Available online at: http://www.ijmtst.com/vol6issue04.html

International Journal for Modern Trends in Science and Technology

ISSN: 2455-3778 :: Volume: 06, Issue No: 04, April 2020

Vehicle Detection and Classification Using Deep Learning: A Survey

Ruchi Patel[1] | Dr. Udesang Jaliya[1] | Dr. Narendra Patel[1]

[1]Department of Computer Engineering, Birla VishvakarmaMahavidyalaya, Vallabh Vidyanagar, Gujarat, India

To Cite this Article

Ruchi Patel, Dr. Udesang Jaliya and Dr. Narendra Patel, "Vehicle Detection and Classification Using Deep Learning: A Survey", *International Journal for Modern Trends in Science and Technology*, Vol. 06, Issue 04, April 2020, pp.:268-273.

Article Info

Received on 25-March-2020, Revised on 15-April-2020, Accepted on 19-April-2020, Published on 23-April-2020.

ABSTRACT

Detection and classification of different types of vehicles is a difficult task, when vehicles are occluded behind other vehicles, because of the background conditions or due to weather conditions. The main aim of vehicle detection is to detect all types of vehicles that are present inside the image, so it can be useful in case of the prevention of accident and traffic analysis. Vehicle Detection and Classification becomes an important task in various applications such as for estimating queue length, analyze traffic speed, tracking individual vehicles, counting vehicles, traffic congestion, prevention of accidents. There are many deep learning algorithms implemented for the detection and classification of vehicles by different researchers. In this paper, various methods are presented which are used for vehicle detection and classification. It includes deep learning algorithms such as Convolutional Neural Network (CNN), Region-based Convolutional Neural Network (R-CNN), Fast RCNN, Faster RCNN, Mask R-CNN, and You Only Look Once (YOLO).

KEYWORDS: *Vehicle Detection, Vehicle Classification, Deep Learning, Object Detection, CNN, R-CNN, Fast RCNN, Faster RCNN, Mask R-CNN, YOLO*

I. INTRODUCTION

Vehicle detection and classification is an important task for many applications such as for tracking individual vehicles, counting vehicles, collection of tall tax, vehicle trajectory prediction, analyzing the traffic speed. Detection and classification of vehicles play a crucial role in the case of traffic monitoring and management. The main objective of vehicle detection and classification is to detect every type of vehicle that are presents inside images and classify detected vehicles into their types such as a car, bus, truck, auto, bicycle, and motorcycle. Detection of vehicles is one of the most challenging tasks when vehicles are occluded by other vehicles or due to lighting conditions[1]. It is difficult to detect vehicles in case of background obstacles, such as trees, road signals, weather conditions such as rain, snow[2].

The goal of vehicle detection is to determine whether the vehicle is present inside the image or not and if the vehicle is present inside the image then return the spatial location and each vehicle instance. Generally, there are two main approaches available for Object Detection: Machine Learning-based approaches or Deep Learning-based approaches[3]. Machine Learning provides various techniques such as Viola-Jones object detection framework based on Haarfeatures, Scale-Invariant Feature Transform (SIFT), and Histogram of Oriented Gradients (HOG) to define the features and then use techniques such as Support Vector Machine (SVM) to classify detected

objects into their types[3]. Convolutional Neural Network (CNN) is used to extract features using Deep Learning Approach. When training with a large amount of data, Deep Learning is popular due to better accuracy[4]. Performance of a Deep Learning approach is better rather than the methods used for feature extraction such as Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), and Viola-Jones, the deep network will automatically learn different features from the train data[5].

There are many Deep Learning methods available to detect different types of vehicles and classify them into various categories of vehicles such as a car, bus, truck, auto, bicycle, and motorcycle. Various Deep Learning methods used for the purpose of vehicle detection and classification such as a Convolutional Neural Network (CNN) that has been a widely used method for object detection, Region-based Convolutional Neural Network (R-CNN), Fast RCNN, Faster RCNN, Mask RCNN, and YOLO. Fig. 1 shows sample images from the IITM-HetraDataset[6].

