Single Stage Detector for Object Detection

Jian Huang, Danyang Wang, Xiaoshi Wang
Department of Computer Science, Stanford University

Introduction

Object detection is a crucial task for computer vision. In advanced driver assistance systems (ADAS), accurately detecting cars and pedestrians plays a crucial rule on the safety of the autonomous actions.

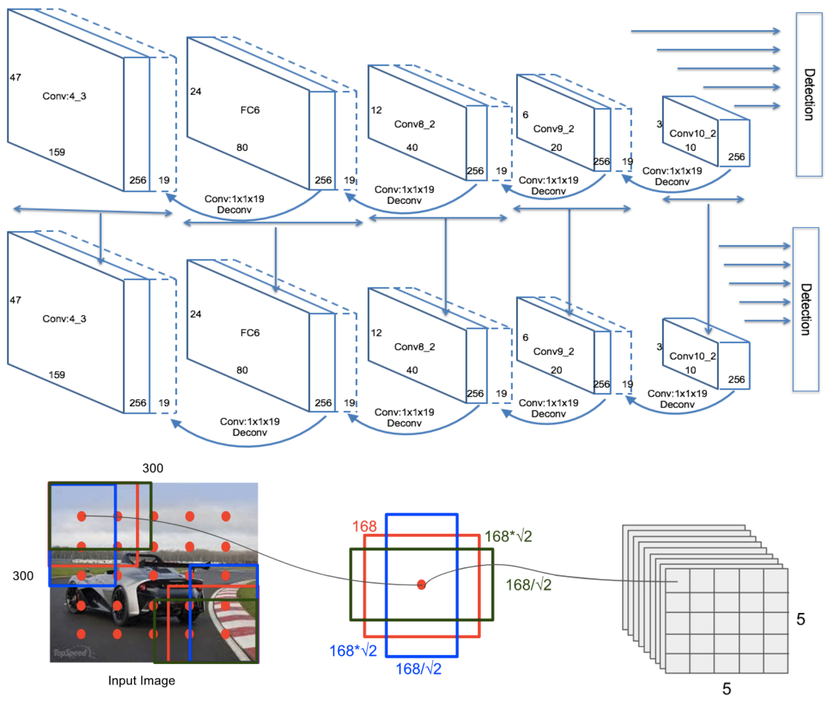
For objection detection, there are two model types applying convolutional neural nets: two stage methods like faster-rcnn and single stage methods like SSD. Traditionally single stage models are faster but less accurate than two stage models. Recently Ren et al published a paper in CVPR claiming they built a single stage model that is fast and accurate. However we find that their models are not fast enough for high resolution images(<1fps for 2560*768 images on a single K80 GPU in testing).

Thus we decide to improve their model and SSD model to achieve a model faster then Ren at al's work and keep high detection accuracy.

Dataset

We use KITTI Object Detection Evaluation dataset to evaluate our model. The dataset containing over 15000 2560*768 RGB images in png format. Images are taken in various real world road settings with multiple objects of different sizes and scales, occlusions and different lighting conditions. The training images are labeled out as images with cars, pedestrians, cyclists or just background with bounding boxes. The dataset and benchmarks is widely used in autonomous driving object detection researches.

Models



Backward Recurrent Rolling Convolution(BRRC)

- One important idea of original RRC model is to add high level abstract info from deeper level feature maps to low level feature maps. This idea is shown in the backward path of recurrent rolling.
- The other idea of RRC is to add "context" info from lower level feature maps to upper level maps. This idea is shown in the forward path of recurrent rolling.
- we argue that the first idea make sense, but second idea is redundant, also making the net harder and slower to train. Thus we only keep the backward path and throw away the forward path.

More Bounding Boxes

- The typical setting of single stage models is to have each point generate 4-6 bounding boxes.
- We increase the number of default boxes to eight so that ideally it is easier to train for a deep network.

Trade off between SSD and RRC

- We would like to invent a single stage model that can achieve a speed between SSD and RRC and has similar accuracy as RRC.
- Currently we finished two experiments, training SSD with more bounding boxes with a large dataset and training BRRC with more bounding boxes on a small dataset.

Results and Future Research

SSD + Bounding Boxs **BRRC** Original RRC

	IoU score	Time (s/iter)
SSD+Bounding Boxes	0.186	8
BRRC	0.518	62
Original RRC	0.530	68

- RRC based models outperform SSD significantly on accuracy.
- SSD is much faster than RRC based models in training speed.
- Our model is comparable to the original RRC model, and moderately faster.

In the next step, we plan to run two experiments:

- Run BRRC on large dataset and compare results.
- Apply some variants of models in image segmentation to detection, for example VGG + conv + deconv, without recurrent rolling.