

Fast R-CNN

Paper Summary

Ross Girshick
Microsoft Research

1 Summary

This paper is proposed by the same author of R-CNN. It's an improved version of R-CNN. It solved the limitations of R-CNN and introduced Region of Interest (ROI) pooling layer that can extract feature for all proposals with same length. Also, replaces the three-stage training by an only one stage network. Fast R-CNN is significantly more accurate than R-CNN and takes low disk space as well. Fast R-CNN trains the VGG16 network on PASCAL VOC 2012 dataset, 9 times faster than RCNN and 213 times faster for testing. It outperforms R-CNN in terms of mAP as well. VOC07, VOC 2010, and VOC 2012 were used to evaluate the network.

2 Contribution

In Fast RCNN, the CNN is fed with the input image to produce the convolutional feature maps. The regions of proposals are extracted using these maps. The proposed regions are then all reshaped into a fixed size using a RoI (Region of Interest) pooling layer so they can be fed into a fully connected layer. In order to introduce scale invariance, input images are resized into a randomly sampled size at train time in the Fast R-CNN multi-scale pipeline. This way back propagation is done using RoI pooling. The FC network is used under a softmax layer. These are the ImageNet models used for experiments: AlexNet Model (Small), VGG CNN M 1024, which has the same depth as S but is broader. M (Medium): VGG CNN M 1024, which has the same depth as S but is wider and L (Large): VGG16 model with a lot of depth.

3 Limitations

Fast R-CNN also uses selective search to identify the Regions of Interest, which is a labor - intensive and time-consuming process. Though compared to RCNN, it detects objects in each image in around 2 seconds, which is a significant improvement. However, even a Fast RCNN lacks its speed when huge real-world datasets are taken into account.

4 Result

Fast R-CNN processes images 45 times more quickly during testing than R-CNN, and 9 times quicker during training. In addition, it trains 2.7 times quicker and executes test pictures 7 times faster than SPP-Net. With continuous usage of shortened SVD and just a 0.3 drop in mAP, the network's detection time is reduced by more than 30