International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT2174106

An Efficient Road Surveillance Approach to Detect, Recognize & Tracking Vehicles Using Deep Learning Methods

Vinod Kumar Yadav*1, Dr. Pritaj Yadav2, Dr. Shailja Sharma2
*1PHD (Research Scholar), Computer Science & Engineering Department, RNTU, Bhopal, Madhya Pradesh, India
2Associate Professor, Computer Science & Engineering Department, RNTU, Bhopal, Madhya Pradesh, India

Article Info
Volume 7, Issue 4
Page Number: 503-512

Publication Issue :
July-August-2021

Article History
Accepted : 07 Aug 2021
Published : 14 Aug 2021

ABSTRACT

In the current scenario on the increasing number of motor vehicles day by day, so traffic regulation faces many challenges on intelligent road surveillance and governance, this is one of the important research areas in the artificial intelligence or deep learning. Among various technologies, computer vision and machine learning algorithms have the most efficient, as a huge vehicles video or image data on road is available for study. In this paper, we proposed computer vision-based an efficient approach to vehicle detection, recognition and Tracking. We merge with one-stage (YOLOv4) and two-stage (R-FCN) detectors methods to improve vehicle detection accuracy and speed results. Two-stage object detection methods provide high localization and object recognition precision, even as one-stage detectors achieve high inference and test speed. Deep-SORT tracker method applied for detects bounding boxes to estimate trajectories. We analyze the performance of the Mask RCNN benchmark, YOLOv3 and Proposed YOLOv4 + R-FCN on the UA-DETRAC dataset and study with certain parameters like Mean Average Precisions (mAP), Precision recall.

Keywords : Deep Neural network, CNN, R-FCN, YOLO, YOLOv4, Deep-Sort Tracking.

I. INTRODUCTION

The Neural networks model was develop by the motivation with biological features of the human brain for that reason the name is neural network. It refers to mathematical functions of neural networks, convolutional neural networks (CNNs), and deep neural networks (DNNs) have lead to various successes in computer vision. Computer vision for Vehicle detection, recognition and tracking plays an important role in intelligent traffic surveillance in Deep learning study. The objective of road surveillance and monitoring systems is to reduce need of human labour for collecting traffic data, counting the vehicles, or implement roadside questionnaires vision based tasks that can performed by a computerized process. The computer vision systems have also applied in different public areas such as roads, agriculture, airports, and retail areas. One such application of computer vision systems is in the task of monitoring and analysing scenes of road surveillance. Camera-based surveillance techniques for road or traffic surveillance have become more common. Cameras are cost-effective, ubiquitous, and easy maintenance. The height of computer vision techniques was holdup of camera-based road or traffic monitoring systems for a long time. In real-world road surveillance for detection, recognition and tracking of vehicles in scenes are required in all-weather conditions at all times. We analyse the performance of the Mask RCNN, YOLOv3 and Proposed YOLOv4 + R-FCN on the UA-DETRAC [24] dataset for vehicle detection and tracking and evaluate certain parameters like Intersection over Union, Precision-Recall and Mean Average Precision. This paper has organized as follows: Section II: Literature Survey describes some research work related to our project. Section III: Describe proposed methodology. Section IV: describes implementation and setup of the project. Section V: Results Analysis discussed. and end with conclusion in section VI.

