

# **Review of “Faster R-CNN: Towards real-time object detection with region proposal networks”**

Riyad Bin Rafiq

## **1. Paper summary**

In the field of object detection, region proposal computation was a bottleneck until the idea of Faster R-CNN came out. This paper [1] introduced a region proposal network (RPN) which was a fully convolutional neural network (CNN) generating ROIs nearly cost-free from an input image. The combination of RPN and Fast R-CNN was trained end-to-end in four steps with shared convolved features. This detection system obtained 73.2% mAP and 70.4% mAP on PASCAL VOC 2007 and 2012 respectively using 300 proposals.

## **2. Contribution**

### **2.1 Region proposal network**

In my opinion, this region proposal network is one of the most notable contributions of the paper. This network outputs the class and anchor boxes and as both RPN and Fast R-CNN share common convolution layers, the whole network can be fine-tuned to detect objects with greater accuracy. Moreover, the generation of proposals was nearly cost-free as a CNN was used instead of an expensive selective search method.

### **2.2 Translation invariant**

Translation invariance means that if the pixels of the image (containing object) move in the same amount in the same direction, the object can be identified as an object. The proposed approach was translation invariant while the MultiBox method [2] was not. That's why nine anchors worked for the Faster R-CNN whereas the MultiBox method needed 800 anchors.

### **2.3 Experiments**

The authors did different kinds of experiments to investigate the behavior of RPN. For example, they stopped training after the second step to see the results on how detector-tuned features influence to fine-tune the RPN, distill RPN from the whole network, etc. These trials provided a better intuition of the RPN and made the network more robust and worthy.

## **Critique**

### **2.1 Joint optimization**

The authors trained the combined RPN and Fast R-CNN separately in 4 steps. They also said that it was not easy to define a single network including RPN and Fast R-CNN and trained the network jointly. But it was not clearly mentioned why they chose 4-steps training instead of jointly training the network. Did they try the simultaneous training first? Although they mentioned an assumption, in my opinion, the reason was not properly explained.

### **2.2 RPN proposals**

The research trained Fast R-CNN with top-N ranked (2k) RPN proposals and evaluated it with different numbers of proposals such as 300, 1000. But they didn't investigate how it performed with fewer proposals in training although they demonstrated it during testing. Moreover, the same number, 2000 proposals were being used for training the network which was also done for the prior detection methods.

## **Reference**

1. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst. 2015;28. Available: <https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html>
2. Szegedy C, Reed S, Erhan D, Anguelov D, Ioffe S. Scalable, High-Quality Object Detection. arXiv [cs.CV]. 2014. Available: <http://arxiv.org/abs/1412.1441>