International Journal of Research in Engineering, Science and Management
Volume-3, Issue-6, June-2020
www.ijresm.com | ISSN (Online): 2581-5792

Selective Area on Modified CNN for Vehicle Detection

Utsav Jain[1*], Shivank Gahlot[2], Sanyam Jain[3], Ajeet Kumar Bhartee[4]

[1,2,3]*Student, Department of Computer Science and Engineering, Galgotias College of Engineering and Technology, Greater Noida, India*

[4]*Professor, Department of Computer Science and Engineering, Galgotias College of Engineering and Technology, Greater Noida, India*

*Corresponding author: utsavjain2698@gmail.com

***Abstract*: Recent work on fast-RCNN and faster R-CNN is done to increase the speed and performance of object detection. This is done by selecting various methods of using the CNN. These systems try to apply the CNN on proposed region multiple times or applying the CNN once and then selecting RoI to detect the object. These systems have high accuracy in general condition. However, when the camera is supposed to be stationary there are several steps which can be considered as non-essential to detect vehicles or objects of various classes. The proposed work is used to reduce the area of application of a system of detection of vehicles for stationary camera. While simultaneously reducing the time to detect vehicles in a frame.**

***Keywords*: R-CNN, Faster R-CNN, Convolutional Neural Network, Feature map.**

1. Introduction

Vehicles on roads need to be monitored in metropolitan or urban areas to regulate the flow of traffic and maintain vigilance. Monitoring traffic and identifying vehicles also helps enforcing authorities to act rapidly. There are several available models for vehicle detection in city to monitor traffic congestions [1] and identify the type of vehicles [2] on road in specific locations. These models are similar to the proposed models in a way, that all these models require the camera to be fixed at a point.

However, the aim of our model is different from [1][2], as our model aims at reducing the time of detection of objects by application of a 2-stage system. The work of Clifford Austin, Carlo Migel Bautista, for detecting vehicles in low resolution videos of traffic is the base of this research.

The common deep learning approach to detect vehicles is, to apply a Convolutional neural network, either one or many layers of a CNN are either fully connected, and every layer of the CNN gives the output as a feature map, which is a 2D matrix. The resultant feature map yj of a convolution layer j is shown by $yj = bj + i(kij * xi)$ where bj is a bias parameter which can be trained, kij is the layer's filter, and xi is the image used as an input.

Therefore, convolutional operation represented as '*' for any node is given as,

$$G[m,n] = (f * h)[m,n] = \sum_j \sum_k h[j,k] f[m-j, n-k] \quad (1)$$

Here the kernel is denoted by "h" and the image denoted by "f". "m", "n" are the row and column indexes.

2. Faster RCNN for Vehicle Detection

In Faster R-CNN, (RPN) shares the convolutional layers with object detection network, therefore reducing the proposal costs. In this model few extra convolutional layers are used to regress regional bounds [4] This algorithm has led to a significant increase in speed and it has also proven to improve performance of object detection. Faster R-CNN was the basic model for winning entries in ImagNet [6] detection and localization at ILSVRC 2015 at the COCO2015 competition [3].

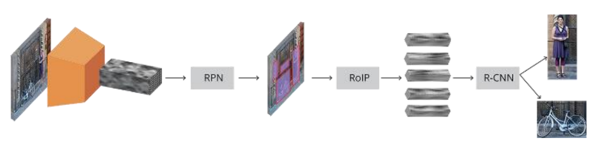


Fig. 1. The structure of Faster R-CNN. RPN and the classifier shared fully convolutional layers, which are trained jointly. The RPN acts as attention director, for the optimal-bounding box for a wide range of scales

3. Overview of Model

The faster RCNN models is a 2 stage models similar to our model, however the first stage of faster R-CNN instead of using selective search algorithm on the feature map to identify the region proposals like conventional RCNN or fast RCNN [5], faster RCNN uses a completely separate network to predict regional proposals. then reshaping these proposals using a RoI pooling layer and therefore classifying the image in the proposed region to predict the bounding boxes respective offset values.

From the above test it is clear that the two stage faster R-CNN is much faster than traditional regional CNN.

Our model however unlike faster R-CNN does not use does not use the same separate network for regional proposal. The model proposed in this paper like RCNN uses a regional proposal algorithm first, for vehicle detection, only the lower half of the image frame is selected to run a separate network to identify the feature map to identify regions, where the probability density of vehicle (i.e. the road in the image in lower half) is maximum. The selection of lower half of the image means the network to find regional proposals will have look only into half of the grid on the image.

Deselection of upper half of the image to find regional proposal on a SxS grid results in a (S/2 x S) grid thereby reducing the input layer and consequently the inner layer of the network to be reduced by half.

