Project 3: Fast R-CNN for object detection

Yufei Hu (404944367)

November 23, 2017

1 Objectives

Recently, deep CNN have significantly improved image classification and object detection accuracy. Compared to image classification, object detection is a more challenging task that requires more complex methods to solve. Due to this complexity, current approaches train models in multi-stage pipelines. In this project, we learn a popular model for object detection, Fast R-CNN. A Fast R-CNN network takes as input an entire image and a set of object proposals. The proposals are obtained by using a region proposal algorithm as a pre-processing step before running the CNN. The proposal algorithms are typically techniques such as EdgeBoxes or Selective Search, which are independent of the CNN.

The network first processes the whole image with several convolutional (conv) and max pooling layers to produce a conv feature map. Then, for each object proposal a region of interest (RoI) pooling layer extracts a fixed-length feature vector from the feature map. Each feature vector is fed into a sequence of fully connected (fc) layers that finally branch into two sibling output layers: one that produces softmax probability estimates over K object classes plus a catch-all background class and another layer that outputs four real-valued numbers for each of the K object classes. Each set of 4 values encodes refined bounding-box positions for one of the K classes. In this project, we use pre-trained fast R-CNN model to do object detection task.

2 Single-class object detection in one image

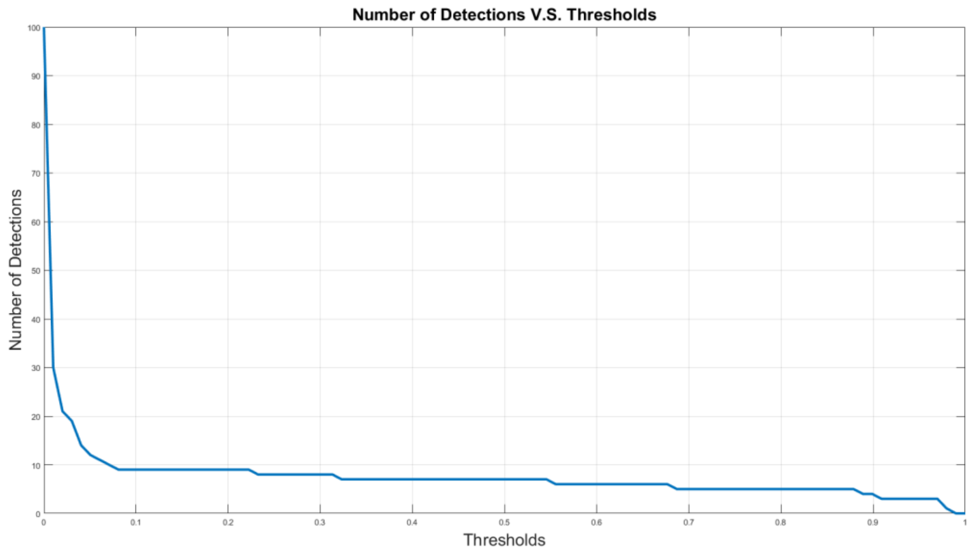

Figure 1: Number of Detections V.S. Thresholds

Before thresholding, 100 detections with highest scores are chosen. As the threshold increases, the number of positive detection decreases. Finally, I chose 0.6 as threshold. The detection result is shown below. Notice I have chosen 30% as the overlapping threshold for non-maximum suppression (NMS).

