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

Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun

1   Summary

By substituting Region Proposal Network for Faster R-CNN, the problems with Selective Search is resolved (RPN). Using ConvNet, they first extract feature maps from the input picture, and these maps are then sent through an RPN to provide object recommendations. These maps are finally classed, and the bounding boxes are projected.Feeding the ConvNet an input image, will return feature maps for the image. After that, they used the Region Proposal Network (RPN) on these feature maps to obtain object proposals and applied RoI pooling layer to reduce the size of each proposal to the same level. In order to classify any predictions of the bounding boxes for a image, a fully connected layer is used in the end.

2   Contribution

The Faster R-CNN fixes previous issue by generating the region proposals using the RPN, which is another neural network. The region proposal method was the only remaining component of the network in Fast R-CNN. Also, R-CNN and Fast R-CNN both were CPU-based region proposal techniques, such as the Selective search algorithm, which works on CPU computing and takes around two seconds per picture. But Faster R-CNN uses a region proposal network instead of a selective search algorithm to speed up the whole process.This improves feature mapping overall by reducing the region proposal time per picture from 2 seconds to 10 milliseconds and enabling the region proposal step to share layers with the succeeding detection stages.So, basically the foundation of faster RCNN is standing on - RPN: To generate region suggestions, and Rapid R-CNN: for item detection in the proposed locations.

3   Limitations

The RPN is trained using a single picture to extract all anchors in the mini-batch of size 256, which is one disadvantage of Faster R-CNN. Because all samples from a single picture could be associated, it might take the network a while to reach convergence.Object proposal requires time, and because there are multiple stages operating simultaneously, each system's performance is influenced by the performance of the one before it.

4   Result

PASCAL VOC 2007 benchmark dataset was used for the experiment. This dataset includes around 5,000 test and 5,000 trainval images from 20 item categories. Additionally, they offer PASCAL VOC 2012 benchmark data for a few models.The RPN approach beats Selective search and edgebox by 1.3 mAP on the PASCAL VOC 2007 test set using Fast R-CNN with ZF detectors but multiple different proposal methods.