II. LITERATURE SURVEY

Vehicle Detection: Detection of target in images or video is a difficult, due to the nature of objects in images is those are often of unusual sizes, different orientations, and partly cover target that causes occlusion of the object of interest to be detected. These problems solution require a detection algorithm that has several properties, such as different positions of object of interest in the image, rotation of the object in the image, and scale invariance. From the Traditional or beginning of vehicle detection, researchers have proposed several traditional vehicle detection methods. Handcrafted features are used to determine the performance of methods. Histogram of Oriented Gradient (HOG) [32] and Haar-like [21] features are the most common used features. One of the earliest real-time detectors is a cascaded detector [34], which achieves competitive accuracy. Deformable part-based models (DPM) and Support Vector Machines (SVM) [35] are two well-known models of the part-based model approach. CNN has achieved good results in object detection; that has a strong capability to learn image features and can achieve multiple related tasks, such as classification and bounding box regression [3]. The two-stage object detection methods generate a candidate box of the object via a selection of algorithms and then classify the object by a convolutional neural network but one-stage method does not generate a candidate box although straight converts the positioning problem of the object bounding box into a regression problem processing. In two-stage object detection method, Region-CNN (R-CNN) [25] uses selective region search [29] in the image or video input of the convolutional network must be fixed-size and the deep organization of the network time-consuming for training time and consumes a large amount of storage memory. SPP NET [26] allows the system to input images of different sizes and to have fixed outputs. R-FCN, FPN, and Mask RCNN are two stage methods that improved the feature extraction methods, feature selection, and classification capabilities of convolutional networks in particular behaviour. A CNN model proposed by Sanjay Saini [18] that handles the traffic light detection for autonomous vehicles, to address the problem of large variation of light. This model takes image as input data, extracts candidate region, and performs final traffic light detection and recognition. Phan[12] proposed their method to handle the dense occlusion from static surveillance cameras, that consists of background subtraction, occlusion and vehicle detection; it is to extract the occluded vehicles independently based on the external properties. This proposed model get better accuracy of vehicle detection from 61.19% to 84.09% compared to Ha's method [22]. Lu [5] proposed Region Proposal Network (RPN) known as a scale-aware RPN, to address the problem of detecting vehicles at different scales. The scale-aware RPN consists of two sub-networks: one detects large feature and another detects small feature, and then feeds features into two separate XGBoost [19] classifiers to make final predictions.

Among the one-stage methods, the most important are the Single Shot Multibox Detector (SSD) [17] and You Only Look Once (YOLO) [18] frameworks. The MutiBox [27], Region Proposal Network (RPN) and multi-scale representation methods are use in SSD, which uses a default set of anchor boxes with different aspect ratios to more accurate position the object. YOLO is a one-stage object detection method that proposed by Redmon et al. after Faster RCNN. YOLO network first step calculates floating number coordinates of an object in the feature map. Secondly, YOLO frames detection as a regression problem, that architecture can extract features from input images straight to predict bounding boxes and class probabilities.

YOLOv2 [17] is a next version of YOLO, which works with new concepts to improve YOLOs speed and precision. YOLOv2 include the Batch Normalization layer, which makes the network normalize the input of each layer and accelerate the network convergence speed. YOLOv2 method uses a multi-scale training technique select at random a new image size for each ten batches. It is normalize the complete training set because the optimization step uses stochastic gradient descent. As SGD uses mini-batches during training, each mini batch produces approximate of the mean and variance of activation. Calculate the mean and variance value of the mini-batch of size m, and then normalize the activations of number m to have mean zero and variance one. Lastly, the elements of each mini-batch sampled from the similar distribution. YOLOv2 insert a BN layer [6] in front of each convolutional layer that accelerates the network to get convergence and helps regularize the model. Batch normalization gets more than 2% improvement in mAP High Resolution Classifier.YOLOv2 concatenates with Fine-Grained Features that higher resolution features with the low resolution features by stacking adjacent features into different channels which gives a modest 1% performance increase Multi-Scale Training.

YOLOv3 [4] is an improved version of YOLOv2. YOLOv3 method use multi-label classification to adjust more complex datasets that have many overlapping label. That utilizes three different scale feature maps as 3-d tensor encoding class predictions, object and bounding box to predict the bounding box. YOLOv3 recommend a deeper and strong feature extractor, called Darknet-53 that inspired by ResNet. YOLOv4 is the latest and most advanced iteration method till date. It has excellent speed for use in invention systems and for optimization in parallel computations. YOLOv4 are include some new techniques like Weighted-Residual-Connections, Cross-Stage-Partial-Connections, Cross mini-batch, Normalization (CmBN), Self-adversial-training, Mish-activation, etc. to obtain higher values for accuracy.

