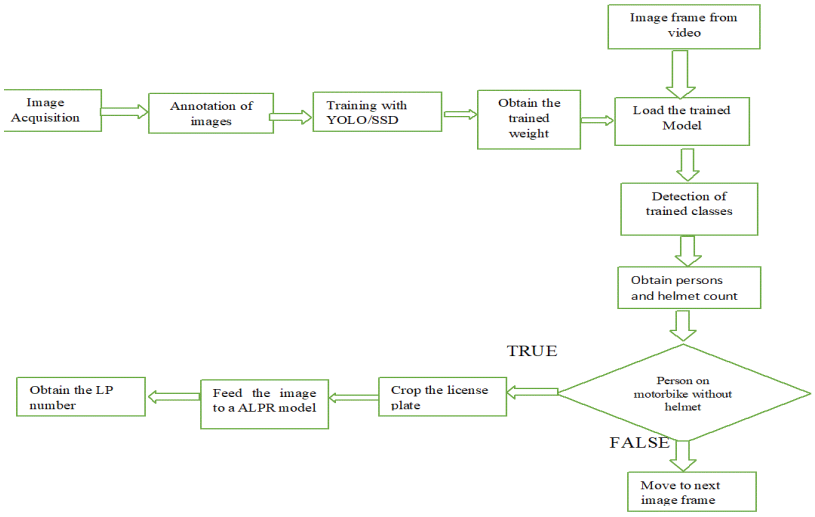

Fig No.01 Flowchart for Helmet Detection and LP Recognition.(Allamki, Panchkshri & Sateesha,2019).

 Copyright © The Online Journal of Distance Education and e-Learning

The system is developed with the help of a CCTV camera 's-based model. YOLO is chosen for accuracy & speed. YOLO is a real-time object detection system. Here, for the helmet detection process; objects are categorized into five classes Helmet, Non-Helmet, Motorbike, the person (sitting on the bike), and License plate. To recognize those features from other images, a classifier based on the CNN network from deep learning is used. A license plate is extracted similarly. Once the coordinates of LP are found, LP is cropped & saved (Allamki, Panchkshri & Sateesha,2019).

To an OCR model; The extracted new image of LP is given. The text in the given image recognized by OCR and the OCR module will give an output of predicted LP numbers along with confidence value. For recognizing given LP accuracy; the confidence value is used. It indicated how confident the OCR module is. A system that captures video from CCTV cameras at road junctions by making use of ML & CV techniques. Classifiers are built using CNN. Optical Character Reader (OCR) is used for number plate recognition & to penalize the concerned motorcyclists. For a non-helmet motorcyclist, detection accuracy was found 98.72%. Here, training and testing datasets are used for 2 hours and 1 hour respectively. The accuracy calculated by the category of motorcycle vs nonmotorcycle classifier is 99.68%. Whereas, the helmet vs nonhelmet classifier gave 99.04%. The following Table 2 shows the performance of each classifier in percentage on the test (Kulkarni, Bodkhe, & Patil,2018).

Classifier	**Performance (%)**		
	1. Accuracy	2. Precision	3. Recall
Vehicle: Motorcycle vs non-motorcycle classifier	99.68	99.5	99.5
Helmet vs non-Helmet classifier	99.04	99.30	98.92
OCR	99.36	96.84	96.51

Table No.02 performances of each classifier on the test. (Kulkarni, Bodkhe & Patil, March 2018)

(Dorathi, Joel, Sinha, & Malathi,2020) A method for recognizing the riders that are not wearing helmets & automate the detection of traffic violators using the CNN-based algorithm YOLOv3(2020). For Fast single-stage, YOLOv3 is an object detection algorithm. It is based on darknet architecture and trained with the help of the MSCOCO dataset.

