Faster R-CNN

Title: Faster R-CNN

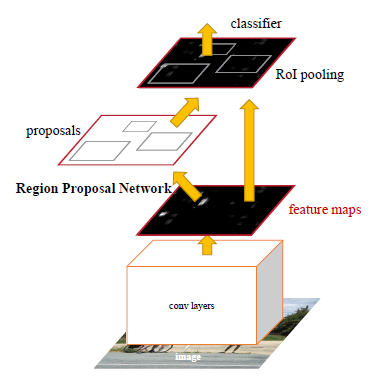
Initial Goal

Region Proposal step in Fast R-CNN is still slow. Because it uses the Selective Search algorithm. The initial goal of this paper is to use a newer technique instead of the Selective Search algorithm which is really slow.

Contributions

1. Region Proposal Networks: Instead of using the Selective Search Algorithm, it is employed to discover region suggestions. RPN is a fully convolutional network that produces region proposals in a variety of sizes and aspect ratios. The RPN instructs the object detection (Fast R-CNN) on where to look using neural network terminology and attention.
2. Anchor Boxes: Rather than employing image pyramids (many instances of the same picture at different scales) or filter pyramids, this study presents new "anchor" boxes that act as references at various sizes and aspect ratios.
3. This research combines RPNs with Fast R-CNN by presenting a training approach that alternates between fine-tuning for region proposal and then fine-tuning for object recognition while keeping the proposals stable.
4. 4. When compared to the strong baseline of Selective Search with Fast R-CNNs, RPNs with Fast R-CNNs exhibit greater detection accuracy on PASCAL VOC detection benchmarks. With a frame rate of 5 frames per second, the effective running time for this proposal is under 10 milliseconds.
5. 5. Faster R-CNN and RPN are the foundation of multiple first-place submissions in the ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation tracks in the ILSVRC and COCO 2015 competitions.

Model Architecture



Faster R-CNN is made up of two modules:

1. Deep fully convolutional network that proposes regions.
2. The Fast R-CNN detector

Region Proposal Network:

A fully convolutional network called a Region Proposal Network (RPN) accepts an image (of any size) as input and produces a series of rectangular object suggestions, each with an objectness score. As its ultimate objective is to share compute with a Fast R-CNN, the RPN network processes the pictures using the same convolution layers as the Fast R-CNN.

Anchor Boxes:

An anchor is centered on the sliding window in question and has a scale and aspect ratio linked with it. They employed three scales and three aspect ratios by default, resulting in k = 9 anchors at each sliding point.

How does it work

1. Region Proposal Network generates region proposals with help of anchor boxes. RPN guides the Fast R-CNN module to look for objects.
2. The ROI Pooling layer extracts a fixed-length feature vector from each region proposal.
3. Fast R-CNN is then used to classify the extracted feature vectors.

Advantages

1. The pictures are processed by the RPN network using the same convolution layers as the Fast R-CNN. As a result, as compared to Selective Search, RPN takes less time to develop region proposals.
2. Because RPN and Fast R-CNN have several convolution layers in common, they can be combined into one network. As a result, just one training session is required.
3. Faster and more accurate than Fast R-CNN.

Limitations

One disadvantage of Faster R-CNN is that the RPN is trained using a single picture to extract all anchors in the mini-batch of size 256. The network may take a long time to attain convergence since all samples from a single picture may be correlated (i.e. their characteristics are similar).