Real Time Vehicle Detection Using Deep Neural Networks: A Survey

[1]Spoorthi D N, [2]Ashwini B P

[1]Student, [2]Assistant Professor
Department of Computer Science and Engineering,
Siddaganga Institute of Technology, Tumakuru, Karnataka, India - 572103

Abstract: —Today, safety is one of the important key factors for everyone. With the increase in the number of automobiles on the urban roads, highways face many challenges in traffic regulations. Advanced Driver Assistance Systems (ADAS) provides more optimistic way in controlling and managing the accidents and road traffic. It also ensures the safety and provides better driving experience. Vehicle detection is one of the basic and important components in the ADAS technology. Vehicle detection provides a necessary data to solve the problem of traffic management. Precisely detecting the vehicles in real time is difficult and also the challenging task in the field of research. Recently deep learning has manifest favorable in resolving the innumerable computer vision task such as object recognition, detection. In this survey, latest developments of the models for vehicle detection are discussed and compared based on the analysis of architecture and their performance.

IndexTerms **- Vehicle detection, deep learning, object detection, ADAS, traffic management.**

I. INTRODUCTION

Roadaccidentsarethecrucialcauseoflossworldwide.Alloverthe worldinayearitisapproximatedthatmorethan50million are bruised and 1 million people are dead due to road crashes. These accidents cause more economical burden, it includes damage in properties, hospitalization. Automotive assemblers, companies in the market makes the research and development in the fieldof ADAS. ADAS are the system helps the driver in day to day process, the goal of this system is to save the life of driver and pedestrians by giving earlycautions

The traditional approach [1] for target detection is divided into three phases, first few candidate regions are selected from an inputimageandthenfromtheseregionsextractthefeaturesatlastclassifyusingtrainedclassifiers.Toselectthecandidateregions, traditional approach uses the sliding window technique and those are not specific and even high in the time complexity. The features that are hand designed are not really great because of the diversity of background and light variations. Classifiers uses support vector machine (SVM) and Adaboost forclassification.

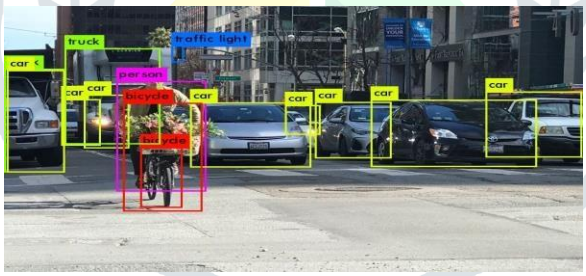

Figure 1: Vehicle detection

Due to the expeditious development in the field of deep learning [2], the classification and detection entered the new stage by encountered the bottleneck which are present in the traditional approaches. Deep learning has the level of consistency for the geometric contortions, varieties, light. It disposes of the disparity of vehicle presence for candidate determination and utilizations the preparation information to separate the highlights for greater adaptability

Framework for detecting the objects are categorized into two stage: two-stage approach and one-stage approach. In two-stage approach [1-3], networks generate a group of candidate object boxes using region proposals or selective search algorithm and then categorized and revert whereas in One-stage approach [4-6] the network generated the many samples over the aspect ratio, scales, locations; these will be classified and regressed at the same time. Real time is the major advantage of the one-stage approach. However,withtheadvantageofrealtimeinonestageapproachtheaccuracyiscomparativelylowthanthetwo-stageapproachdue to the class irregularity. You only look once (YOLO) [4], Single Shot multibox Detector (SSD) [6] are grouped under one-stage approachandRegionbasedConvolutionalneuralnetwork(R-CNN)[1],fastR-CNN[2],fasterR-CNN[3]areallcomesunderthe two-stageapproach.

Thispapercenteraroundthedeeplearningmodelsthatcandistinguishthevehiclesforbettertrafficobservationasappearedin **Fig.1** and furthermore, we are contrasting their execution dependent on mean average precision(mAP) and speed (Frame per second(FPS)).

II. RELATEDWORK

Ross Girshick et al. [3] proposed the paper "Rich feature hierarchies for accurate object detection and semantic segmentation" they propose a simple detection algorithm by integrating the two key insights: 1) In order to segment and localize the object one canapplyconvolutionalneuralnetworkwhicharehighcapacitytobottomupregionalproposal.2)Toyieldthenotablechangesin performance the labeled training data must be supervisely pre-trained followed by fine tuning. Since they integrated the regions proposalwithCNNtheycalledasR-CNN.Inthis,wepresentashortoverviewofR-CNNmethodanditworks.