MS COCO Dataset is capable of detecting at least 80 classes of objects at a time (Dorathi, Joel, Sinha, & Malathi,2020). LeNet architecture captures all the images coming from the frame so it gives less accuracy rather than Motorbike riders. In YOLO Dense backbone method accuracy is 98.78%.The drawback of this module is to capture the images of all persons coming in the video frame rather than motorcyclists. This directly affects the accuracy of the classification (Dorathi, Joel, Sinha, & Malathi,2020). (Xu, Wang, LinShi, & Yuwan,2019). An improved faster RCNN model is used to optimize the original model which combines some techniques like increasing anchors, multi-scale training, & OHEM. In the RCNN model user can learn two modules: Region proposed network (RPN) module that generates candidate regions &a fast RCNN object detection module. The RPN generates ROIs &applies an "attention" mechanism to the object detection module. RPN used a sliding window mechanism.

(Singh & Babu, 2017) Adaptive background removal, which is insensitive to light, video quality, and other factors. This technique is used with moving objects. For identifying motorcyclists and no motorcyclists, these moving objects are fed into a CNN classifier (Singh & Babu,2017; Khan,2021). The suggested method is tested on two datasets with sparse and congested traffic. HOG-SVM performs best, with an accuracy of 99.24 percent Singh & Babu (May 2017), Khan (2021). The experimental evaluation's result reveals that utilizing traditional CNN increases classification performance for both classification tasks, resulting in the reliable detection of violators driving without a helmet. The suggested method is tested on two video datasets with sparse and dense traffic. HOG-SVM has the best performance, with an accuracy of 99.24 percent. (Khan,2021). The experimental results reveal that utilizing traditional CNN increases classification performance for both classification tasks, resulting in the reliable detection of violators driving without a helmet (Khan,2021).

 Copyright © The Online Journal of Distance Education and e-Learning

Deep learning convolutional neural network architecture such as VGGNET and ALEXNET. ALEXNET has usually referred to as the DCNN for object recognition & for achieving good performance. The helmet detection using VGGNET has 86% accurate. Using ALEXNET the accuracy of helmet detection has 96.03% (Mansoor,2019).(Tonge, Chandak, Khiste, Khan, & Bewoor, 2020) The Following modules are introduced module names are: The first module for vehicle detection second, For Helmet classification, the Third for crosswalk violations at traffic signals, and the fourth module for License plate recognition (Bewoor,2020). The work done here; has been used for detecting a vehicle using object detection; This process was performed using the YOLO algorithm, & each vehicle in traffic was checked against appropriate violations. Helmet violation in heavy traffic was detected using CNN based classifier and vehicle numbers are obtained with the help of OCR.Fig.2 below represents the clear architecture of YOLO which has 24 convolutional layers and two fully connected layers. The Dataset used for the research was prepared from CCTV footage of several signals obtained from the traffic police department. (Tonge, Chandak, Khiste, Khan, & Bewoor,2020).

The flowchart in Fig No.02 Tonge, Chandak, Khiste, Khan, & Bewoor,2020 shows the YOLO network model.