Vehicle tracking: Vehicle tracking means assign the same vehicle through multiple successive frames that contain various methods for trajectory assignment, motion modelling, tracking result filtering, and at last vehicle counting. Better multi-object tracking, are also a critical ITS task [28]. MOT (multi-object tracking) methods use Detection-Based Tracking (DBT) and Detection-Free Tracking (DFT) for object initialization. DBT technique uses background-modelling to detect moving objects in video frames before tracking objects. DFT method wants to initialize the object for the tracking but cannot handle the addition of new objects and the departure of old objects. The Bhattacharyya [30] distance is used to calculate the distance of the color histogram between the objects. Presently, detection-level or trajectory-level exclusion can solve this problem by scale changes and illumination changes of moving objects, used SIFT feature [31] points for object tracking, even this is slow. The ORB feature point detection algorithm [29] better extraction feature points at a significantly higher speed than SIFT. IOU tracker is built on the hypothesis of every object is tracked on a per-frame such that there are none or very few gaps present in between detection [7]. In the same way, IOU assumes that there is a larger overlap value for intersection over union while obtaining object detection in successive frames. Similarly, Kalman-IOU (KIOU) tracking has Kalman filter's that capability of performing predictions allows users to skip frames while still keeping track of the object. Skipping frames permits the detector to speed-up the procedure as in a tracking-by-detection job, lesser number of frames wedges lower computational necessity. In the same way, this feature can also improve the performance of Kalman-IOU tracker compared to the IOU tracker. An implementation of tracking-by-detection framework is Simple Online and Real time Tracking (SORT) [20] where the main purpose is to detect objects in each frame and associate them for online and real-time tracking application. Kalman Filter and Hungarian algorithm are use for SORT. The feature of SORT is that it only uses detection information from the earlier and current frames, allow to proficiently performing online and real-time tracking. In Feature-based object tracking, extracts object features from one frame and then matches appearance information with successive frames based on the measure of similarity. The Simple Online and Real time tracking with a Deep Association metric (Deep SORT) make possible for multiple object tracking by integrating appearance information with its tracking components [13]. Kalman Filter and Hungarian algorithm combination are use for tracking. Kalman filtering is performing in image space while Hungarian techniques help frame-by-frame data association using an association metric that computes bounding box overlap. To obtain motion and appearance information, a trained deep convolutional neural network (DCNN) is applied.

Vehicle Recognition: Generally, recognition methods can be divided into two categories: older Hand-crafted feature engineering methods and newer Deep learning approaches. Hand-crafted methods based on human-engineered feature extraction channel to convert the image into a set of features that are tough to variations in both vehicle specific variables such as scale, location ,color and environment variables like pose, illumination and background. The large-scale MIO-TCD [8] dataset, deep learning has become the predominant approach for vehicle type recognition. Kim and Lim [14] choose a convolutional neural network of moderate size, and the samples augmented with flipping and rotations. Lee and Chung [15] proposed 12 local expert networks and 6 global networks, that local expert networks take the GoogLeNet structure, and each network trained on a subset of training samples. The dataset is dividing in view of the resolution and aspect ratio of samples. In [16], two neural networks trained independently with the weighted cross-entropy loss function. Both models based on ResNet, and they differ in the number of layers. Rachmadi et al. [9] introduces a Pseudo Long Short-Term Memory (P-LSTM) classifier for identifying a single image. Xiang et al. [2] suggest a four-stage pipeline that takes the interaction between parts into account. Part detection implemented using a backbone model truncated at an intermediate layer, while part bring together involves point wise convolutional layers that gather associated parts into the same feature map. Afterward, topology constraint comprises depth wise convolutional layers and estimates the probability of the topology relationship between related parts. The finale classification uses a fully connected layer to make predictions.

