International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Vehicle Detection from Images using Deep Fully Convolutional Networks

Vengoti Bhargavi, D.Rajeswara Rao

Abstract*: Vehicle detection provides the facilitation of traffic planning and management. It also helps in finding suspected and stolen vehicles. Although it has many applications, it is a very complex problem due to variations in vehicle type and size. As the amount of vehicle types are very high the models find it hard to classify the correct type of the vehicle. In this paper, we are proposing a vehicle detection model based on YoloV3 convolutional neural network architecture with custom backbone. Our proposed backbone in the Yolov3 architecture helps classify the different types of vehicles accurately. This makes the classification of the images at pixel level and predicts the regression based ROI bounding box for the classified vehicles in the images. The model contains features extracted at different kernel sizes to find the features at multiple scales which will then be concatenated. Experiments were performed on the Kitti vehicle detection dataset have shown the superior performance of our proposed model.

***Index Terms*: Image classification, Regression, Vehicle detection, Yolov3**

I. INTRODUCTION

Vehicle detection has been the main problem for applications such as autonomous driving and traffic control. Having a faster and automatic vehicle detection model can mitigate accidents and ensures safe driving. Major corporations and organizations are spending huge amounts of money on developing robust detection models

Vehicle detection is a special subset of the generic object detection. Object detection models detect the objects of the interest from the images without any human intervention. Recent achievements in the computer vision has boosted the performance and accuracy of the object detection models [1,2,3]. They have surpassed all the previously used approaches to great lengths. A subclass of these models try to predict the rectangular box around the detected object along with the pixel level classification. This boxes are called bounding boxes and they can help visualize the location of the detected object directly on the original image without overlapping with the segmentation masks.

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
Vengoti Bhargavi*, Studying Master of Technology, Department of Computer Science and Engineering, V R Siddhartha Engineering College Vijayawada.
D.Rajeswara Rao, Professor,, Department of Computer Science and Engineering, V R Siddhartha Engineering College, Vijayawada.

Most of the literature for vehicle detection has also used similar kind of approaches. These models have several shortcomings. They cannot effectively predict the vehicles due to variations and occlusions. To account for these convolutional neural networks with large number of layers are used. They are hard to train and takes a considerable amount of time to predict. This can a problem in real world application because in those scenarios the model has to detect vehicles in real time as they pass.

In this paper, we are proposing a new model that is faster and efficient at detecting vehicles. Our proposed model is based on YoloV3 architecture. YoloV3 has been the faster architecture for object detection. We have developed our own custom backbone for extracting features and modified the architecture to make it more efficient. To evaluate our model at detecting vehicles we have implemented the model on the widely used Kitti vehicle detection dataset. Based on the experiments our proposed model has shown better results than other models.

II. RELATED WORK

Bhaskar et al. [4] proposed an image processing based approach for vehicle detection tracking. They used Gaussian mixture model to separate the foreground and background from image frame. Then used blob detection to detect the vehicle by its movement. Tracking is performed by searching for centroids around the centroids detected in earlier frame. Liu et al. [5] focused their work on finding vehicles from airborne lidar data. The problem of vehicle detection was taken as a binary classification problem. Gaussian classification is used to separate the foreground and background of the image and Gradient based segmentation is used to segment the actual vehicles. Xiao-feng gu et al. [6] proposed a real time vehicle detection and tracking neural network model. Their proposed model can obtain vehicle candidates, vehicle probabilities and the coordinates of the vehicles. Convolutional layers, Spatial pyramid pooling and inception modules are used to extract the features while fully connected layers are used to produce the final results.

Wen et al. [7] solution to vehicle detection problem is using haar-like feature selection with adaboost model. The proposed approach uses both the feature values along with the class labels to generate classification locations. Chu et al. [8] developed a deep convolutional neural network with region of interest voting. Offset direction of each ROI boundary is predicted by the proposed CNN model. The Convolutional layers are accompanied by fully connected layers for bounding box regression and offset direction.

