

# **Review of “Fast R-CNN”**

Riyad Bin Rafiq

## **1. Paper summary**

The paper [1] proposed a fast region-based convolutional neural network (CNN) for object detection. This paper improved the performance and modified this research [2]. In this paper, the authors proposed a single-stage training algorithm that can learn object proposals and localisations simultaneously. The algorithm is called Fast R-CNN.

## **2. Contribution**

### **2.1 Speed**

The significant contribution of the paper is speed in training and testing the network. The paper showed that the method trained the very deep VGG16 network 9 times and 213 times faster than R-CNN in training and testing respectively. Fast R-CNN is also speedy compared to SPPnet. In this research, the authors introduced truncated singular value decomposition (SVD) to accelerate the computing of the fully connected layers in detecting objects. They also demonstrated what portion of the layers in a network took time in training and testing with and without the truncated SVD.

### **2.2 Simultaneous process**

Fast R-CNN is a single stage learning algorithm using a multi-task loss. The algorithm has two sibling output layers: one is discrete probability distribution and another one is bounding-box regression over the number of classes. The research introduced a multi-task loss,  $L$  that jointly trains for classification and bounding-box regression. So this approach saved a lot of time in training and testing the network.

### **2.3 Answers to several questions through experiments**

Another important contribution of the paper is that it experimented on a specific dataset to evaluate design decisions as well as to understand the comparison between Fast R-CNN, R-CNN and SPPnet. For example, whether the introduced multi-task training improves detection accuracy or not, if more proposals are always better or not and so on. This research discussed these kinds of important issues with the result. In my opinion, this provides a great intuition in the computer vision community.

### **3. Critique**

#### **3.1 Region proposals**

In the research, it was stated that a Fast R-CNN takes an entire image and a set of object proposals as input. But it was not mentioned how many object proposals should be considered for such a network though the authors discussed a number of proposals with varied average recall (AR) and mean average precision (mAP). In my opinion, this object proposal idea was a bottleneck as each proposal needed to be processed in the network.

#### **Reference**

1. Girshick R. Fast r-cnn. Proceedings of the IEEE international conference on computer vision. 2015. pp. 1440–1448.
2. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. pp. 580–587.