IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.117116

Vehicle Detection using Mask Regional Convolution Neural Network (MRCNN)

Kushagra Gupta[1], Rajot Saha[2], Rekha B S[3]

Department of Information Science and Engineering, R V College of Engineering,

Mysore Road, Bengaluru – 560059[1,2,3]

Abstract: The recent advancement in artificial intelligence approach or deep learning techniques explored the ways to facilitate automation in various sectors. The application of deep learning with the computer vision field has resulted in the realization of intelligent systems. World Health Organization (WHO) estimates the road traffic death in India to be 22.6 per 1,00,000 population. Several factors like lousy drivers, appalling road conditions, ignorance of the traffic rules, inability to understand the present road situation and making correct decisions instantly, may contribute to the road crashes and the eventual deaths. Thus an Intelligent Vehicle System has become very important in today's world which will provide an aid to the driver for dealing with road classification, identifying complex road situations and alert him or her beforehand about the probable
crash.
The technique to be used for vehicle detection is Mask -RCNN. The Mask R-CNN model predicts the class label, bounding box, and mask for the objects in an image. It will be accomplished using an online GPU and cloud services provided by Google Colab by using Tensorflow and Keras framework.
The project should successfully detect each car in the image and mark them with independent masks and a bounding box of just the accurate size to fit in the segmented object. We will be able to observe that almost all the vehicles are recognized by the trained model. The model performs satisfactorily for occluded and small sized objects as well.

Keywords: Computer Vision, Image Classification, Scene Detection, Machine Learning, Convolution Neural Network, Places Dataset.

I. INTRODUCTION

I. In the advancement of technology in the field of artificial intelligence and computer vision have provided a lot of opportunities in research. In the last few decades we have seen a lot of changes in the field of technologies from a smart phone to commercial space travel. Intelligent Vehicle System and automatic cars are one among them. Road accidents are one which poses significant challenge to it. According to World Health Organization (WHO) estimates the road traffic death in India to be 22.6 per 1,00,000 population. There are various reason like lousy drivers, appalling road conditions, ignorance of the traffic rules, inability to understand the present road situation and making correct decision instantly, may contribute to the vehicle crashes and eventual deaths. Therefore an Intelligent Vehicle System has become very important in the today's world which will provide an aid to the driver for dealing with traffic analysis, identifying complex road traffic and alert him or her beforehand about the probable vehicle.

A. Vehicle Detection

An efficient vehicle detection and segmentation system along with a system to perceive road scenes involving lanes [3], pedestrians, traffic divider, potholes [4] and speed bumps [5] forms the basis of the realizing an intelligent vehicle system.

B. Convolution Neural Network (CNN)

Convolution neural network is a type of deep neural network architectures [4]. It is also known as multilayer neural network that inspired by human perception. CNN has variety of application such as computer vision, natural language processing etc.

C. Regional Convolution Neural Network (RCNN)

A naive approach to solve vehicle detection problem would be to take different regions of interest from the image, and use a CNN to classify the presence of the object within that region. The problem with this approach is that the objects of interest might have different spatial locations within the image and different aspect ratios. Hence, you would have to select a huge number of regions and this could computationally blow up. To bypass the problem of selecting a huge number of regions, R-CNN was proposed where we use selective search to extract just 2000 regions from the image and it is called region proposals. Therefore, now, instead of trying to classify a huge number of regions, you can just work with 2000 regions. These 2000 region proposals are generated using the selective search algorithm.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.117116

The Problems With R-CNN are that It still takes a huge amount of time to train the network as you would have to classify 2000 region proposals per image and it cannot be implemented real time as it takes around 47 seconds for each test image.

D. Faster Regional Convolution Neural Network (FRCNN)
Selective search is a slow and time-consuming process affecting the performance of the network. Therefore Faster RCNN was proposed that eliminates the selective search algorithm and lets the network learn the region proposals. The image is provided as an input to a convolutional network which provides a convolutional feature map. Instead of using selective search algorithm on the feature map to identify the region proposals, a separate network is used to predict the region proposals.
The Problems with Faster R-CNN is that the algorithm requires many passes through a single image to extract all the objects and as there are different systems working one after the other, the performance of the systems further ahead depends on how the previous systems performed.