Figure 2: Car detection result

3 Object detection on Pascal VOC 2007 dataset

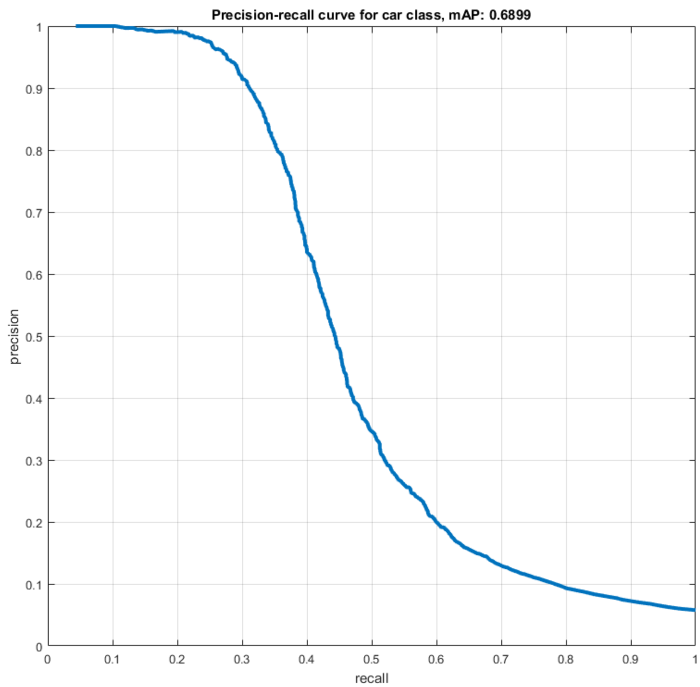

Figure 3: Precision-Recall Curve for the Car Class

As I gradually increase the threshold for judging car class, precision increases while recall decreases. Notice I have chosen 30% as the overlapping threshold for non-maximum suppression (NMS).

Figure 4: Detection result 1

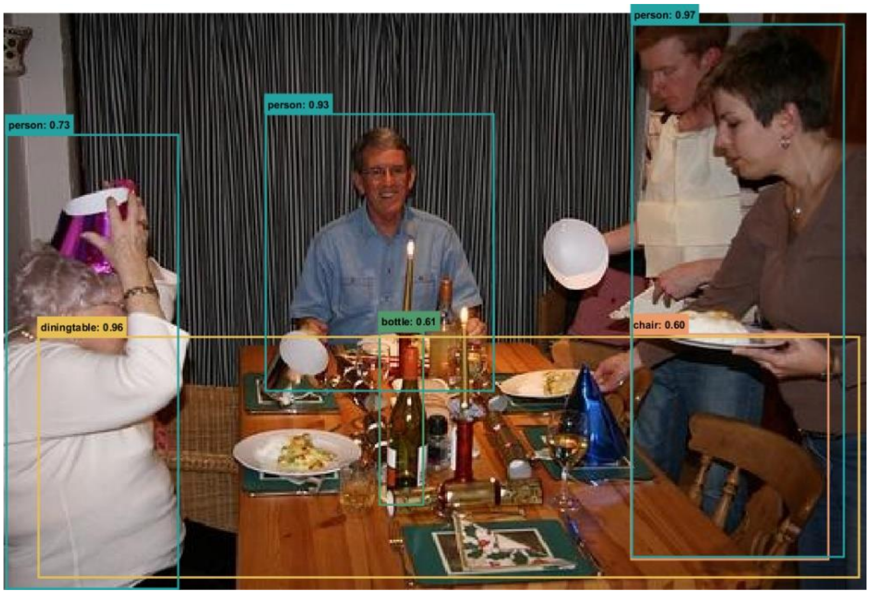

Figure 5: Detection result 2

Figure 6: Detection result 3

The detection results shown above are quite reasonable though there could be some missing detections.

Finally, the average precision and the total mean average precision are listed in the table below:

I have to say the precisions are pretty good, except for detecting tv monitor, potted plant, and bottle.

Table 1: Average Precisions for VOC 2007

Class	AP
aeroplane	0.869767441860465
bicycle	0.868913857677903
bird	0.785074626865672
boat	0.695340501792115
bottle	0.415343915343915
bus	0.721973094170404
car	0.758893280632411
cat	0.790190735694823
chair	0.512871287128713
cow	0.686567164179105
dining table	0.658730158730159
dog	0.842666666666667
horse	0.721354166666667
motorbike	0.797872340425532
person	0.797419646899049
potted plant	0.347923681257015
sheep	0.727272727272727
sofa	0.627027027027027
train	0.792452830188679
tv monitor	0.345885634588564
Total	0.689900093662192