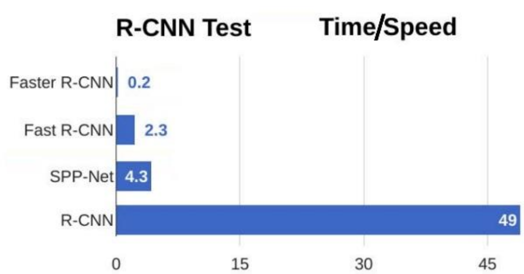


Fig. 2. Time vs. Speed test comparison of faster R-CNN, Fast R-CNN, SPP-net, traditional R-CNN

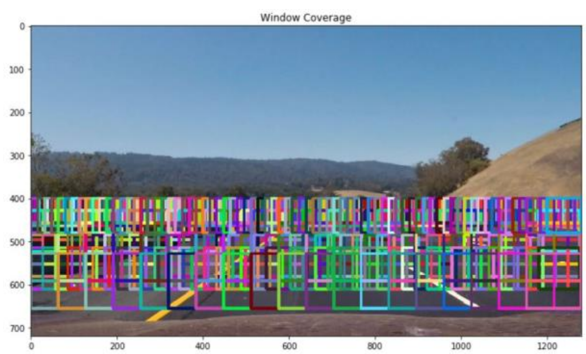


Fig. 3. The regional proposal network running on only lower half of the image, avoiding to run the region with least probability of object density

For the image with the (S/2 x S) grid with bounding boxes, every single bounding box gives the outputs an offset values as well as class probability for the boxes. The bounding boxes having these class probabilities above a given threshold value are selected and therefore selected to localise the object in the frame.

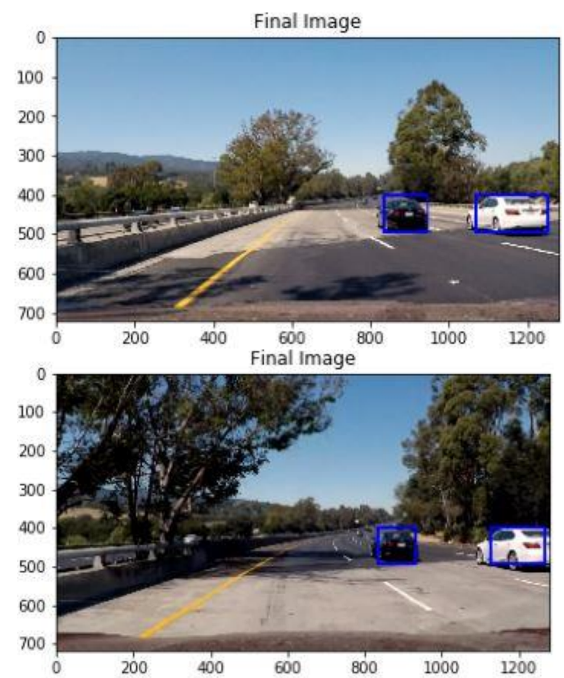
Fig. 4. Grid with green as high-class probability of vehicle i.e. above threshold

4. Comparison with Other Models of R-CNN

In RCNN after the RoI is generated, 2000 regions are generated as output and convolutional network is applied to every region thus the time for detection in every frame is about 45-50 second. However, in our model which is derived from faster RCNN and YOLO model a separate network initially find the RoI similar to that of RCNN but the region of proposals can be calculated to be 2000/2, also our model uses the grid for bounding box of region and does not uses "selective search methods" which performs no learning, rather uses a network to learn about the high probability region in the frame the regions selected for application of convolutional network is highly selective and adaptive.

The network looks at the parts of the half image which must be having high probabilities of holding the object, rather than looking at half image at once.

5. Validation



6. Conclusion

We organized a range of comparative-study experiments and gave a comprehensive analysis on the performance of the proposed model of vehicle-detection, which includes a study of both testing and training size scale, localization of object as compared to recognition and training. We derived an approach better than the RCNN and combined the model with YOLO model of object detection to identify RoI[7], and restructured the faster RCNN to identify region for object detection first and then apply the convolutional layer so that the model can return 0 (zero) value or no output when a vehicle is not in the frame without even the application of a convNet and save time for such iterations.

References

[1] N. Aguirre, J. Alcantara, and J. Trinidad, "TraVis: Web-Based Vehicle Counter with Traffic Congestion Estimation Using Computer Vision", Undergraduate thesis, 2014.
[2] Z. Chen, N. Pears, M. Freeman, and J. Austin, "Road vehicle classification using Support Vector Machines," in Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference, Shanghai, 2009, pp. 214-218.
[3] S. Ren, K. He, R. B. Girshick, and J. Sun. Faster R-CNN: towards real-time object detection with region proposal networks. CoRR, 2015
[4] J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," International Journal of Computer Vision, 2013.
[5] R. Girshick, "Fast R-CNN," in IEEE International Conference on Computer Vision (ICCV), 2015.
[6] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in Neural Information Processing Systems (NIPS), 2012.
[7] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, University of Washington, Allen Institute for AI, Facebook AI Research," May 2016
[8] http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf
[9] https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e
[10] https://towardsdatascience.com/gentle-dive-into-math-behind-convolutional-neural-networks-79a07dd44cf9