RossGirshick[4]proposedthepaper"FastR-CNN",inthishepresentedanewmodelcalledFastRegion-basedConvolutional neural Network (Fast R-CNN) for detecting the objects. Fast R-CNN is built on R-CNN to classify the object proposal efficiently. It employs some innovations in order to increase the speed of training, testing and reliability. From this paper, focused on the innovations of Fast R-CNN and theirlimitations.

Shaoqing Ren et al. [5] proposed the paper "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", this proposed work mainly focused on detecting the vehicles in real time by introducing region proposal network (RPN). RPN enables the cost-free for region proposals. In this we present an overview frame work with RPN and workflow of faster R-CNN and also the results.

Joseph Redmon et al. [6] presented the paper "You Only Look Once: Unified, Real-Time Object Detection", in this paper they approach a new method for object detection called YOLO. This model forecast the multiple bounding boxes for each single grid cell. The final layer of this model will predict the coordinate for bounding box and also the class probabilities and the main advantageofthisisextremelyfastdetection,soitiswellsuitedforrealtime.Eventhoughitisfasttherearesomelimitations.From thispaper,overviewtheworkingofYOLOalongwithbothadvantagesandlimitations.

Joseph Redmon, Ali Farhadi [7] proposed the paper "YOLO9000: Better, Faster, Stronger", in this paper they introduced two realtimedetectionsystemnamelyYOLOv2andYOLO9000.TheymadeseveralimprovementsovertheexistingYOLOmethodin terms of prior art and novel. From this paper we overview the advantages, improved accuracy and the limitations of two real time detectionsystem.

III. MODELS FOR VEHICLEDETECTION

Vehicle detection has pulled in such a large number of research considerations and has been utilized in numerous zones like open wellbeing, traffic control. A portion of the modern object detectors are recorded here.

3.1 Region based Convolutional neural network(R-CNN)

To encounter the problem of selecting the vast region proposed the R-CNN [3] framework by using CNN and region proposals.Inthismethodtheyapplytheselectivesearchalgorithmtoaninputimagetoextractthe2000regionsthesearecalledas regionproposal.EachregioniswrappedandgiveasaninputimagetoCNNasshowninthe**Fig.2**.ProblemwiththeRCNNisthat ititerates2000timestotrainthenetworkforeach2000regions.HenceR-CNNhasbothspaceandtimecomplexity.Itisnotsuited for real time due to the slow detection ofobject.

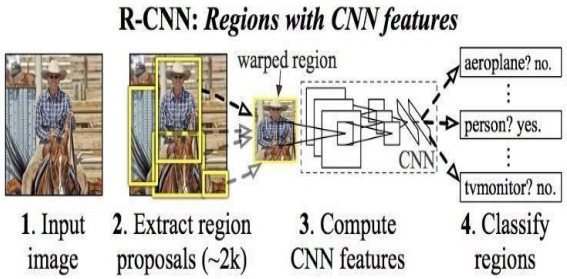

Figure 2: R-CNN architecture

3.2 Fast Region based Convolutional network

Fast R-CNN [4] model is almost similar to the R-CNN, as the name says it is faster than previous one. Here in this model it feeds the input image to the pre-trained CNN once instead of feeding the regional proposal every time and creates the convolutionalfeaturemapasshowninthe**Fig.3.**From this map it recognizes the regional proposal and bind them in a square using region of interest (RoI) pooling layer. Here there is a need to mold them in a size which are fixed to feed into the fully interconnected layer as an input. They used a softmax layer from RoI feature vector to predict the proposed region classes. The reason for fast computation than R-CNN is that it need not to feed each of 2000 region proposals every time to CNN instead just done with the convolutional operation once per image and generate the feature map and the offset value for boundary box regressor.

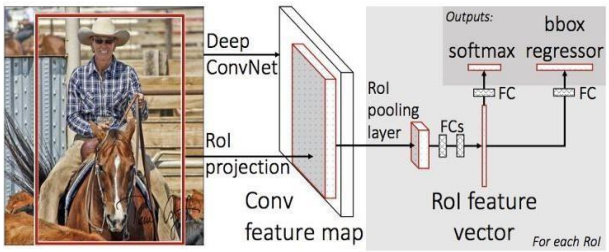

Figure 3: Fast R-CNN architecture

The main advantage of this model is higher and more accurate detection. Some other advantages are training is done at single stage and it update all the network layers and there is no need for disk storage to cache the features also reduces the mis localization of object Even though it is faster in terms of both testing and training but when comes to performance it is low due to the bottleneck of regional proposal. Same as R-CNN it also uses selective search algorithm for generating the regional proposal which is slow and time consuming.