III. PROPOSED METHODOLOGY

With merge of one-stage (YOLOv4) and two-stage (R-FCN) detectors methods to improve vehicle detection accuracy and speed results. Two-stage detectors have high localization and object recognition precision, while one-stage detectors achieve high inference and test speed. Completely inherit the advantages of one-stage and two-stage detectors while overcoming their disadvantages, this achieves better accuracy than two-stage detectors and maintains comparable efficiency of one-stage detectors. Deep-SORT [13] tracker applied on detected bounding boxes to estimate trajectories. First, images and annotation files are passing into the pre-processing stage. The pre-processed images are passed through a YOLOv4 model that consists of Backbone: CSPDarknet53 [1], Neck: SPP [23], PAN [6] Head: YOLOv3 [4]) + R-FCN for vehicle detection. We are using these bounding box locations fed to Deep-SORT tracker, which gives us the trajectory of each vehicle.

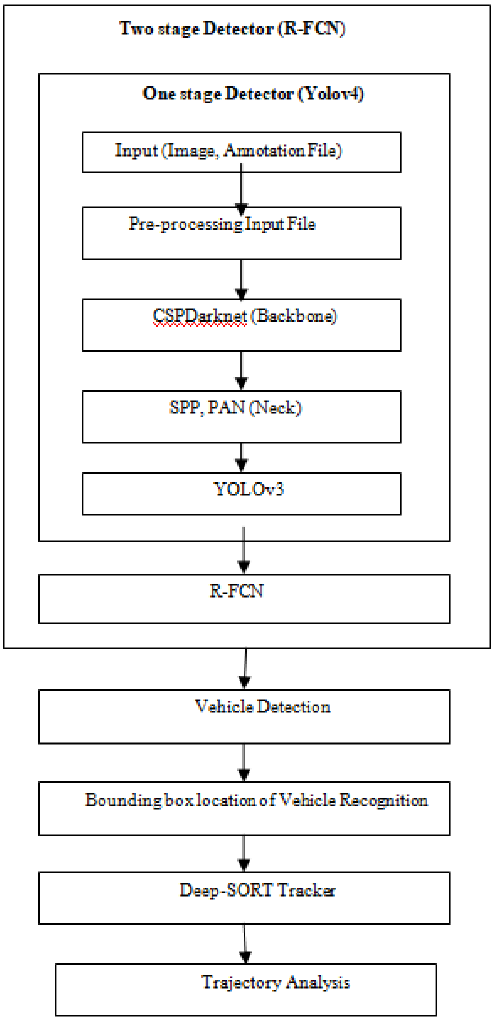

Figure 1 : Block Diagram of Proposed Methodology

IV. IMPLEMENTATION SETUP

We perform experiments on the UA-DETRAC dataset and evaluate some parameters like IoU, mAP. All the training and testing phases were performed on a laptop with Intel Core i7-6700HQ 2.8GHz CPU, GPU GTX 1050m 4 GB graphics, RAM 8 GB, Memory 500 GB drive.

A. Dataset

UA-DETRAC dataset have multiple vehicles, multiple views, weather, scale and light. The dataset consists of 140232 frames in total out of which 138353 frames labelled and have annotations. For the simplicity of algorithm and experiment, we only consider two categories namely “Car” and “Bus”. All the bounding boxes from ground truth data that have “van” or “others” as their label, they ignored. There are 70% sequences of data use for training and 30% sequences in testing. As working with UA-DETRAC dataset, we found missing annotations for several images. In the testing portion, we found missing annotations only for sunny and night data images. The sequence number and frame number details of these images for test portion have provided in Table I. In the training section, we found some missing annotations for night, cloudy and rainy data images. The sequence number and frame number details of these images for training portion have provided in Table II. No missing annotation has observed in Sunny data images in training portion.

