Object Detection and Its Implementation on Android Devices

Stanford | ENGINEERING
Electrical Engineering

Rao Zhang, Zhongjie Li
(zhangrao/jay2015)@stanford.edu

Stanford University

Introduction

Object detection is a very important task for different applications including autonomous driving, face detection, video surveillance, etc. CNN based algorithm could be a great solution for object detection with high accuracy. Besides, most current deep learning applications are running on servers or desktop computers. Considering there are a lot of mobile computing devices available, we implemented the CNN based object detection algorithm on Android devices. The model architecture is based on SqueezeNet and further improved to find bounding boxes for recognized objects. The total model size is around 8 MB and makes it run fast, especially on mobile devices.

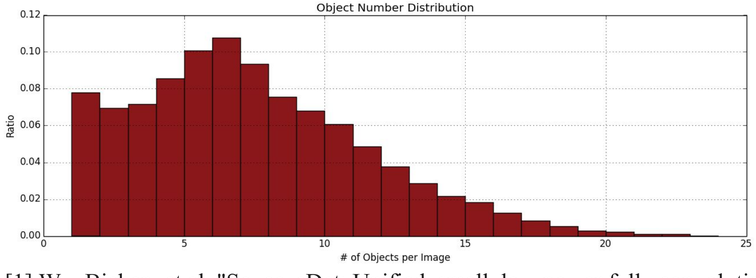
Problem Statement

To find and implement an object detection algorithm that's suitable for implementation on mobile platforms. The key requirements include:

- Low computation cost
- Low memory demand
- High efficiency

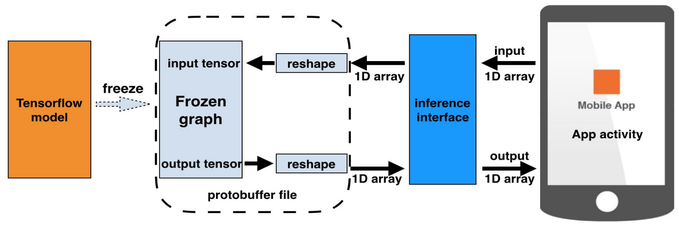
Datasets and Preprocessing

- Data Sets
  The dataset we use is the object detection data-set of KITTI Vision Benchmark Suite, which contains 7481 training images and 7518 test images. Total 80256 objects are labeled for this dataset and the 3 classes used for evaluation are cars, pedestrians and cyclists. The distribution of object number in the training data-set is shown below. 51865 objects are labeled, including 28742 cars, 4487 pedestrians and 1627 cyclists. On average: 3.8 cars, 0.6 pedestrian and 0.2 cyclist per image.
- Data processing
  - Image flipping
  - Random cropping
  - Batch normalization
  - Image variations of brightness/contrast/blur, etc.

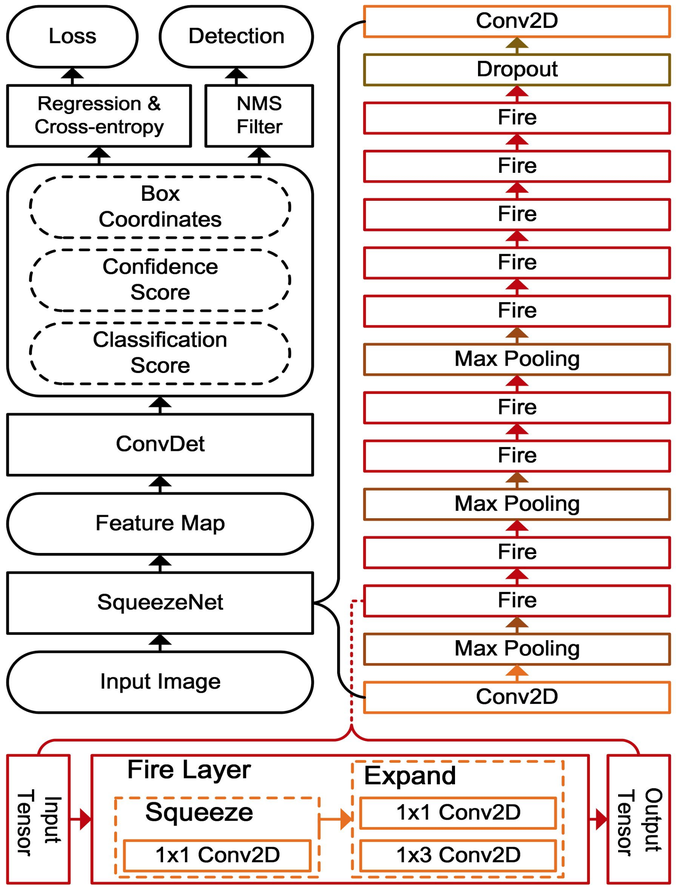


Approach

1. Android App Architecture



2. SqueezeNet + ConvNet Graph[1]

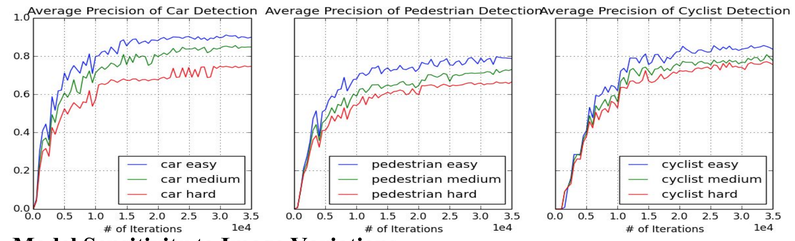


Results

1. Training Results

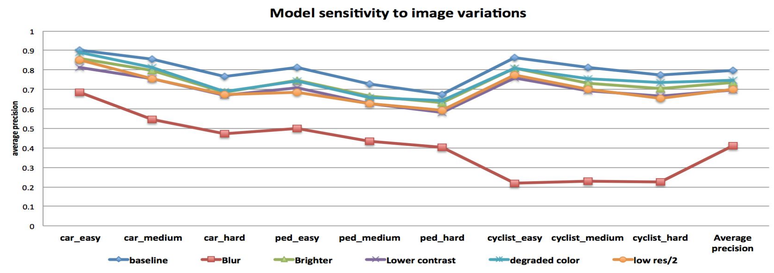
The squeezeDet model is trained with KITTI detection dataset. After ~35k steps of training, the overall recall can get 81%. The detection precisions are as following.

detection accuracy	car	cyclist	pedestrian
easy	90%	86%	80%
medium	85%	80%	74%
hard	75%	77%	67%



2. Model Sensitivity to Image Variations

The model is validated against multiple image viations, including blur, brightness, contrast, color and resolution. The results showed that the model is most sensitive to blur, especially for cyclist classifications.



[1] Wu, Bichen, et al. "SqueezeDet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving." *arXiv preprint arXiv:1612.01051* (2016).
[2] Iandola, Forrest N., et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size." *arXiv preprint arXiv:1602.07360* (2016)