3.3 Faster Region with Convolutional neuralnetwork

In faster R-CNN [5] the author designed the model by introducing a new network called regional proposal network (RPN)formoreaccurateandefficientregionproposalgenerationasshowninfig.4**Fig.4**.FasterR-CNNcontainstwocomponents: deep fully convolutional network for proposing a region and fast R-CNN detector, these uses the already proposed regions. The entire system is a unified network for detecting the objects using deep neural network (DNN) with the attention mechanism. The RPNisinstructedtogeneratethegoodqualityofregionalproposalandtheseareusedfordetectionbyfastR-CNN.TheRPNtakes the input image of any size and gives the output as set of object proposals with objectness score. The goal of RPN is to share computation for object detection network with fast R-CNN. For extracting the features, they used the VGG16 model. They train their network on PASCAL VOC 2007 and 2012 dataset to increase the performance of vehicle detection and got the mean average precision (mAP)69.9%.

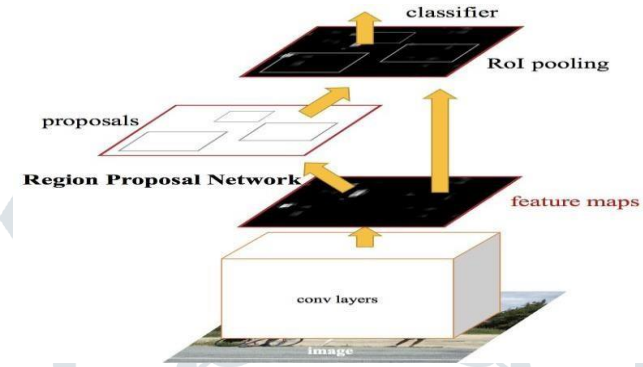

Figure 4: Faster R-CNN architecture

Faster RCNN is not using selective search to generate a region of proposal instead it is using a separate network called RPNtopredicttheregionofproposalshenceitismuchfasterthanpreviousoneandmainadvantageisitcanbeusedforrealtime. The drawback is, it uses low resolution images to improve the detection accuracy and the network will not look for the complete image.

3.4 You Only Look Once(YOLO)

YOLO [6] is one stage approach where it generates the samples and do the classification and regression simultaneously. Its architecture is same as fully convolutional neural network (F-CNN) and uses the concept of RPN to detect the region of proposals. Here RPN perform the multiple time detection for various regions in an input image as shown in the **Fig.5**. It uses the features from complete image to forecast the bounding box it predicts the boundary box for all classes of an input image simultaneously.YOLOenablesendtoendtraining.Thisarchitecturesplitstheimageintos*sgridandeachgridwillgenerateclass probabilities and two bounding boxes. YOLO gives over all mAP 63.2% on VOC 2007 dataset. After extending the model YOLO to YOLOV2 [5] got mAP 78.6% on same dataset.

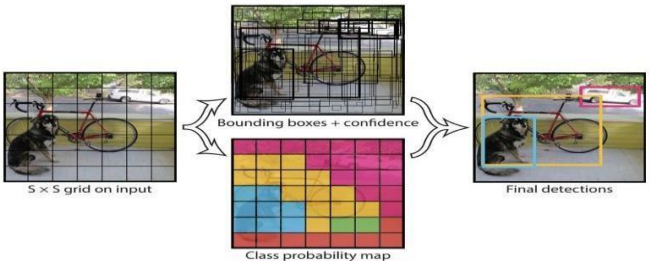

Figure 5: YOLO framework

The biggest advantage of this system is speed. It takes 45 frames per second(fps) which is really need for real time detection.FormoreaccurateandfastdetectionofobjectextendtheYOLOmodeltoYOLO9000[7].YOLO9000detectsmorethan 9000 object categories. The limitations with the YOLO system with all advantages is it struggle to detects the small object due to the spatial constraints and also to generalize the object which are in unusualratio.