E. Mask Regional Convolution Neural Network (MRCNN)
It is the extension of Faster R-CNN. It goes pixel by pixel matching instead of bounding boxes. It produces an exact mask around the vehicle in an image. Advantage of Mask R-CNN is that Mask R-CNN is simple to train.
Performance: Mask R-CNN outperforms all existing, single-model entries on every task.
Efficiency: The method is very efficient and adds a overhead to Faster R-CNN.
Flexibility: Mask R-CNN is easy to generalize to other tasks.

The general and conceptual architecture of Mask Regional convolution neural network shows in figure 1 and 2.

Fig.1 General layer architecure of CNN

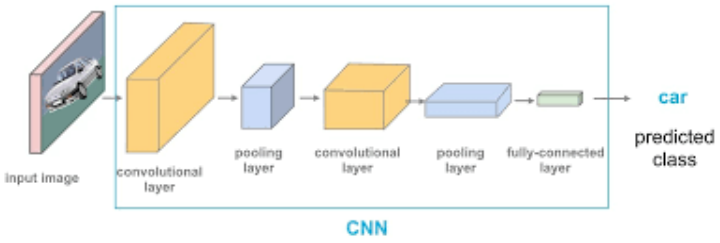
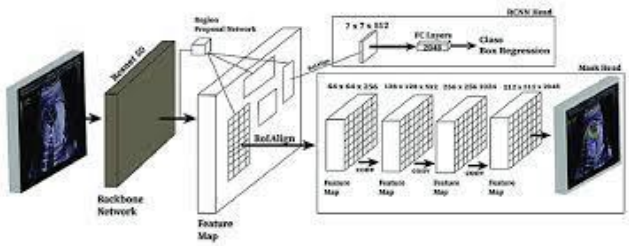


Fig.2 Conceptual Design of MRCNN



II. RELATED WORK

In every domain, research is continuous process. Though, some work on vehicle detection also suggested by researchers. This section explored some related work to scene detections.
An attempt happened to solve one of the most common problems of vehicle detection through Instance Segmentation using Mask R-CNN for Intelligent Vehicle System [3]. Vehicle detection plays a key role in Intelligent Vehicle System and Intelligent Transport S ystem as it assists critical components of these systems like detecting obstacle vehicles to find an unhindered pathway, and even preventing accidents.
A new HOG Feature Extraction and KNN Classification for Detecting Vehicle in The Highway [2]. It uses the feature extractor Histogram of oriented gradient(HOG). In HOG feature descriptor the input image is of size 64 x 128 x 3 and the output feature vector is of length 3780. In the HOG feature descriptor, the distribution (histograms) of directions of gradients (oriented gradients) are used as features.
A model Context Aware Vehicle Detection using Correlation Filter [1] was proposed using Canny and then Hough transform. In this paper authors have modeled correlation filter to detect the vehicles by first using Canny and then Hough

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.117116

transform to identify lane and extract patches then finally carry out correlation analysis to achieve 96.4% accuracy. Canny Edge Detection is a popular edge detection algorithm. It was developed by John F. Canny in the Hough Line Transform is a transform used to detect straight lines.

A paper on the MRCNN architecture was released in 2018[4] in which the author compared R-CNN, Fast R-CNN and Faster R-CNN. The introduction of AlexNet in this paper in ImageNet competition opened doors for use of Convolution Neural Network (CNN) for object detection. After that many state of the art techniques like R-CNN, Fast R-CNN, Faster R-CNN achieves Mean Average Presision (mAP) of 62%, 66%, 75.9% respectively on PASCAL VOC 2012 dataset.

An analysis of thermal images for vehicle detection has been done using improved YOLO[9] V3-tiny to attain 78.77% mAP in This paper uses YOLO(You Only Look Once) algorithm for object detection.

