Report - 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

[old]The goal of this project is to detect the cars and pedestrians 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.

[old]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.

[old]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].

[old]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

[old]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.

[old]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.

[old]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.

[old]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.

In many tasks, since the number of images and windows to evaluate is huge, we often rely on a weak classifier to get proposals for the more expensive classifier. Selective Search [1] is a successful algorithm, which emphasize recall to include all image fragments of potential relevance.

3 Methods

4 Experiments

4.1 Data set

We use the KITTI data set for Object Detection [2], which contains 7481 color images with ground truth bounding box labels, including 28782 cars and 4487 pedestrians. We partition the data set to a training set contains 5237 images(70%), and a testing set contains 2244 images(30%). Then we run different algorithms on the data set and analyze the results.

4.2 Selective Search + SVM

4.2.1 Pre-process

First, we crop the cars and pedestrians from the images to generate the positive examples. We noticed that we have far more cars than pedestrians in the data set. To balance the difference, we multiply the pedestrian image set by adding a random bias to the color map (-20% to +20%) repeatedly. We also reverse the original image horizontally to make full use of the training set. Since the KITTI benchmark only evaluate objects larger than 25 pixels (height), we ignore these small objects in the training set. After all of these augmenting and filtering strategies, we have 22576 cars and 20960 pedestrians as positive examples for training.

We use the Selective Search [1] to generate the negative examples. We randomly select images from the training set, run Selective Search, and find out background segments, which have a 0% to 30% overlap with a positive example. We generate 20000 negative examples for training.

4.2.2 Training

4.2.3 Validation

4.2.4 Testing

4.3 Zhicong: TODO

5 Conclusion

References

[1] K. E. Van de Sande, J. R. Uijlings, T. Gevers, and A. W. Smeulders, “Segmentation as selective search for object recognition,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1879–1886.

[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*. IEEE, jun 2012, pp. 3354–3361. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6248074

Old 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.

[8] K. Sande, J. Uijlings, T. Gevers, and A. Smeulders, Segmentation as Selective Search for Object Recognition, in 2011 IEEE International Conference on Computer Vision, pp. 1879-1886.