Overlapping Vehicles Detection Using Different Object Detection Methodologies

Minhazul Abedin[1], Rubel Uddin[2], Kamrul Hasan Tuhin[3] and Nur Mohammad Al Kawsar[4]

[1,2,3]Noakhali Science and Technology University
[4]International Islamic University Chittagong
E-mail: mabedin06@gmail.com

1. INTRODUCTION

In any urban area traffic congestions are increasing day by day. It is becoming harder to manage the traffic system in developing countries. In Southeast Asia Dhaka city is facing severe crisis with currently average traffic speed of 6.4 kmph. In upcoming year's it will be difficult to manage the traffic problems. To solve this problem we can use AI-based traffic control systems.

Object detection in deep learning is most popular topic in recent periods. Various types of object detection algorithms were introduced by many researchers and they are able to yield remarkable results. However, most of them fail when it comes to detecting overlapping and small objects in images and that's why it is very challenging task in urban area like Dhaka city to detect multiple overlapping vehicles.

The purpose of this paper is to train a vehicle detector using YOLOv3, YOLOv4, RetinaNet, Faster R-CNN and EfficientDet methods and make experimental analysis comparison between them with the result obtained from these methods. In addition, we tried different settings and approaches to maximize the result.

1.1 Objectives

- Assessing the ability of state-of-the-art methods to detect and recognize traffic vehicles.
- Creating new opportunities to adopt interesting applications such as self-driven cars, drone vision, robot navigation and video surveillance systems.

2. RELATED WORK

Object detection methods can be divided into two categories: (i)Two stage detectors: Faster -RCNN[1] (ii)One stage detectors: YOLOv3[2], YOLOv4[3] , RetinaNet[4] etc. Two stage detectors are very slow compared to one stage detectors. As a result, it is not efficient to use two stage detector methods in real time application. Among all the existed object detection methods, one stage detector model, which are represented by YOLOv4 [3], is advantageous in both detection precision and speed.

3. METHODOLOGY

One-stage detectors process object detection as a simple regression problem by taking an input image and learning the class likelihood and bounding box coordinates at the same time.

An one-stage anchor-based detector is normally made up of a backbone network, a detection neck, a feature pyramid network (FPN), and a detection head for object classification and localization. They are also common components in most of the one-stage detectors.

On the other hand, in two-stage detectors firstly, a Region Proposal Network is used to generate regions of interest that ensure high probability of being an object. Next, they perform the final classification and bounding-box regression of objects by taking these regions as input.

A two stage detector i.e. Faster R-CNN, takes image as an input to a convolutional network which provides a convolutional feature map. Instead of using selective search algorithm on the feature map to recognize the region proposals, a separate network is used to predict the region proposals. The predicted region proposals are then reshaped using a ROI pooling layer which is then used to classify the image within the proposed region and predict coordinates of the bounding boxes.

YOLOv4[3] added some new features that improves detection accuracy than YOLOv3[2].The universal features are: WRC, CSP, CmBN, SAT, Mish activation, Mosaic data augmentation, CmBN, DropBlock regularization, and CIoU loss, and combine some of them to achieve better results.

4. RESULTS

In this work, we used YOLO series(YOLOv3 and YOLOv4), RetinaNet, Faster-RCNN and EfficientDet algorithms to detect different types of vehicles.We will present the comparison results in this section.

4.1 Dataset

DhakaAI 2020 recently organized a traffic detection challenge for the solutions of automatic Dhaka traffic detection problems on optical images. They provide a dataset composed of vehicle images, where an image contains a vehicle of one or more of 21 different classes of vehicle. This makes the dataset useful for multiple vehicle detection and recognition. The considered vehicle classes are: ambulance, auto-rickshaw, bicycle, bus, car, garbage van, human hauler, minibus, minivan, motorbike, Pickup, army vehicle, police car, rickshaw, scooter, Suv, taxi, three-wheelers (CNG), truck, van, wheelbarrow.
The dataset contains 3003 train images with annotation and 500 test images without annotation.