Step 1: Collect KITTI Object Dataset
Step 2: Perform Data Augmentation on the dataset and the corresponding labels to generate larger dataset from existing dataset
Step 3: Train Modifies YoloV3 network
Step 4: Initialize all weights and biases of the network to small random value
Step 5: Set learning rate α such that 0<α<1
Step 6: for epoch = 1 to M do
  Propagate batch of images through the network
  For k=1 to number of neurons in the output layer
  Calculate Error
Step 7: for layers L-1 to 1 do
  Calculate error factor to be back propagated
  End for
Step 8: for i=1 to L do
  Update weights and biases
  End for
  End for
Step 9: Save the most significant weights of the network at a particular epoch
Step 10: Predict the class labels and bounding boxes along with the confident score on test data
Step 11: Visualize the bounding boxes on top of the original testing images.
Step 12: Report the different evaluation measures along with the visualization images.

Figure 1. Algorithm of the proposed model

I. METHODOLOGY

The proposed algorithm was given in Figure 1.

A. Kitti Vehicle Detection Dataset

The proposed model is evaluated on a publically available benchmark dataset Kitti[9]. The dataset is collected from video cameras as bird's eye view and save as png files. The dataset contains approximately 7481 training images and 7518 testing images. All the objects are labeled and can be directly used to train. Data augmentation is performed to generate more number of images so that the model can generalize for broad variations of vehicles. Image padding is performed to make all the images in to constant size.

B. Modified Backbone

The original backbone for YoloV3 is called Darknet-53. It contains normal convolution layers with pooling layers along skip connections. This backbone is used to extract the features from the images. We have made our own custom backbone for vehicle detection. As the vehicle detection requires a lot of minute features that have to propagate through the deeper layers. We have used dense connections between the blocks. The dense connection extracts more features with higher feature reuse and it takes less number of parameters than the darknet-53 model. The architecture is given in Figure 2. After the two initial convolutions each of the following convolution block have dense connections from the previous blocks. The features are reduced in size by a factor of 2 at each layers. Each convolution block contains a convolutional layer, maxpooling layer.

C. YoloV3

YoloV3 is a fully convolutional network that has been the faster and efficient architecture for object detection in recent literature. It performs object detection at three different scales. Three different sizes of features are taken at three different places from the backbone network which are up sampled feature sizes. A 1x1 detection kernel is applied on the three feature maps to produce the predicted output. The shape of the detection kernel is given as

$$1 \times 1 \times \left(B \times (5+C)\right) \quad (1)$$

Where B is the number of bounding boxes, 5 are the 4 bounding box attributes and 1 confidence, and C is number of classes.

Detections performed at different layers helps in detecting the very small details which in turn can detect smaller objects from the images. The dense layers preserve the fine grained features from all the previous layers. Bounding boxes are predicted using the logistic regression. Multiple bounding boxes are generated for the given vehicle. The bounding box that overlaps best with the groundtruth at a certain threshold of .5 is taken as the best bounding box. All the other generated boxes less the threshold values are ignored.

Its architecture facilitates for more than one class prediction which is also called a multilabel prediction. Softmax is not used as it is done in previous architectures and each classification is done using logistic classifications. The loss function for class prediction is taken as binary-cross entropy.

II. EXPERIMENTS

The proposed model was implemented in Pytorch library using python programming language. stochastic gradient descent is used to optimize the model and updates the weights at each layer. Binary cross entropy is used as the loss function for classification and mean squared error was used for the logistic regression.

III. RESULTS

The proposed model is evaluated by using the mean average precision(AP). Mean average precision computes