III. PROPOSED METHODOLOGY

Instance segmentation not only classify and localize the object of interest in an image but it also assigns label to each pixel that belongs to the particular object instance. It refines the detection process by assigning the mask along with the bounding box on the object. Thus by employing instance segmentation we can achieve better vehicle detection results as it will allow us to deal with crowded scenes and identify partially occluded vehicles. Convolution Neural Networks (CNN) based models and its derivative model like RCNN, have been significantly used in computer vision tasks like object detection, classification, tracking, and segmentation. Mask RCNN [4] is a state of art technique which provides an extension to the Faster RCNN [7] and realizes instance segmentation using region proposals and identifies each instance of the object of interest in an image at pixel level. We have employed Matterport's implementation of Mask RCNN for vehicle detection using Keras and Tensorflow framework.

A. System Architecture

For the Vehicle detection system, architecture shown in figure 3 is proposed that consist of training and test module. Training module deals with training of model or detector based on image dataset. After the training, Test module invoke that predict scene category for given input image. Test module comprises of input image, feature extraction, Trained classifiers etc.

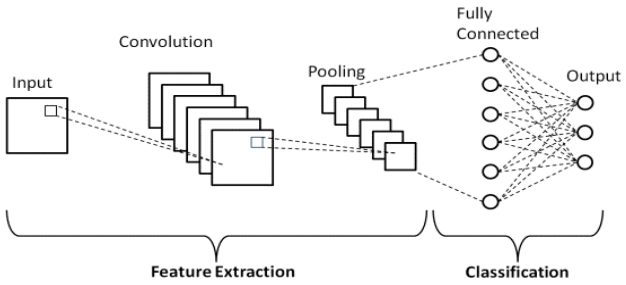


Fig.3 System architecure

B. System High Level Design

There are two main parts to a CNN architecture A convolution tool that separates and identifies the various features of the image for analysis in a process called as Feature Extraction A fully connected layer that utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages.

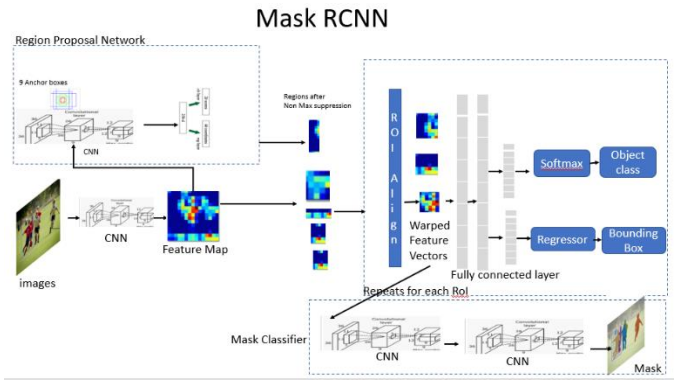


Fig.4 High level system design

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.117116

C. System Low Level Design

Figure5 shows low level system design which presents detail working of each phase of detection system.

Convolutional Layer

This layer is the first layer that is used to extract the various features from the input images. In this layer, the mathematical operation of convolution is performed between the input image and a filter of a particular size MxM. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image with respect to the size of the filter (MxM).

Pooling Layer

In most cases, a Convolutional Layer is followed by a Pooling Layer. The primary aim of this layer is to decrease the size of the convolved feature map to reduce the computational costs. This is performed by decreasing the connections between layers and independently operates on each feature map. Depending upon the method used, there are several types of Pooling operations.