Fig. 1 Vehicles images from IITM-HetraDataset[6]

This paper provides a survey of various Object Detection methods for classifying vehicles.

II. DETECTION AND CLASSIFICATION

Detection and Classification of the object is an important task[7]. Object Detection provides object class and its location information from the image, and Classification of the object gives information about the object belongs to which particular type of vehicles[7]. Many researchers used a Deep Learning Approach for the purpose of detection and classification of vehicles. C. N. Aishwarya et al.[8], in this paper, YOLO method is used for faster object detection. Deep Learning Classifier layers are trained using a dataset that contains both types of images static and dynamic images in order to give better prediction accuracy. It uses CNNs for classification purpose. They have also compared the accuracy of Tiny YOLO with YOLO. Tiny YOLO require less prediction time compared to YOLO because it has a few numbers of layers and it was trained to identify only some objects. B. Hicham et al.[9] proposed a CNN method to perform vehicle type classification. These systems consist of two methods: The Data Augmentation method is used for the purpose of reducing the imbalanced dataset problem and the CNN model is used for image classification. CNN structure comprises of series of convolutive layers and fully connected layers. Convolution layer performs feature extraction and fully connected Artificial Neural Networks (ANN) layer work as a classification. Precision, Recall and Accuracy metrics used for the purpose of the model evaluation.

M. V. et al.[10] presents the R-CNN method that is used as a classification of various moving vehicles with region proposals. Box filter is used for smoothening the rapid variation that occurs because of the movement of the vehicles. Remove the background information from the frame to identify the Region of Interest.

K. Shi et al.[11] proposed a detection model based on Fast R-CNN for the purpose of vehicle detection. An author has used a new way of efficient vehicle detection through incremental learning and pre-processing approach in order to optimize the training process, during training process training parameters adjusted in order to achieve the best state. The vehicle detection model comprises of two stages: training and testing. In the training stage, after pre-training on ImageNet, CNN of initial parameters is re-trained. It includes three types of networks in the pre-training process: CaffeNet, VGG_Cnn_M_1024, and VGG-16. To initialize all the layers before the RoI layer uses a network model that was obtained by training ImageNet. The CaffeNet is used to initialize Fast R-CNN, so for that, the RoI pooling layer is added by the last convolution layer. The multi-task loss function is used and bounding box regression is added directly to the CNN network. They have combined the BUU-T2Y dataset and KITTI dataset[12] with incremental learning. Using an

incremental process, the performance of vehicle detection is improved.

D. Mittal et al.[5] presented Faster RCNN for vehicle detection using a limited heterogeneous traffic dataset. Deep Learning Approach for object detection requires a large amount of dataset for training because the network has to learn millions of parameters. The collection of a huge amount of dataset is time-consuming because it requires the labeling task. So data augmentation technique is proposed in order to solve this problem and they have augmented the Pascal VOC dataset[13] with the IITM-HeTra[6] dataset. After that, add new class auto-rickshaw to Faster RCNN model and train with the augmented dataset, so using augmentation technique it will able to detect all types of vehicle and gives better results. B. Benjdiraet al.[14] proposed a Faster RCNN model and YOLOv3 model for car detection. Region Proposal Network (RPN) is used to generate regions with different scales and aspect ratio. Fast RCNN uses these regions of interest as input. They have performed an experimental comparison between Faster RCNN and YOLOv3. The performance matrix is prepared using the F1 score, precision, recall, processing, and quality time. Both methods have high precision that means both are able to recognize the car in the image accurately, but in the case of a recall, YOLOv3 outperforms Faster R-CNN. That means YOLO V3 is able to detect all the cars that are present inside the image with better accuracy. G. Prabhakar et al.[15] presents Faster RCNN that is used to detect and classify on-road objects. The authors have also provided a description of the conventional CNN and different variants of R-CNN. The Mean Average Precision (mAP) is used to measure the detection accuracy. Training of the Region Proposal Network (RPN) is performed using Stochastic Gradient Descent (SGD) for classification and regression. There are two types of training for Object Detection: Approximate Joint Training and Alternating Training. In Approximate Joint Training, both RPN and Fast RCNN networks are trained at the same time and in Alternating Training, proposals that are generated using RPN are used to train fast R-CNN. M. C. Olgunet al.[16] presents Faster RCNN and Haar Cascade Classifier. The system is trained on Faster RCNN to detect objects and vehicles. Haar Cascade Classifier is used to detect traffic light and stop sign. The distance between Camera and the detected objects are calculated during the detection of a traffic light or stop sign. For lane tracking, CNN based on NVIDIA's PilotNet is used. Faster RCNN for the purpose of vehicle tracking provides better performance.