Weather	ID	Frame numbers
Cloudy	MVI_38966	679-935, 987-1015
	MVI 40462	1617-1620
	MVI 40565	333-370
	MVI 40614	233-245, 527-530, 547-555
	MVI 40716	1167-1170, 1287-1290
	MVI 39861	542-610, 897-1080, 1297-1380

Night	MVI 39981	337-340
	MVI 39911	207-260, 342-510, 597-655
	MVI 40951	169-315
	MVI 41291	282-345, 1667-1755
Rainy	MVI 63844	992-1195

Table I. Distribution of images whose annotations are missing in training portion

Weather	ID	Frame numbers
Night	MVI 40858	1635
	MVI 41087	1752-1765
	MVI 40993	1970
Sunny	MVI 39251	996-1066
	MVI 39223	479-579
	MVI 39558	87-90

Table II. Distribution of images whose annotations are missing in testing portion

B. YOLOv4 + R-FCN Implementation

We decided to run YOLOv4 + R-FCN using OpenCV deep neural networks (DNN) module because it gives much more flexibility and control on the information returned by proposed algorithm like bounding box locations, confidence scores etc. We utilized pre-trained YOLOv3 weights and configuration file into OpenCV Deep neural network (DNN) module. Figure 2 shows qualitative results of applying YOLOv4 + R-FCN on a road image.

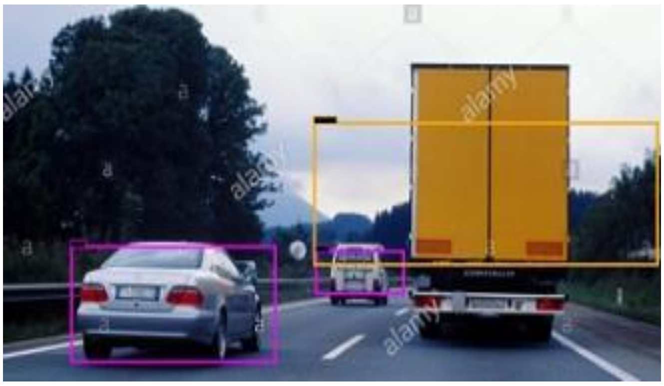

Figure 2 : Applying YOLOv4 + R-FCN on a pre-processed image

C. Deep-SORT Tracker

The Deep-SORT tracking method cannot work without the results from the vehicle detection stage. The YOLOv4 + RFCN detection performed to get information regarding the bounding box, the class, and the features of detected objects. Next, this information about each object feed to the Deep-SORT module through a queue. Deep-SORT process, there are two inputs that are the video frame and the information of the detected objects, respectively. After pre-processing processes, the Deep-SORT module removes the overlapping bounding boxes. The objective of the tracking is to assign the detected bounding boxes with the object IDs already appearing in the earlier frame. If a bounding box cannot assigned to any earlier object ID, a new object ID will assigned to this bounding box. Deep-SORT tracks objects by two metrics that are the location metric and the appearance metric. The detected object is tracked using the IoU parameter, If the bounding box cannot be allot to any object ID using the IoU parameter, the bounding boxes will be processed in the features component. In checking component, if the bounding box already assigned to an object ID, the information concerning the object ID and corresponding bounding box will sent to the resolution module.

V. RESULTS AND DISCUSSIONS

Figure 2 show the results of YOLOv4 + RFCN algorithms on a challenging real scenario road image from rainy portion of UA DETRAC test dataset. Black boxes indicate ground truth objects and in total, there are 22-labelled cars, 1 bus and 1 van according to ground truth annotation file. YOLOv4 + RFCN detect 21 cars in total one of them is actually a van, which considered as car by YOLOv4 + RFCN algorithm. As a result, YOLOv4 + RFCN fail to detect the bus and 2 cars. We evaluate algorithms on several parameters that have detailed as follows.

A. Intersection over Union (IoU)

First, we measure the area of intersection between predicted and ground truth bounding boxes. IoU is computed by taking the ratio of Area of intersection by total area of boxes. We determine Intersection over union values with respect to all ground truth-bounding boxes in each frame and take arithmetic mean of all of them.