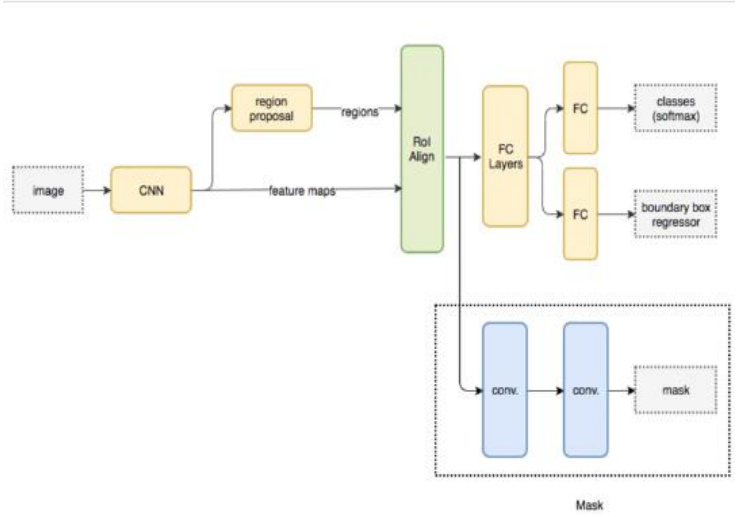


Fig.5 Low level system design

1V. TESTING & RESULT ANALYSIS

Various tests were applied where the system was evaluated. The tests will be of various degrees of priorities ranging from high to low. It is very important that the system passes all the suites of higher priority, failure in doing so will result in revamp of the developmental approach. Passing of test cases of low priority may not be evaluated very strictly. Black-box testing is a method of software testing that examines the functionality of an application without peering into its internal structures or workings. This method of test can be applied virtually to every level of software testing: unit, integration, system and acceptance White-box testing is a method of software testing that tests internal structures or workings of an application, as opposed to its functionality. In white-box testing an internal perspective of the system, as well as programming skills, are used to design test cases.

A. Test Cases

Fig. 6 Case 1: Image only having single car and is fully visible

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.117116

Fig 7. Case 2: Image with all kinds of vehicle(fully or partially visible)

B. Result Analysis

We have implemented vehicle detection using MRCNN and Resnet. We have compared their Mean Average precision and average recall. For MRCNN we have got **0.667 and 0.881** respectively. And for Resnet we have got **0.526 and 0.699**.

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.531
Average Precision  (AP) @[ IoU=0.50       | area=   all | maxDets=100 ] = 0.721
Average Precision  (AP) @[ IoU=0.75       | area=   all | maxDets=100 ] = 0.576
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.369
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.573
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.667
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.392
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.646
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.689
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.534
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.731
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.818
```

Fig. 8 mAP of MRCNN

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.335
Average Precision  (AP) @[ IoU=0.50       | area=   all | maxDets=100 ] = 0.515
Average Precision  (AP) @[ IoU=0.75       | area=   all | maxDets=100 ] = 0.358
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.125
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.386
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.526
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.288
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.451
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.475
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.200
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.557
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.699
```

Fig. 9 mAP of Resnet

V. CONCLUSION

After implementing Vehicle detection using MRCNN and Resnet we have come to a conclusion that MRCNN is much more better than Resnet because its accuracy is much higher and it makes complete box around the vehicle while resnet makes only bounding box. The final trained model successfully visualizes the accurate localization and masks for cars on the images from different scenes and situations.

In future this can be used at highways to detect what car is coming and how many cars are passing by a highway or a road. Permanent cameras across city can help to tell traffic on different areas.

REFERENCES

[1] C. Son, S. Park, J. Lee, and J. Paik, “Context Aware Vehicle Detection using Correlation Filter,” 2019 IEEE Int. Conf. Consum. Electron. ICCE 2019, pp. 5–6, 2019, doi: 10.1109/ICCE.2019.8661942.

[2] F. A. I. Achyunda Putra, F. Utaminingrum, and W. F. Mahmudy, “HOG Feature Extraction and KNN Classification for Detecting Vehicle in The Highway,” IJCCS (Indonesian J. Comput. Cybern. Syst., vol. 14, no. 3, p. 231, 2020, doi: 10.22146/ijccs.54050.