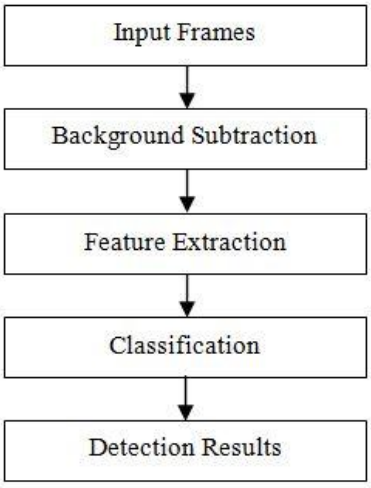

Fig. 2 Vehicles Detection and Classification[7]

Fig. 2 shows the steps of vehicle detection and classification. To perform vehicle detection and classification extract frame from the video. Apply Background Subtraction Technique on each frame, it separates out the foreground objects from the background one, and it is also useful in the case of the detection of moving vehicles from the video sequence. Then extract features and classify extracted objects into their types.

Lu Lou et al.[17], in this paper, Mask RCNN is used to detect vehicle contour and Kalman filter used for vehicle target tracking. Using Mask RCNN existing training models such as COCO some of the vehicles are not detected. So Fine Tuning is performed, and after performing this some of the vehicles are detected that are not detected before the Fine Tuning. They have shown a comparison between the Mask RCNN original model and trained Fine Tuning Model for vehicle detection, and found that the train Mask RCNN model gives better detection results. In a better daytime traffic scenario, this algorithm achieves 98% counting accuracy while detecting and counting the vehicles. Kalman filter is enhanced for the purpose of identifying and counting the moving vehicles. The authors have tested this algorithm on the diversity of traffic situations and under different weather conditions. And found that it solves the problem of partial vehicle occlusion and it improves the recognition tracking accuracy under a variety of traffic scenarios and different climates situations.

Y. He et al.[18] focuses on various tasks: Coordinate Space Conversion, Vehicle Detection,

and Data Fusion. To detect vehicle object, they have used Deep Learning Algorithm based on CNN. The author proposes YOLOv2 method for the detection of vehicles from the video. Coordinate Space Conversion is used to acquire the pixel positions of radar detection of vehicle objects in a video sequence. Data Fusion is the process of integrating video data and radar data to obtain the exact location of the object, which is an important component for performance tracking. Appearance information is also required to obtain better tracking results. Centroid Distance needs to be calculated in order to improve the location precision and detection accuracy in the data fusion module, it is important to calculate in order to differentiate whether the nearby object is the same or not. By combining Video Detection and Radar Information, the performance of vehicle detection is improved. The results using the real datasets show that the algorithm is very efficient for vehicle detection and tracking.

Yi-Hsuan Hsu et al.[19], in their paper, proposed a PVA-lite Deep Learning model for the detection of vehicles. For matching detection results with the current trackers, the Data Association method is used and the Kalman Filter is used to predicting the position of the lost tracker.

III. SURVEY OF EXISTING METHODS

A. Detection and classification

Deep Learning methods are used for the detection and classification of objects including vehicles. It provides an extracted frame from the video to the network and is passed through the different layers of the Neural Network. The final results are available in the form of a class label, bounding box and in the case of Mask RCNN, an object mask is an important result.

