Accepted Manuscript

HybridNet: A fast vehicle detection system for autonomous driving

Xuerui Dai, Xueye Wei

PII: S0923-5965(18)30377-1
DOI: https://doi.org/10.1016/j.image.2018.09.002
Reference: IMAGE 15439

To appear in: *Signal Processing: Image Communication*

Received date : 18 April 2018
Revised date : 4 September 2018
Accepted date : 4 September 2018

Please cite this article as: X. Dai, X. Wei, HybridNet: A fast vehicle detection system for autonomous driving, *Signal Processing: Image Communication* (2018), https://doi.org/10.1016/j.image.2018.09.002

HybridNet:A Fast Vehicle Detection System for Autonomous Driving

***Abstract*—For autonomous driving system, vehicle detection is an import part as well as a challenging problem due to the large intra-class differences caused by occlusion, truncation and different viewpoints. The detection system should be fast and accurate enough to support real-world applications. Most of the existing deep convolution neural network(CNN) based object detection methods can be roughly categorized into two streams: single-stage and two-stage modes. These single-stage methods are usually extremely fast and easy to train with losing some precision. As for two-stage methods, they often get high performance in object detection competitions, however, they are not competitive for real-world applications because of the speed limits. The detection system with high degree of precision and fast computation speed is desirable. In this paper, a new two-stage regression based cascade object detection system is proposed. This system can be fast detection of the vehicles which concentrated the advantages of the two aforementioned methods, denoted by HybridNet. In our design, the first and the second stage are both regression modes. We add a transitional stage to map proposals(generated in the first stage) on high resolution feature maps to get exact features for decision refinement in the second stage. The challenging KITTI and PASCAL VOC2007 data sets are used to evaluate our proposed method. The experimental results show that our approach is more fast and more accurate in vehicle detection than other state-of-the-art methods.**

***Index Terms*—Vehicle detection, CNN, two-stage, decision refinement.**

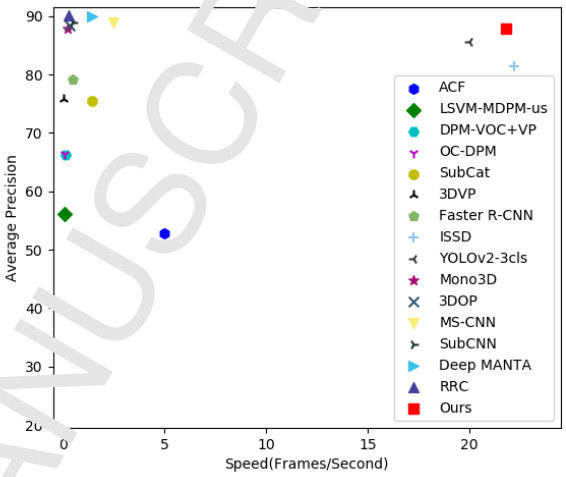

Fig. . Comparisons of HybridNet along with previous traditional, single-stage and two-stage detectors in accuracy and speed on KITTI object detection benchmark.

I. INTRODUCTION

With the increasing size of the city, the number of vehicles increases significantly and the population density also becomes higher[1]. Therefore, the rapid increase in the number of road traffic accidents has become a worrying phenomenon. There are tens of thousands of passengers and drivers died from road traffic accidents each year[2], [3]. In recent years, vehicle detection has always been a topic of great interest to researchers[4], [8]. They are aimed to save lives and reduce the number of traffic accidents. There are various sensing modalities have been used for on-road vehicle detection, such as radar[5], lidar[6], [7] and computer vision[9], [10], [11]. Computer vision is getting more and more attention because cameras are becoming cheaper, smaller, and the quality is higher than ever before. Vision-based vehicle detection systems are very challenging problems[12], [13]. For real-world applications, they should be fast and robust enough to deal with different situations such as different viewpoints, illumination changes, saturation, occlusion and so on[14]. Fig.2 shows a few of examples under different situations in PASCAL VOC2007 car dataset[38] and KITTI object detection benchmark[14].

In recent years, CNN has been widely applied to all kinds of computer vision problems, such as image classification[15], [16], [17], [18], [19], [20], object detection[22], [42], [44], semantic segmentation[36], [50] owing to the powerful abilities of learning for discriminative features. Considerable advances have been made in object detection by the successful application of CNN. We can categorize these CNN-based object detection methods into two categories: two-stage and single-stage methods. In the first two-stage detectors[22], [42], [44], thousands of regions of interesting(RoIs) are proposed in the first stage, the decision refinement is performed in the followed second stage. These methods often get high performance in some benchmarks such as PASCAL VOC[38] and ImageNet[21]. Nevertheless, they are time-consuming and would be even worse when a great many of proposals are utilized. Apropos of the single-stage methods[23], [48], they frame object detection as a regression problem like human vision system. Without region proposal stage, bounding boxes and classification are straight from image pixels in a single network. They are extremely fast but low in accuracy when compared to those state-of-the-art detection methods.

In this paper, we propose an efficient yet accurate two-stage cascade detector which combines the advantages of the two kinds of object detection system. For real-world applications, computation time is as important as accuracy. To rapidly and accurately detect vehicles, the network is designed like two-stage detectors(accurate). What is new is that our network is designed as regression based(fast). The first stage provides a rough estimate of bounding boxes in each grid cell of the feature maps. Because of the existing of multiple Max-