EE596: Project Proposal

Team: Daniel Cashon-Smith (solo)

04/29/19

1 Project Idea

The goal, broadly speaking, is to implement an object detection and localization algorithm in the context of autonomous driving. Then, as a practical test, I plan to use my own car along with a single front-view camera to examine the effectiveness of the result. The goal is to gain a practical understanding of how this is done via reading papers and hands-on implementation.

2 Brief Background

The most natural questions to ask about an image are the following: Is there an object I am interested in present in the image? What exactly is that object, if it is present? Are there other objects, if so, can I also identify them? Where are these objects in the image? These questions form the basis for object detection, classification, and localization tasks. It is clear that any autonomous vehicle must possess a system that is able to do this with high accuracy - it must be able to identify and interpret objects such as stop signs, traffic lights, other vehicles, and pedestrians, to name a few possibilities, in order to make decisions about how to navigate. In addition, this system must be able to run in real-time, as the driving environment is in constant flux. One approach to accomplish this is to combine a multi-scale sliding window approach with a convolutional neural net (CNN) for object classification and bounding box identification. This approach seems to produce impressive results, and computational tricks that are enabled by use of the CNN to classify seem to circumvent the typically high computation time required by the sliding window technique [2]. The basic pipeline I have in mind is illustrated in Fig. 1:

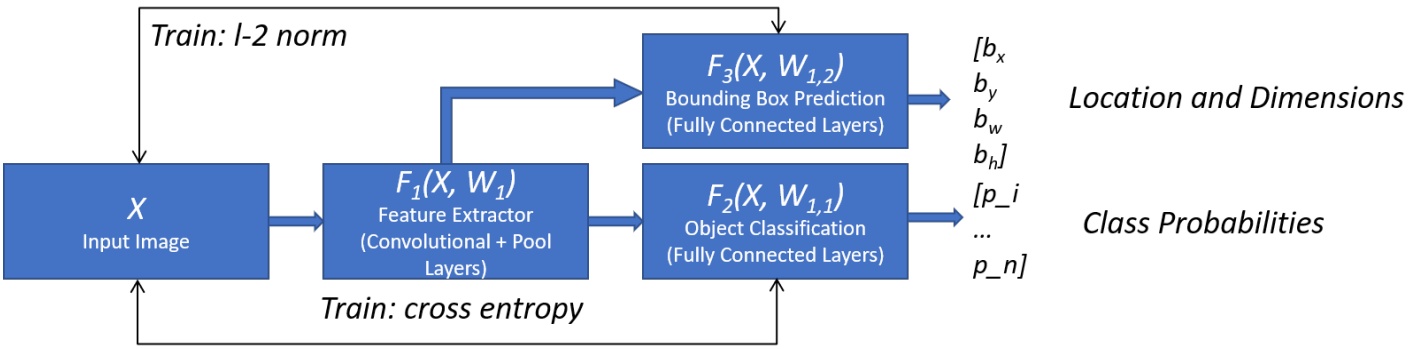

Figure 1: Basic Idea for General Object Detection via Multi-Scale Sliding Window

3 Relevant Datasets

There is not a shortage of data for this task, as it is a very active field. The two listed here seem particularly useful:

1. **Berkeley Deep Drive BDD100k** : (https://bdd-data.berkeley.edu/) Huge dataset of annotated images from hundreds of hours of driving.
2. **COCO**: (http://cocodataset.org/home) COCO includes hundreds of thousands of annotated images. There is a sufficient number which include common road objects: stop signs, people, traffic lights, bicycles, cars, etc.

References

[1] J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, Los Alamitos, CA, USA, jul 2017. IEEE Computer Society.

[2] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Robert Fergus, and Yann Lecun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In *International Conference on Learning Representations (ICLR2014), CBLS, April 2014*, 2014.