Convolution Neural Network (CNN)[20] is a technique to detect objects, in this case, detect different types of vehicles and classify them into classes. CNN classification takes the image as an input, process the image and classifies it into various categories such as a car, bus, truck, auto, bicycle, and motorcycle. It passes the image into a sequence of convolution layers. The structure of a CNN consists of Convolution Layer, a Pooling Layer, and Fully Connected Layers. The Convolution Layer is used to extract features from the image. Convolution of image and kernel results in Feature Maps. When the image is too large, the Pooling Layer is used to reduce the number of parameters. Down-sampling and Up-sampling are called spatial Pooling; which is used to reduce each Feature Map dimension while maintaining its essential information. A Fully Connected Layer, flatten the matrix into a vector and is provided to the Softmax or Sigmoid Activation Function to classify vehicles into their categories.

You Only Look Once (YOLO)[21] is used to detect objects in the image. The image is divided into an s*s grid and predict the confidence and bounding box for each grid. The confidence value affects the bounding box accuracy and determines whether the object is there in a bounding box or not. For each box, it predicts the classification score. Combine both classes to compute the probability of each class being present in the predicted box.

Region-based Convolutional Neural Network (R-CNN) is used to improve bounding box quality and to extract high-level features[1]. R-CNN[22] is a combination of two steps: Selective Search method used to generate regions from the image and extract features. CNN is pre-trained for image classification purposes. The Selective Search method is used to generate a Region of Interest. The generated regions are of different size. Regions are then reshaped so that all regions are of the same size as required by the CNN. These regions are fed to CNN for generating a feature vector. These feature vectors are used by SVM and for each class train SVM independently. The regression model is trained to generate a tighter bounding box for each object that is found in the image.

Fast R-CNN[23] is a method from the category of Region-based CNN. In R-CNN, generate regions from the image and then these generated regions are fed to CNN. Instead of running CNN for each region, apply CNN once per image. After extracting the feature map uses the RoI pooling layer to reshape all regions into a fixed size. These regions are fed into the Fully Connected Layer that classifies using the Softmax activation function and generates a bounding box using the Regressor activation function.

Faster R-CNN[23] is popular for object detection and classification. Fast RCNN uses a Selective Search method to find a region of interest which is slow and time-consuming process. Faster RCNN uses Region Proposal Network to generate region proposals. The input image is passed through CNN which generates a feature map. On these feature maps, the Region Proposal Network is applied and it gives the object proposals as an

output. Apply RoI pooling to convert the proposal of the same size and pass these proposals to the Fully Connected Layer, it classifies object using Softmax Layer and generating a bounding box using Linear Regression Layer.

Mask R-CNN[24] is an extension of Faster R-CNN. Fig. 3 shows the steps of Mask RCNN. Mask RCNN use ResNet 101 architecture to extract features from the image. A Region Proposal Network (RPN) is performed on the extracted feature maps. RPN predicts whether the object is present inside the region or not. Region Proposal Network returns those feature maps or regions that contains the object that is predicted by the model. RoI alignment is used to fix the misalignment problems and keeps an exact spatial location. It converts cells of the same size but does not digitize the boundary of the cell[25]. To achieve a better feature map, it applies interpolation. RoI alignment improves accuracy. After applying RoI alignment, these regions are fed into Fully Connected Layers to perform the classification and predict the boundary box using the Softmax and Regression model respectively and apply two more convolution layers (as in Fig. 3) after RoI align in order to generate object masks.

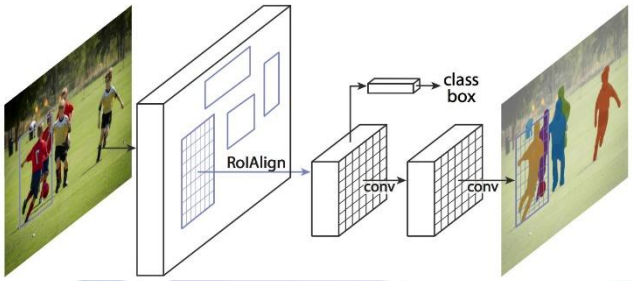

Fig. 3 Mask RCNN framework[26]

Mask R-CNN output for each candidate object: class label, bounding-box and the object mask. Fig. 4 shows the output of Mask RCNN.