B. Precision Recall

Precision and recall rates were used as the measure of the system's overall performance. A true positive (TP) represents a detected and correctly classified vehicle that has a corresponding manually tagged object in the test database. A false positive (FP) represents a detected and classified vehicle that has no matching tagged object in the test database. A detected but not classified vehicle denoted as a false positive even if it had a corresponding manually tagged object. A false negative (FN) represents objects that missed by the vision system. Precision is the ability of the model to identify correct positive predictions.

$$Precision=TP/ TP+ FP$$

Table III : Average Precision (AP) compares with YOLOv3, Mask RCNN and Proposed Method for different IoU thresholds (0.6, 0.7, 0.8 and 0.9) weathers on UA-DETRAC dataset

Method	Vehicle	Weather Data	0.6	0.7	0.8	0.9
YOLOv3	CAR	Cloudy	0.653	0.543	0.299	0.021
		Night	0.654	0.543	0.247	0.012
		Rainy	0.544	0.439	0.194	0.008
		Sunny	0.625	0.507	0.240	0.008
	BUS	Cloudy	0.382	0.255	0.113	0.006
		Night	0.802	0.736	0.496	0.079
		Rainy	0.578	0.444	0.119	0.012
		Sunny	0.501	0.399	0.185	0.005
Mask-RCNN	CAR	Cloudy	0.658	0.593	0.445	0.083
		Night	0.684	0.611	0.394	0.056
		Rainy	0.585	0.503	0.336	0.047
		Sunny	0.667	0.604	0.429	0.049
	BUS	Cloudy	0.566	0.488	0.320	0.022
		Night	0.712	0.561	0.360	0.061
		Rainy	0.578	0.445	0.334	0.002
		Sunny	0.505	0.430	0.302	0.030
Proposed Method (YOLOv4 + R-FCN)	CAR	Cloudy	0.661	0.578	0.453	0.076
		Night	0.692	0.621	0.401	0.063
		Rainy	0.598	0.512	0.341	0.049
		Sunny	0.672	0.618	0.436	0.051
	BUS	Cloudy	0.577	0.493	0.332	0.023
		Night	0.819	0.572	0.319	0.081
		Rainy	0.596	0.452	0.314	0.016
		Sunny	0.665	0.445	0.299	0.021

Recall is the ability of object detector to find all relevant ground truths.

$$Recall= TP/ TP+ FN$$

C. Average Precision (AP)

Average precision is corresponds to area under the precision-recall for a particular class. Total areas of these rectangles correspond to average precision (AP). Tables III present the average precision values for different weathers at different IoU thresholds using both techniques.

D. Mean Average Precision (mAP)

Mean average precision is simply an arithmetic mean of average precisions for all classes. The evaluation protocol mAP at 0.8 IoU threshold is considered as evaluation metric. We take the arithmetic mean of average precision at 0.8 IoU thresholds for all weathers. It is absolutely clear that YOLOv4 + R-FCN in all weather conditions in terms of accuracy.

E. Tracking Evaluation

The performance of tracker analysed by the reporting mean absolute error (RMSE) between detected trajectory and ground truth trajectory. For each trajectory point, we calculate the Euclidean distances between point of trajectory and ground truth points of all frames corresponding to the trajectory. The minimum distance corresponds to the loss for this trajectory point. Hence, total loss of each frame is calculated by the sum of losses for all trajectory points and can be described by the following equation:

$$l = \sum_{c=1}^{M} min(\sqrt{(x_c - x_g) + (y_c - y_g)})_{g=1}^{N}$$

where, M is the number of trajectory points in a frame and N is the total number of frames The total loss of tracker is the sum of losses for all frames.

$$L = \sum_{i=1}^{N} l_i$$

The results reported in Table IV. The values represent the root mean square distance error per detection in units of pixels between detected and ground truth trajectories for different weathers on UA-DETRAC dataset.

Table IV: RMS error per detection in units of pixels between detected and ground truth trajectories for different weathers on UA-DETRAC dataset

Tracker	Detector	Cloudy	Night	Rainy	Sunny
SORT	YOLOv3	15	16	20	16
	Mask RCNN	13	14	15	13
Deep-SORT	Proposed Method	**17**	**17**	**22**	**18**

