CSC 2515 Project Proposal
- Autonomous Driving: Object Detection

Zhen Li
Department of Computer Science
University of Toronto
Toronto, ON, M5S 2E4
zhen@cs.toronto.edu

Zhicong Lu
Department of Computer Science
University of Toronto
Toronto, ON, M5S 2E4
luzhc@cs.toronto.edu

1 Introduction

The goal of this project is to detect the cars, pedestrians and cyclists in the given image, locating the 2D bounding box of the object correctly. Such detection techniques would make vision-based autonomous driving systems more robust and accurate, hence increasing the possibility of adopting them in the real life.

The basic idea of this project came from The KITTI Vision Benchmark Suite [2]. According to the website, an overlap of 70% is required for a detection of a car, while an overlap of 50% is required for the pedestrians and cyclists.

To achieve our goals, we will download the left color images of object, camera calibration and training labels data from `http://www.cvlibs.net/datasets/kitti/eval_object.php`, as well as the development kit, upon which we will train the model and evaluate our methods. The labeled training data will be splitted into training, validation, and testing, which takes 60%, 10% and 30% respectively. Besides KITTI, we would also use some other datasets to test our methods, especially on pedestrain detection, including Caltech Pedestrian Detection Benchmark[7] and Daimler Pedestrian Segmentation Benchmark Dataset[1].

Inspired by the ranking of different methods on the KITTI website, we would try to implement those with very good performence and without using other sensor data, for the reason that such methods are more accessible and have better potential to be implemented even on mobile devices. We would try to improve the training model based on the findings of the state-of-the-art to come up with our own method, and evaluate the performance of it.We expect that our method can reach a high accuracy on the test set with an optimized speed.

2 Related Work

Object detection, especially cars and pedestrains detecions for autonomous driving systems, has been a hot topic in computer vision for recent years. Convolutional Neural Network(CNN or ConvNet)[6] is able to learn the features of the object and handle variations such as poses, viewpoints, and lightings, with high accuracy and high efficiency. However, it doesn't perform well when occlusion occurs, which is often the case in pedestrian and cyclists detection.

DeepParts[3] is a method to solve such problems, which consists of extensive strong part detectors to detect pedestrian by observing only a part of a proposal. It can be trained on weakly labeled data and performs very well on the task of pedestrain detection.

Recently, the Fully Convolutional Neural Network (FCN) based methods[5], with end-to-end approach of learning model parameters and image features, further improves the performance of object detection. DenseBox[4] is a unified end-to-end FCN that directly predicts bounding boxes and object class confidences through all locations and scales of an image with great accuracy and efficiancy. It

also incorporates with landmark localization during multi-task learning and further improves object detection accuracy. It has the best accuracy on car detection on KITTI by the time the proposal is finished. However, it has not been tested on the tasks of pedestrain or cyclists detection.

Our method will be based on CNN and FCN, with some techniques to deal with occlusion and other issues. We will try to make it accurate on all of the three detection tasks. The performance of our methods will be compared with that of DeepParts and DenseBox.

References

[1] C. G. Keller, M. Enzweiler, and D. M. Gavrila, A new benchmark for stereo-based pedestrian detection, in 2011 IEEE Intelligent Vehicles Symposium (IV), 2011, pp. 691696.

[2] A. Geiger, P. Lenz, and R. Urtasun, Are we ready for autonomous driving? The KITTI vision benchmark suite, in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 33543361.

[3] Y. Tian, P. Luo, X. Wang, and X. Tang, Deep Learning Strong Parts for Pedestrian Detection, in 2015 IEEE International Conference on Computer Vision, 2015.

[4] L. Huang, Y. Yang, Y. Deng, and Y. Yu, DenseBox: Unifying Landmark Localization with End to End Object Detection, arXiv:1509.04874, 2015.

[5] J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, arXiv1411.4038, 2014.

[6] A. Krizhevsky, I. Sutskever, and G. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25, 2012.

[7] P. Dollar, C. Wojek, B. Schiele, and P. Perona, Pedestrian detection: A benchmark, in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 304311.