Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

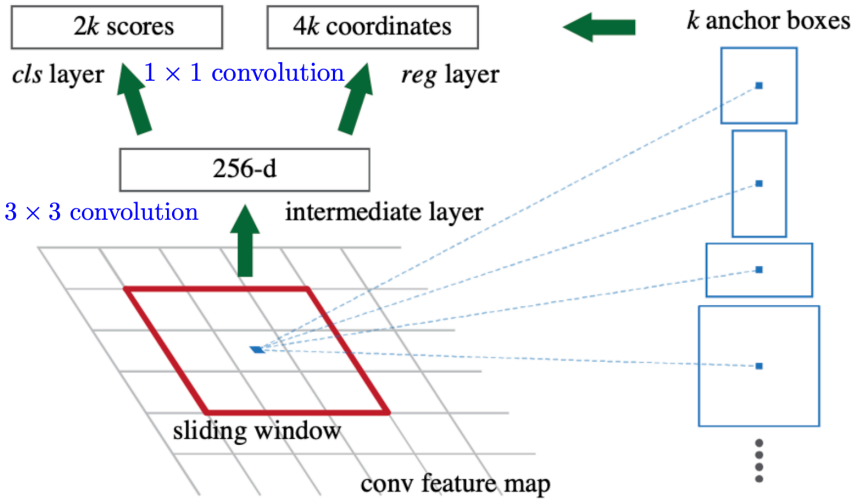
Boulder

YouTube Playlist

Region Proposal Network (RPN)

An RPN is a fully-convolutional network that simultaneously predicts object bounds and objectness scores at each position.

RPN and Fast R-CNN can be trained using a simple alternating optimization to share convolutional features.



Translation-Invariant Anchors

9 anchors: 3 scales & 3 aspect ratios

$WHk \rightarrow$ total anchors for a feature map of size $H \times W$

Loss Function (Region Proposals)

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)$$

$i \rightarrow$ index of an anchor in a minibatch

$p_i \rightarrow$ predicted probability of being an object

$p_i^* = 1 \rightarrow$ if anchor is positive

$p_i^* = 0 \rightarrow$ if anchor is negative

For anchors, use 3 scales with box areas of 128 x 128, 256 x 256, and 512 x 512 pixels, and 3 aspect ratios of 1:1, 1:2, and 2:1.

Positive label is assigned to two kinds of anchors: (i) the anchor/anchors with the highest Intersection- over-Union (IoU) overlap with a ground-truth box, or (ii) an anchor that has an IoU overlap higher than 0.7 with any ground-truth box. Negative label is assigned to a non-positive anchor if its IoU ratio is lower than 0.3 for all ground-truth boxes.

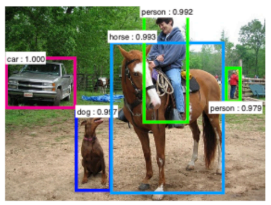
Alternating Training

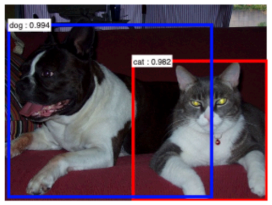
First step: Train the RPN initialized with an ImageNet-pre-trained model and fine-tuned end-to-end for the region proposal task.

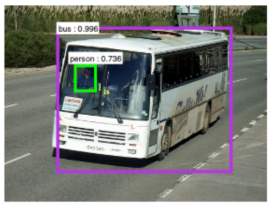
Second step: Train a separate detection network by Fast R-CNN using the proposals generated by the step-1 RPN. This detection network is also initialized by the ImageNet-pre-trained model.

Third step: Use the detector network to initialize RPN training, but we fix the shared conv layers and only fine-tune the layers unique to RPN.

Finally: Keeping the shared conv layers fixed, fine-tune the fc layers of the Fast R-CNN.







Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems*. 2015.