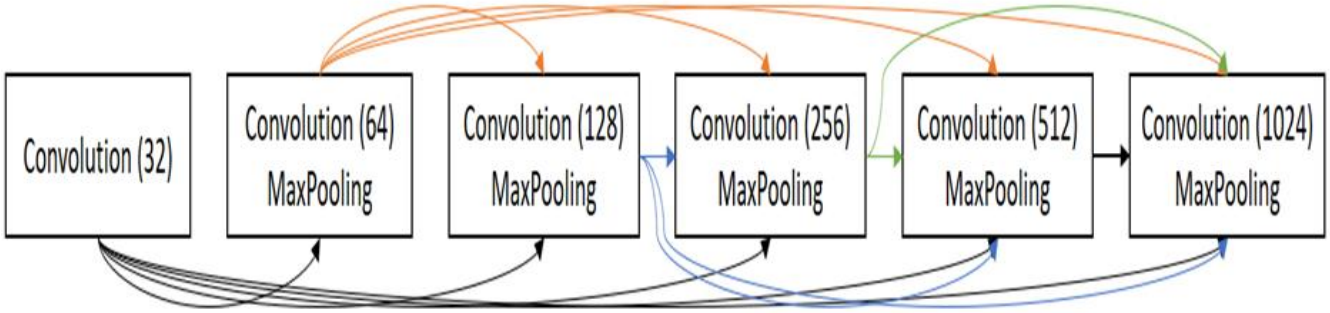

Figure 1. Network Backbone Architecture

the average precision value for recall value. The results for different state of the art models are given in Table 1. As it can be seen that our proposed model has achieved better AP than the rest of the state of the art models. Some of the predictions performed on the model using the KITTI test dataset were given in Figure 3, 4 and 5. Our model has predicted vehicles with higher confidence. It was also able to detect vehicles with occlusions and from side angles. The vehicles that are far in the image were also be able to detected.

Table 1. Evaluations metrics of proposed and comparison models

Model	AP
Faster R-CNN[3]	86.5
MSCNN[10]	89.3
Googlenet	89.8
SubCNN[11]	90.3
SDP+RPN[12]	90.14
Proposed	91.2

I. CONCLUSION

Vehicle detection from images is a very complex problem. The model has to be fast at detecting the vehicles to be able to cope with the real world applications. Most of the models fail to predict images in real time and even if they do, the accuracy of the predictions is heavily decreased. We have proposed an approach for vehicle detection in this paper that is faster and more efficient at detecting vehicles. Our custom backbone extracts feature with maximum feature reuse and less training features. Our research can save a path for more applications of automatic road detection that complements the autonomous vehicle driving. It ensures safety and reliability of those applications.

REFERENCES

1. P. Viola and M. J. Jones, “Robust real-time face detection,” Int. J. Comput. Vis., vol. 57, no. 2, pp. 137–154, 2004.
2. W Y. Li, S. Wang, Q. Tian, and X. Ding, “Learning cascaded shared-boost classifiers for part-based object detection,” IEEE Trans. Image Process., vol. 23, no. 4, pp. 1858–1871, Apr. 2014.
3. S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards realtime object detection with region proposal networks,” in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 91–99.
4. P. K. Bhaskar and S. Yong, "Image processing based vehicle detection and tracking method," 2014 International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2014, pp. 1-5.
5. Y. Liu, S. T. Monteiro and E. Saber, "Vehicle detection from aerial color imagery and airborne LiDAR data," 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, 2016, pp. 1384-1387.
6. X. Gu, Z. Chen, T. Ma, F. Li and L. Yan, "Real-Time vehicle detection and tracking using deep neural networks," 2016 13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, 2016, pp. 167-170.
7. X. Wen, L. Shao, W. Fang and Y. Xue, "Efficient Feature Selection and Classification for Vehicle Detection," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 3, pp. 508-517, March 2015.
8. W. Chu, Y. Liu, C. Shen, D. Cai and X. Hua, "Multi-Task Vehicle Detection With Region-of-Interest Voting," in IEEE Transactions on Image Processing, vol. 27, no. 1, pp. 432-441, Jan. 2018.
9. A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in CVPR, pp. 3354–3361, 2012.
10. Z. Cai, Q. Fan, R. Feris, and N. Vasconcelos. A unified multi-scale deep convolutional neural network for fast object detection. ECCV, 2016
11. Y. Xiang, W. Choi, Y. Lin, and S. Savarese. Subcategoryaware convolutional neural networks for object proposals and detection. arXiv:1604.04693, 2016.
12. F. Yang, W. Choi, and Y. Lin. Exploit all the layers: Fast and accurate cnn object detector with scale dependent pooling and cascaded rejection classifiers. CVPR, 2016.

AUTHORS PROFILE

Vengoti Bhargavi, Studying Master of Technology, Department of Computer Science and Engineering, V R Siddhartha Engineering College Vijayawada.

D.Rajeswara Rao, Professor,, Department of Computer Science and Engineering, V R Siddhartha Engineering College, Vijayawada.

Figure 3. Sample Image 1 and Prediction from KITTI Test Dataset

Figure 4. Sample Image 2 and Prediction from KITTI Test Dataset

Retrieval Number: B3103078219/19©BEIESP
DOI: 10.35940/ijrte.B3103.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Figure 2.Sample Image 3 and Prediction from KITTI Test Dataset