VI. CONCLUSIONS AND FUTURE SCOPE

In this paper proposed a new computer vision-based an efficient approach to vehicle detection, recognition and Tracking. We merge with one-stage (YOLOv4) and two-stage (R-FCN) detectors methods to improve vehicle detection accuracy and speed then tracking results from the Deep-SORT. Proposed detection could miss the vehicle object so that no detected bounding boxes will forwarded to the Deep-SORT component, which leads to no tracking results from the Deep-SORT. When the YOLOv4 + R-FCN detect again these objects in subsequent frames, the objects will assignee a new object ID. The next contribution of this paper concerns the operating speed. In our proposed architecture, any detected object from YOLOv4 + R-FCN is sent immediately to the Deep-SORT tracker. Hence, vehicle detection and tracking conducted in similar. In addition, the vehicle appears in several frames. Therefore, the vehicle detection in the proposed model conducted in some frames only so that we can reduce the YOLOv4 detection time and R-FCN improve vehicle detection accuracy. This research tests the proposed architecture with different videos to examine different weather situations. The accuracy of the tracking still depends on the YOLOv4 + R-FCN detection.

VII. REFERENCES

[1]. Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping,Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh., “CSPNet: A new backbone that can enhance learning capability of cnn”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2020.

[2]. Xiang, Y., Fu, Y., Huang, H.,”Global topology constraint network for fine-grained vehicle recognition”, IEEE Transactions on Intelligent Transportation Systems, 2019.

[3]. Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X., “Object detection with deep learning: A review.” arXiv e-prints, arXiv: 1807.05511, 2018.

[4]. Redmon, J., & Farhadi, A., “Yolov3: An incremental improvement”, arXiv preprint arXiv: 1804.02767,2018.

[5]. Ding Lu, Yong Wang, Robert Laganière, Xinbin Luo, and Shan Fu. "Scale-Aware RPN for Vehicle Detection." In International Symposium on Visual Computing, pp. 487-499. Springer, Cham, 2018.

[6]. Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia., “Path aggregation network for instance segmentation”, In Proceedings of the IEEE Conference on Computer Vision and

Pattern Recognition (CVPR), pages 8759–8768, 2018.

[7]. E. Bochinski, T. Senst, and T. Sikora, "Extending IOU based multi-object tracking by visual information." In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1-6. IEEE, 2018.

[8]. Luo, Z., Branchaud-Charron, F., Lemaire, C., Konrad, J., Li, S., Mishra, A., Achkar, A., Eichel, J., Jodoin, P.M.," MIOTCD: a new benchmark dataset for vehicle classification and localization", IEEE Transactions on Image Processing, 27(10), 5129–5141, 2018.

[9]. Rachmadi, R.F., Uchimura, K., Koutaki, G., Ogata, K.," Single image vehicle classification using pseudo long short-term memory classifier", Journal of Visual Communication and Image Representation, 2018,56, 265–274.

[10]. Redmon, J., & Farhadi, A.," Yolo9000: Better, faster, stronger: IEEE", 2017.

[11]. Sanjay Saini, S. Nikhil, Krishna Reddy Konda, Harish S. Bharadwaj, and N. Ganeshan. "An efficient vision-based traffic light detection and state recognition for autonomous vehicles." In 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 606-611. IEEE, 2017.

[12]. Hung Ngoc Phan, Long Hoang Pham, Duong Nguyen-Ngoc Tran, and Synh Viet-Uyen Ha. , "Occlusion vehicle detection algorithm in crowded scene for Traffic Surveillance System." In 2017 International Conference on System Science and Engineering (ICSSE), pp. 215-220. IEEE, 2017.

[13]. N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in 2017 IEEE international conference on image processing (ICIP). IEEE, 2017, pp. 3645–3649.

[14]. Kim, P.K., & Lim, K.T., "Vehicle type classification using bagging and convolutional neural network on multi view surveillance image". In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, (pp. 41–46).

[15]. Taek Lee, J., & Chung, Y., "Deep learning-based vehicle classification using an ensemble of local expert and global networks. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, (pp. 47–52).

[16]. Theagarajan, R., Pala, F., Bhanu, B.,"EDen: ensemble of deep networks for vehicle classification". In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 33–40.

[17]. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C. "Ssd: Single shot multibox detector, In 2016 European conference on computer vision",: Springer International Publishing, 2016 ,pp. 21–37.