3.5 Single shot multibox Detector(SSD)

SSD[8]achievesabetterbalancebetweenquicknessandprecision.ItisthestrongcompetitorwithYOLOtodemonstrate thehigherFPSinrealtime.ThearchitectureofSSDcomposedtwoparts:extractfeaturesmapsandapplyconvolutionfilterfor

object detection. SSD uses vgg16[9] for feature extractor and detects the objects using convolutional 4-3 layers. It encloses all the calculation in a unified network this made training easy. At the divination time, for each object category the network generates a scoreandmakesmallalterationstothedefaultboxforbettermatch.Itattainstherealtimeprocessingspeedandbeatsthefaster R- CNN [3] accuracy by eliminating the RPN. To recuperate the accuracy, it made some improvements in default boxes and multi scale features. It handles the various sizes of objects as shown in the**Fig.6**.

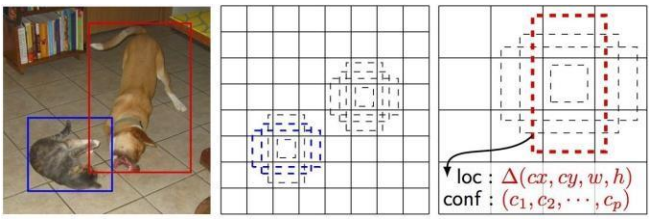

Figure 6: Detection using SSD

The main problem is it is worse than Faster R-CNN [3] for detecting the small-scale objects. SSD has more classification error because of using the same boundary box for making multiple class predictions. SSD512[6] gives better accuracy SSD300. It achieved above 70% mAP for Pascal VOC 2007 dataset.

3.6 Reverse Connection with Objectness PriorNetworks

RON [9] is one of the effective and efficient end to end framework for detecting generic objects. The motivation of this RON is to be associated with the best region free (SSD) and Region based (Faster R-CNN) methodologies. It mainly concentrates on following two problems: multi scale object localization and negative sample mining. To overcome the object localization problem, they introduced the reverse connection which allows the network to recognize the object on multistage in CNN as shown in the **Fig.7**. To deal with the negative sample mining they propose object prior to lower the searching space. Thus, it can directly foretell the final detection results.

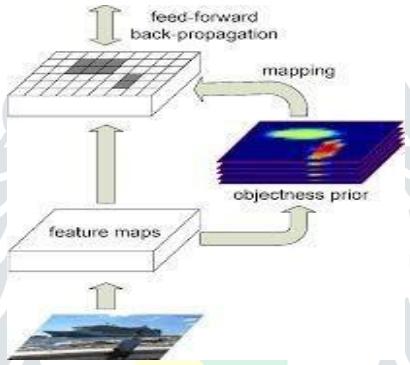

Figure 7: RON framework

RON320[9] achieves 74.2% and RON384[9] achieves 81.3% mAP on the PASCAL VOC 2007 test set. RON is time and resource efficient and reverse connection aid the previous layer of CNN with additional semantic information. They used training strategies like mining and augmentation to increase accuracy and high detection.

IV. Comparative Study of Vehicle DetectionModels

ComparisonofeachofthesevendiverseprominentmodelsresultareoutlinedinthispaperareappearedinTable4.1..R-CNN is the pioneer in the object detection based on deep learning it encountered the problem of selecting a vast region. But it is not possible to implement real time as it takes only 6FPS. Fast R-CNN resolves some drawbacks of R-CNN by feeding the entire image to CNN. The drawback is quite same as R-CNN it also uses same algorithm, and this also cannot be used for real time. Faster R-CNN is the first model that we can implement real time, the mAP and FPS are also comparatively more than previous model.Allthesearetwo-stageapproach,thereisbottleneckintheregionproposals.YOLOisregionfreeproposal,extremelyfaster model with 45FPS and can apply for real time but having strong spatial restrictions and hard to detect small objects. Compared to allothermethodsSSD hasbetteraccuracyandfasterthanpreviousmodelsbutitperformsworse for smallscaleobjects.However, the single stage detectors are quite impressive. RON model uses both best region free and region-based methodologies and improved the accuracy of 6.5% over previous models and achieved better accuracy over all the models, but it also takes less FPS. We referred the results of VOC 2007 dataset for each vehicle and made our analysis that every model has their own limitations. Fromthe**Fig.8**, wecananalyzethatFasterR-CNNaccuracyisquitegood,butspeedisonly18FPSwhereasYOLOachievedgreat speed i.e. 45FPS but accuracy is comparatively less i.e. 63.2%. In case of SSD300 the speed is high i.e. it takes 59FPS, but accuracy achieved is less than SSD500. In contrast SSD500 achieves 75% accuracy but speed is low as it takes 22FPS. RON achieves the remarkable accuracy of 81% but takes only 15FPS for detecting objects. However, the single stage detectors arequite impressive, but the problem is with accuracy. Therefore, for better detection results both accuracy and speed must be high. The general synopsis of various models are illustrated in the Table4.2