Fig. 4 Image Output of Mask RCNN[6]

IV. DATASET DETAILS

AAU-RainSnow[27] dataset is used for vehicle detection and classification. It contains 5 different types of vehicles such as a car, bus, truck, bicycle, luxury bus and contains 22 different road videos. IITM-HeTra[6] dataset contains different types of images of vehicles in order to detect vehicles. It has 1417 images with different classes including car, bus, motorcycle, bicycle, and auto-rickshaw. PASCAL VOC[13] dataset has around 10000 images that contain 20 different classes with a few relevant classes such as a car, truck, and bus. Kitti_drive0005[12] dataset is used for object detection. BrnoCompSpeed[28] includes a dataset of 18 full-HD videos.

The commonly used metric to measure object detection accuracy is mean Average Precision (mAP)[15]. This is calculated by taking Average Precision (AP) of each class and calculating the average of all the Average Precisions. False Positive Rate (FPR) is also used as the evaluation criteria for vehicle detection[10]. Sometimes, Precision, Recall, and Accuracy metrics are also used to measure the performance of vehicle type classification[9].

V. CONCLUSION

This paper presented a review of various methods used for Vehicle Detection and Classification. Various Deep Learning methods are discussed with its challenges. This variety of Object Detection methods comes with certain merits and demerits. Among these methods, Mask RCNN provides better results along with the object mask by using different dataset images.

REFERENCES

[1] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object Detection With Deep Learning: A Review," IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 11, pp. 3212–3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.

[2] Vehicle Detection and Tracking Techniques : A Concise Review." [Online]. Available: https://www.researchgate.net/publication/307668193_Vehicle_Detection_and_Tracking_Techniques_A_Concise_Review. [Accessed: 06-Jan-2020]

[3] "Object detection," Wikipedia. 26-Dec-2019.

[4] S. Mahapatra, "Why Deep Learning over Traditional Machine Learning?," Medium, 22-Jan-2019. [Online].

[5] D. Mittal, A. Reddy, G. Ramadurai, K. Mitra, and B. Ravindran, "Training a deep learning architecture for vehicle detection using limited heterogeneous traffic data," in 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, 2018, pp. 589–294, doi: 10.1109/COMSNETS.2018.8328279.

[6] "IITM-HeTra." [Online]. Available: https://kaggle.com/deepak242424/iitmhetra. [Accessed: 25-Nov-2019].

[7] S. Kul, S. Eken, and A. Sayar, "A concise review on vehicle detection and classification," in 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, pp. 1–4, doi: 10.1109/ICEngTechnol.2017.8308199.

[8] C. N. Aishwarya, R. Mukherjee, and D. K. Mahato, "Multilayer vehicle classification integrated with single frame optimized object detection framework using CNN based deep learning architecture," in 2018 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, 2018, pp. 1–6, doi: 10.1109/CONECCT.2018.8482366.

[9] B. Hicham, A. Ahmed, and M. Mohammed, "Vehicle Type Classification Using Convolutional Neural Network," in 2018 IEEE 5th International Congress on Information Science and Technology (CiSt), 2018, pp. 313–316, doi: 10.1109/CIST.2018.8596500.

[10] M. V., V. V. R., and N. A., "A Deep Learning RCNN Approach for Vehicle Recognition in Traffic Surveillance System," in 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0157–0160, doi: 10.1109/ICCSP.2019.8698018.

[11] K. Shi, H. Bao, and N. Ma, "Forward Vehicle Detection Based on Incremental Learning and Fast R-CNN," in 2017 13th International Conference on Computational Intelligence and Security (CIS), 2017, pp. 73–76, doi: 10.1109/CIS.2017.00024.