Datset Link: https://doi.org/10.7910/DVN/POREXF

4.2 Evaluation Metric

This DhakaAI competition is evaluated on the mean average precision at a different intersection over union (IoU) thresholds. The threshold values range from 0.5 to 0.75 with a step size of 0.05.
At each threshold value t, a precision value is calculated based on the number of true positives (TP), false negatives (FN), and false positives (FP) resulting from comparing the predicted object to all ground-truth objects:

$$\text{precision}(t) = \frac{TP(t)}{TP(t)+FP(t)+FN(t)} \quad [1]$$

The average precision of a single classification label
is calculated as the mean of the above precision values at each IoU
threshold:

$$\frac{1}{|\text{IoU Thresholds}|} \sum_t precision(t) \quad [2]$$

4.3 Tables and figures

The comparison between the state of the art detectors are shown below:

Table 1. **Comparison between different object detectors for overlapping vehicle dataset**

Methods	mAP@0.5:0.75
YOLOv4	0.1606
EfficientDet	0.1103
YOLOv3	0.0708
RetinaNet	0.0672
Faster-RCNN	0.0552

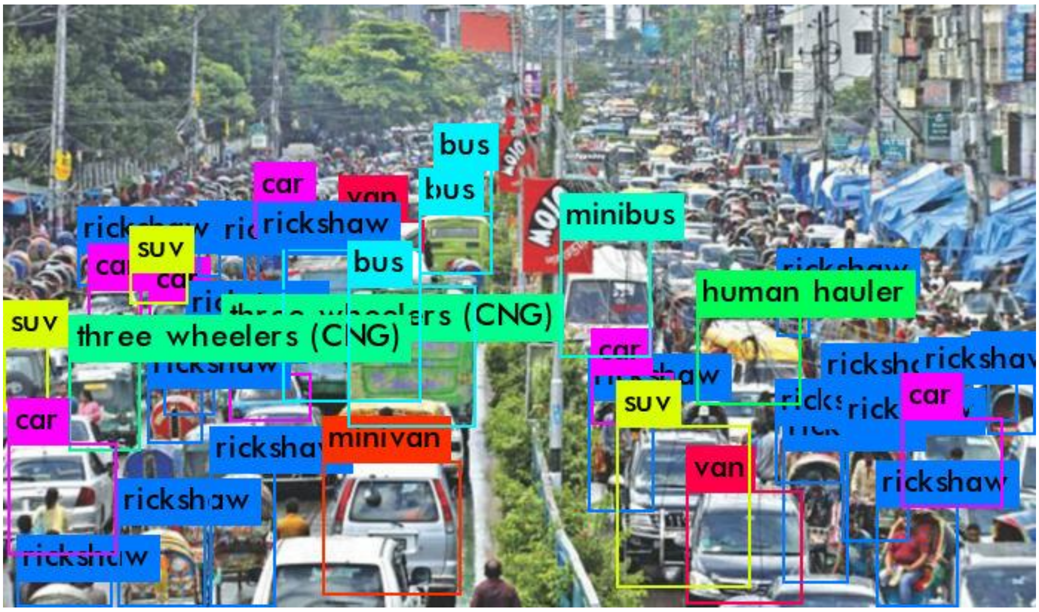

Figure 1. Most accurate and fast detection of YOLOv4

4.4 Experimental analysis

As we see in Table 1, YOLOv4 has the highest mAP for most classes of the dataset. Indeed YOLOv4 is most accurate and fast. Figure 1 shows the YOLOv4 predictions.We can see that the result is outstanding and accurate.We find this best result at IOU threshold 0.5 and confidence score 0.2. The other methods can not detect overlapping vehicles too much accurately.

5. DISCUSSIONS

Object detection and drive state recogination help to reduce traffic congestion,which enable a easy communication in any remote place that may enrich our economic growth.

6. CONCLUSION

The proposed vehicle detector has been successfully trained by using Faster R-CNN, EfficientDet, YOLO and RetinaNet deep learning methods on the sample vehicle datasets and the vehicle detection process has been successfully performed by the trained vehicle detector being tested on the test data set.
The performance of YOLOv4 surpasses all other state of the art detectors in terms of overlapping vehicles detection.
The challenge is ongoing.So we are trying to increasing accuracy by object detector combination[5]

REFERENCES

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017.

[2] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.02767.

[3] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.10934.

[4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," Aug. 2017, [Online]. Available: http://arxiv.org/abs/1708.02002.

[5] K. Drid, M. Allaoui, and M. L. Kherfi, "Object detector combination for increasing accuracy and detecting more overlapping objects," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12119 LNCS, pp. 290–296, 2020.