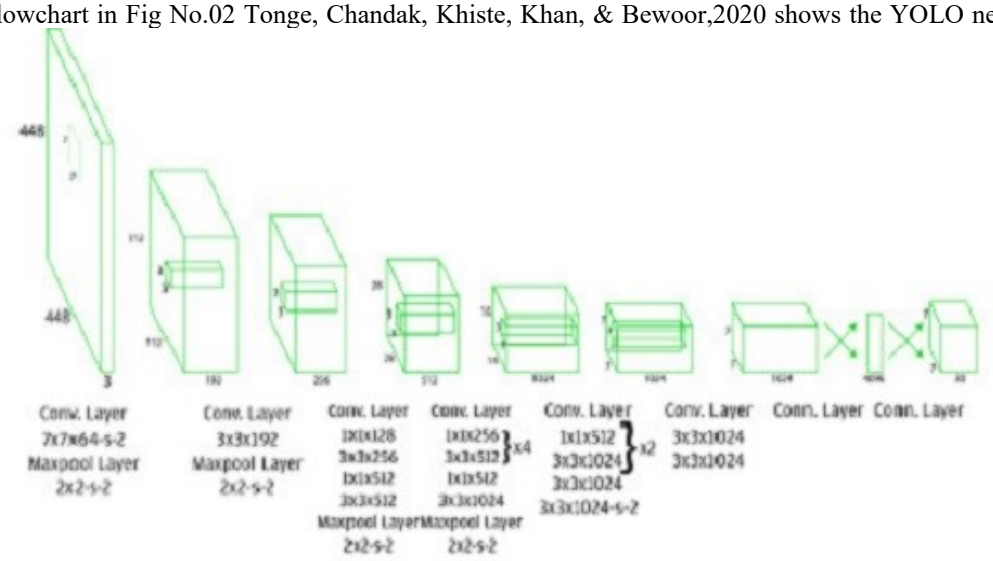

Fig No.02 YOLO Network Model

Summary of a Few contextual gaps

- The study did not address classification accuracy. (Rumuere, Kelson & Thiago,2013; Wen, Yuan, Liu, & Zhao, 2007).
- The study did not investigate the accurate result due to a lack of a training dataset. (Babu, Rathee, Kalita, & Deo,2018).
- This study addresses that different network model gives the different highest accuracy. For accuracy need to change the CNN network model. (Rose, 2020; Narong, Wichai, 2018).
- This study addresses that for the helmet detection process; objects are categorized into five classes Helmet, Non-Helmet, Motorbike, the person (sitting on the bike), and License plate. To recognize those features from other images, a classifier based on the CNN network from deep learning is used(Allamki, Panchkshri & Sateesha 2019).
- This study investigates the Optical Character Reader (OCR) is used for number plate recognition & to penalize the concerned motorcyclists (Kulkarni, Bodkhe, & Patil,2018).
- This study did not investigate the motorcyclists rather it captures the images of people. The classification accuracy problem(Dorathi, Joel, Sinha, & Malathi,2020; Xu, Wang, LinShi, & Yuwan,2019).
- This study did not address what computational power it requires for image segmentation and object detection (Tonge, Chandak, Khiste, Khan, & Bewoor, 2020).
- This study did not address all the factors for better performance & accuracy (Singh & Babu,2017; Khan,2021).

 Copyright © The Online Journal of Distance Education and e-Learning

Findings

- ✓ Captured images aren't in better resolution so the quality of the resolution can be bettered to get an accurate result of the bracket.
- ✓ Deep Neural Network didn't perform better than Random Forest due to lower trained data. Classifier delicacy on test images using Random Forest in the vehicle bracket has 91 and Helmet discovery results using the Random Forest has 92.
- ✓ The stylish CNN MobileNet network model for feting motorcyclists wearing helmets and those who aren't wearing helmets.
- ✓ The accuracy calculated by the category of motorcycle vs nonmotorcycle classifier is 99.68%. Whereas, the helmet vs non-helmet classifier gave 99.04%.
- ✓ The confidence value is used for recognizing given LP accuracy; It is used to indicate how confident the OCR module is.
- ✓ LeNet architecture captures all the images coming from the frame so it gives less accuracy rather than Motorbike riders. In YOLO Dense backbone method accuracy is 98.78%.
- ✓ The helmet detection accuracy using VGGNET & ALEXNET has 86% & 96.78% respectively
- ✓ The suggested method is tested on two video datasets with sparse and dense traffic. HOG-SVM has the best performance, with an accuracy of 99.24 percent.

Conclusions

Recent studies have challenged the real-time image detection application. It has several approaches but using the perfect model or architecture; the user can increase the accuracy of detecting images and classifying images from datasets. In this paper, numerous methods are discussed to detect images from the dataset but the number of datasets used for training and testing purposes has a promising task and it's one of the challenges in the field of Neural Networks and Computer Vision. This paper provides a systematic review of deep learning-based object detection frameworks that handles various kinds of task such as classification, image recognition, and image detection viz. The CNN will give the best result if you train a huge amount of data or video frames. In the end, several promising future directions and guidelines are proposed to understand the landscape of object detection techniques. (Siebert, F. W, & Lin H,2020).