[12] "The KITTI Vision Benchmark Suite." [Online]. Available: http://www.cvlibs.net/datasets/kitti/raw_data.php. [Accessed: 29-Dec-2019].

[13] "The PASCAL Visual Object Classes Challenge 2012 (VOC2012)." [Online]. Available: http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html. [Accessed: 29-Dec-2019].

[14] B. Benjdira, T. Khursheed, A. Koubaa, A. Ammar, and K. Ouni, "Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3," ArXiv181210968 Cs, Dec. 2018.

[15] G. Prabhakar, B. Kailath, S. Natarajan, and R. Kumar, "Obstacle detection and classification using deep learning for tracking in high-speed autonomous driving," in 2017 IEEE Region 10 Symposium (TENSYMP), Cochin, India, 2017, pp. 1–6, doi: 10.1109/TENCONSpring.2017.8069972.

[16] M. C. Olgun, Z. Baytar, K. M. Akpolat, and O. KoraySahingoz, "Autonomous Vehicle Control for Lane and Vehicle Tracking by Using Deep Learning via Vision," in 2018 6th International Conference on Control Engineering Information Technology (CEIT), 2018, pp. 1–7, doi: 10.1109/CEIT.2018.8751764.

[17] "Detecting and Counting the Moving Vehicles Using Mask R-CNN - IEEE Conference Publication." [Online]. Available: https://ieeexplore.ieee.org/document/8908877. [Accessed: 07-Jan-2020].

[18] Y. He and L. Li, "A Novel Multi-source Vehicle Detection Algorithm based on Deep Learning," in 2018 14th IEEE International Conference on Signal Processing (ICSP), 2018, pp. 979–982, doi: 10.1109/ICSP.2018.8652388.

[19] "A Multiple Vehicle Tracking and Counting Method and its Realization on an Embedded System with a Surveillance Camera - IEEE Conference Publication." [Online]. Available: https://ieeexplore.ieee.org/document/8448869. [Accessed: 24-Dec-2019].

[20] Prabhu, "Understanding of Convolutional Neural Network (CNN) — Deep Learning," Medium, 21-Nov-2019. [Online]. Available: https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148. [Accessed: 29-Dec-2019].

[21] "Zero to Hero: Guide to Object Detection using Deep Learning: Faster R-CNN,YOLO,SSD," CV-Tricks.com, 28-Dec-2017. [Online]. Available: https://cv-tricks.com/object-detection/faster-r-cnn-yolo-ssd/. [Accessed: 08-Jan-2020].

[22] "Object Detection for Dummies Part 3: R-CNN Family," Lil'Log, 31-Dec-2017. [Online]. Available: https://lilianweng.github.io/2017/12/31/object-recognition-for-dummies-part-3.html. [Accessed: 29-Dec-2019].

[23] "A Step-by-Step Introduction to the Basic Object Detection Algorithms (Part 1)," Analytics Vidhya, 11-Oct-2018. [Online]. Available: https://www.analyticsvidhya.com/blog/2018/10/a-step-by-step-introduction-to-the-basic-object-detection-algorithms-part-1/. [Accessed: 24-Dec-2019].

[24] "Step-by-Step Implementation of Mask R-CNN for Image Segmentation." [Online]. Available: https://www.analyticsvidhya.com/blog/2019/07/computer-vision-implementing-mask-r-cnn-image-segmentation/. [Accessed: 25-Nov-2019].

[25] J. Hui, "Image segmentation with Mask R-CNN," Medium, 21-Feb-2019. [Online]. Available: https://medium.com/@jonathan_hui/image-segmentation-with-mask-r-cnn-ebe6d793272. [Accessed: 25-Nov-2019].

[26] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.

[27] "AAU RainSnow Traffic Surveillance Dataset." [Online]. Available: https://kaggle.com/aalborguniversity/aau-rainsnow. [Accessed: 25-Nov-2019].

[28] Jakub Sochor et al., BrnoCompSpeed: Review of Traffic Camera Calibration and Comprehensive Dataset for Monocular Speed Measurement. arXiv.org, 2017.