[18]. Redmon, J., Divvala, S., Girshick, R., Farhadi, A., "You only look once: Unified, real-time object detection", In 2016 IEEE conference on computer vision and pattern recognition, 2016, p.p. 779–788: IEEE.

[19]. Tianqi Chen and Carlos Guestrin., "A scalable tree boosting system." In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. ACM, 2016.

[20]. A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking." In 2016 IEEE International Conference on Image Processing (ICIP), pp. 3464-3468. IEEE, 2016.

[21]. Xuezhi Wen, Ling Shao, Wei Fang, and Yu Xue. "Efficient feature selection and classification for vehicle detection." IEEE Transactions on Circuits and Systems for Video Technology, 25, no. 3, 2015: 508-517.

[22]. Synh Viet-Uyen Ha, Long Hoang Pham, Hung Ngoc Phan, and Phong Ho-Thanh., "A robust algorithm for vehicle detection and

classification in intelligent traffic system." In 16th Asia Pacific Industrial Engineering & Management Systems Conference (APIEMS 2015), no. 1, pp. 1832-1838. 2015.

[23]. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition". IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 37(9):1904–1916, 2015.

[24]. L. Wen, D. Du, Z. Cai, Z. Lei, M.-C. Chang, H. Qi, J. Lim, M.-H. Yang, and S. Lyu, "Ua-detrac: A new benchmark and protocol for multi-object detection and tracking," arXiv preprint arXiv: 1511.04136, 2015.

[25]. Girshick, R., Donahue, J., Darrell, T., Malik, J." Rich feature hierarchies for accurate object detection and semantic segmentation", IEEE Conference on Computer Vision and Pattern Recognition, 2014.81: IEEE.

[26]. Kaiming, H., Xiangyu, Z., Shaoqing, R., Jian, S. "Spatial pyramid pooling in deep convolutional networks for visual recognition". IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9), 1904–16.

[27]. Erhan, D., Szegedy, C., Toshev, A., Anguelov, D., "Scalable object detection using deep neural networks", In 2014 IEEE conference on computer vision and pattern recognition. https://doi.org/10.1109/cvpr., 2014, pp. 2147–2154, IEEE.

[28]. Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., Zhao, X., Kim, T.K., "Multiple object tracking: A literature review". arXiv preprint arXiv:1409.7618., 2014.

[29]. Uijlings, J.R.R., van de Sande, K.E.A., Gevers, T., Smeulders, A.W.M."Selective search for object recognition. International Journal of Computer Vision", 2013, 104(2), 154–171.

[30]. Xing, J., Ai, H., Lao, S.,"Multi-object tracking through occlusions by local track lets filtering and global track lets association with detection responses", In 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1200–1207.

[31]. Zhou, H., Yuan, Y., Shi, C., "Object tracking using sift features and mean shift. Computer Vision & Image Understanding", 2009 113(3), 345–352.

[32]. Pedro F. Felzenszwalb, David A. McAllester, and Deva Ramanan. "A discriminatively trained, multiscale, deformable part model." In Computer Vision and Pattern Recognition (CVPR),2008, vol. 2, no. 6, p. 7.

[33]. Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection." In International Conference on Computer Vision & Pattern Recognition (CVPR'05), 2005, vol. 1, pp. 886-893. IEEE Computer Society.

[34]. Paul Viola and Michael Jones. "Robust real-time object detection." International Journal of Computer Vision, 4, no. 34-47, 2001: 4.

[35]. ohan AK. Suykens, and Joos Vandewalle. "Least squares support vector machine classifiers." Neural Processing Letters, 9, no. 3, 1999: 293-300.

Cite this article as :

Vinod Kumar Yadav, Dr. Pritaj Yadav, Dr. Shailja Sharma, "An Efficient Road Surveillance Approach to Detect, Recognize & Tracking Vehicles Using Deep Learning Methods", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 4, pp. 503-512, July-August 2021. Available at
doi : https://doi.org/10.32628/CSEIT2174106
Journal URL : https://ijsrcseit.com/CSEIT2174106