Table 4.1: Results on PASCAL VOC 2007 DATASET

Models	mAP	Bike	Bus	Car	Motorbike	FPS
R-CNN	53.5	64.1	56.4	57.2	69	6
Fast R-CNN	70.0	78.1	81.6	78.6	76.6	12
Faster R-CNN	73.2	79.0	83.1	84.7	73.5	18
YOLO	63.2	63.2	68.3	55.9	80.6	45
SSD 300	72.1	79.8	79.8	80.8	80.9	59
SSD 512	75.1	79.6	79.5	85.6	71.3	22
RON 384	81.3	82.4	86.3	84.3	82.4	15

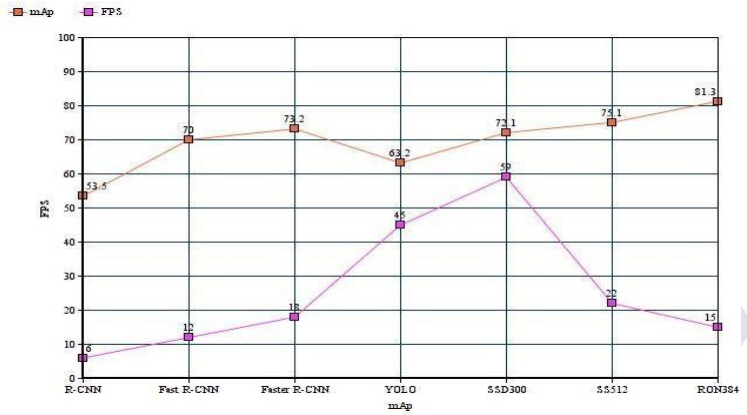

Figure 8: mAP and FPS comparison

Table 4.2: Model Summary

Models	Real Time	Method for RoI	Approach	Vehicle Detection	Region Approach	Disadvantages
R-CNN	No	Selective Search	Two-stage	Slow	Region-Based	Expensive in terms of space and time
Fast R-CNN	No	Selective Search	Two-stage	Slow	Region-Based	Performance is low due to the bottle neck of regional proposal
Faster R-CNN	Yes	RPN	Two-stage	Fast	Region-Based	Uses low resolution images to improve the detection accuracy
YOLO	Yes	RPN	One-stage	Faster	Region-Free	Low recall, difficult to detect small objects
SSD	Yes	Small Convolutional Filter and default boxes	One-stage	Faster	Region-Free	Worse than faster R-CNN for small objects, classification error
RON	Yes	Mining and Augmentation	One-stage	Faster	Both Region-Based and Region-Free	Takes less FPS leads to slow detection

V. CONCLUSION

Deep learning has showed up as a most noticeable in the field of vehicle detection and grouping. Vehicle detection has advancedthroughoutthedecade'smostprominently.Inthisexamination,weintroducedasurveyontheproposeddetectionmodel dependent on their exactness and speed. Each reviewed model in this paper has both preferred standpoint and inconvenience. For continuous, the speed must be high to identify the items with in a second and in the meantime not to overlook the exactness. Thus, decisionofrightlocationshowisbasicandreliesupontheissuethatyouareattemptingtodetermineandsetup.Accomplishingthe extraordinaryprecisionandFPSforincreasinglyexactdetectionofthesemodelisaworkforfutureresearch.

REFERENCES

[1] LI Suhaoa, LIN Jinzhaoa LI Guoquana, BAI Tonga, WANG Huiqiana, PANG Yua "Vehicle type detection based on deep learning in traffic scene" 8th ICICT(2018)

[2] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition" pp.1–14(2014)

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik Rich feature hierarchies for accurate object detection and semantic segmentation(2014)

[4] R. Girshick, "Fast R-CNN," in Proceedings of the IEEE International Conference on Computer Vision, vol. 2015. 1440–1448 Inter, pp (2015)

[5] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: NIPS.(2015)

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection,"(2015)

[7] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," (2016)

[8] W. Liu, D. Anguelov, D. Erhan, and C. Szegedy, "SSD: Single Shot MultiBox Detector," no. 1(2016)

[9] T. Kong, F. Sun, A. Yao, H. Liu, M. Lu, and Y. Chen, "RON: Reverse Connection with Objectness Prior Networks for Object Detection," p. 2017,(2017)