[3] Apoorva Ojha, Satya Prakash Sahu, Deepak Kumar Dewangan, “Vehicle Detection through Instance Segmentation using Mask R-CNN for Intelligent Vehicle System” IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-0-7381-1327-2

[4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 2, pp. 386–397, 2020, doi: 10.1109/TPAMI.2018.2844175.

DOI: 10.17148/IJARCCE.2022.117116

[5] R. Girshick, J. Donahue, T. Darrell, J. Malik, U. C. Berkeley, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 1, p. 5000, 2014, doi: 10.1109/CVPR.2014.81.

[6] R. Girshick, "Fast R-CNN," Proc. IEEE Int. Conf. Comput. Vis., vol. 2015 Inter, pp. 1440–1448, 2015, doi: 10.1109/ICCV.2015.169.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards RealTime Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[8] Y. Li, K. He, J. Sun, and others, "R-fcn: Object detection via regionbased fully convolutional networks," Adv. Neural Inf. Process. Syst., no. Nips, pp. 379–387, 2016, [Online]. Available: http://papers.nips.cc/paper/6465-r-fcn-object-detection-via-regionbased-fully-convolutional-networks.pdf.

[9] J. Gong, J. Zhao, F. Li, and H. Zhang, "Vehicle detection in thermal images with an improved yolov3-tiny," Proc. 2020 IEEE Int. Conf. Power, Intell. Comput. Syst. ICPICS 2020, pp. 253–256, 2020, doi: 10.1109/ICPICS50287.2020.9201995.

[10] T. Wang, Y. Y. Hsieh, F. W. Wong, and Y. F. Chen, "Mask-RCNN Based People Detection Using A Top-View Fisheye Camera," Proc. - 2019 Int. Conf. Technol. Appl. Artif. Intell. TAAI 2019, pp. 2019–2022, 2019, doi: 10.1109/TAAI48200.2019.8959887.

[11] N. Bhattacharya, D. K. Dewangan, and K. K. Dewangan, "An efficacious matching of finger knuckle print images using gabor feature," Adv. Intell. Syst. Comput., vol. 653, pp. 153–162, 2018, doi: 10.1007/978-981-10-6602-3_15.

[12] P. Pandey, K. K. Dewangan, and D. K. Dewangan, "Enhancing the quality of satellite images by preprocessing and contrast enhancement," Proc. 2017 IEEE Int. Conf. Commun. Signal Process. ICCSP 2017, vol. 2018-Janua, pp. 56–60, 2018, doi: 10.1109/ICCSP.2017.8286525.

[13] P. Pandey, K. K. Dewangan, and D. K. Dewangan, "Satellite Image Enhancement Techniques- A Comparative Study," Int. Conf. Energy, Commun. Data Anal. Soft Comput., pp. 597–602, 2017, doi: 10.1109/ICECDS.2017.8389506.

[14] P. Pandey, K. K. Dewangan, and D. K. Dewangan, "Enhancing the Quality of Satellite Images using Fuzzy Inference System," Int. Conf. Energy, Commun. Data Anal. Soft Comput., pp. 3087–3092, 2017, doi: 10.1109/ICECDS.2017.8390024.

[15] D. K. Dewangan and Y. Rathore, "Image Quality estimation of Images using Full Reference and No Reference Method," Int. J. Adv. Res. Comput. Sci. ., vol. 2, no. 5, 2011. [16] D. Dewangan and Y. K. Rathore, "Image Quality Costing of Compressed Image Using Full Reference Method," no. February, 2016.

[17] F. A. I. Achyunda Putra, F. Utaminingrum, and W. F. Mahmudy, "HOG Feature Extraction and KNN Classification for Detecting Vehicle in The Highway," IJCCS (Indonesian J. Comput. Cybern. Syst., vol. 14, no. 3, p. 231, 2020, doi: 10.22146/ijccs.54050.

[18] Mithi, "Vehicle Detection with HOG and Linear SVM," Medium, vol. 1, no. June, pp. 6–9, 2021, [Online]. Available: https://medium.com/@mithi/vehicles-tracking-with-hog-and-linear-svmc9f27eaf521a.