References

Allamki L., Panchkshri M., Sateesha A., and K. P. (2019),” Helmet detection using machine learning and automatic License Plate Recognition” in the International Research Journal of Engineering and Technology (IRJET), 6(12).

Babu R., Rathee A., Kalita K., Deo M. (2018),” Helmet Detection on two-wheeler Riders Using Machine Learning Proceeding of ARSSS in the International Conference, Volume-5, Issue-9, India (PP.14-17).

Boonsirisumpun N., Wichai P., & Wairotchanaphuttha P. (2018),” Automatic Detector for Bikers with no Helmet Using Deep Learning, Electronic” ISBN:978-1-5386-8164-0 Print on Demand (PoD) ISBN:978-1-5386-8165-7, (pp.21-24).

Dorathi J.D., Joel J., Sinha S., D. M(2020),” Detection of two-wheeler riders without Helmets using YOLOv3 and Image classifier” in the International Journal of Advanced Science and Technology Volume.29, No.7, (pp. 899-907).

Draz A., Khan H., M.Z, & Khan M (2021),” Automatic Helmet Violation Detection of Motorcyclists from Surveillance Videos using Deep Learning Approaches of Computer Vision” (2021) in the international conference on Artificial Intelligence (ICAI). (pp.252-257).

Gu, Y., Xu, S., Wang, Y., & Shi, L. (2019), “An advanced deep learning approach for safety helmet-wearing detection” in the International Conference on Internet of Things (Things) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) (pp. 669-674). IEEE.

Iamudomchai P., Seelaso P., Pattanasak S., Wibool. (2020), ”Deep Learning Technology for Drunks Detection with Infrared Camera” in the international conference on Engineering, Applied Sciences, and Technology (ICEAST) (IEEE) 978-1-7281.

Mansoor K,R., M. M, S. M. (2019), “Deep Learning Helmets-Enhancing Security at ATMs” in the International Conference on Advanced Computing & Communication Systems (ICACCS) (pp. 1111-1116). IEEE.

Kulkarni Y., Bodkhe, S., Kamthe A., & Patil A.(2018),” Automatic number plate recognition for motorcyclists riding without a helmet.” in the International IEEE Conference on Current Trends towards Converging Technologies (ICCTCT) (pp. 1-6).

Liang, C.W., & Juang, C.F. (2015), “Moving object classification using local shape and HOG features in wavelet-transformed space with hierarchical SVM classifiers”. Applied Soft Computing, 28, (pp.483–497). http://sersc.org

 Copyright © The Online Journal of Distance Education and e-Learning

Rose A., Siby S., Vettom S., Raju S., Jose G. (2020), "Automated Helmet Detection using Image Processing Algorithms in the International Journal of Science Technology & Engineering | Volume 7 | Issue 1 ISSN (online): 2349-784X(IJSTE) (pp.44-48).

Rumuere S., Kelson A., and Thiago S.(2013)," Automatic detection of motorcyclists without a helmet" at the IEEE XXXIX Latin American Computing Conference (CLEI) (pp. 1-7).

Siebert, F., & Lin H. (2020),"Detecting motorcycle helmet use with deep learning. Accident Analysis & Prevention", 134, 105319.

Singh D., and Babu S.(2017), "Detection of motorcyclists without a helmet in videos using convolutional neural network". International IEEE Joint Conference on Neural Networks (IJCNN) (pp. 3036-3041).

Tonge, A., Chandak, S., Khiste, R., Khan, U., & Bewoor , L. A. (2020)." Traffic Rules Violation Detection using Deep Learning"in the IEEE International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1250-1257).

Wen X., Yuan H., Liu W, and Zhao H.(2007)," An improved wavelet feature extraction approach for vehicle detection", in ICVES, pp. 1–4.

Zhao, Z.Q., Zheng, P., Xu, S.T., & Wu, X. (2019), "Object Detection with Deep Learning: A Review". IEEE Transactions on Neural Networks and Learning Systems, (PP.1–21)

 Copyright © The Online Journal of Distance Education and e-Learning