SSD: Single Shot MultiBox Detector

Link

arXiv

Summary

- This paper introduces SSD (Single Shot Detector) network for object detection which is significantly faster than previous state-of-the-art detectors. It does not require region proposals and is as accurate as slower techniques that does explicit region proposals and pooling like faster R-CNN. The core of SSD is predicting category scores and bounding box offsets for a fixed set of default bounding boxes. To achieve high accuracy they produce predictions of different scales from feature maps of different layers in the network and explicitly separate predictions by aspect ratio.
- Like the anchor boxes faster R-CNN they use default boxes of different size and aspect ratios and the network predicts box offsets and per class score for each box. However unlike Faster R-CNN they apply them to several feature maps of different resolutions.
- Since the default boxes don't align perfectly with ground truth boxes they need to determine which default boxes correspond to a ground truth box. To do so they first match each ground truth box to the default box with highest jaccard overlap. Then the match each default box to any ground truth box with jaccard overlap higher than a threshold 0.5.
- For hard negative mining they sort the negative examples using highest confidence loss (probability of being picked as an object) for each default box and pick the top ones so that the ratio between negatives and positives is at most 3:1.
- Compared to R-CNN, SSD has less localization error. However, SSD has more confusion with similar object categories, partly because locations for the objects are shared during training. SSD also has much worse performance on smaller objects than bigger objects. It performs well on bigger objects and is robust to different aspect ratios because default boxes of various aspect ratios are used during training.
- Use of data augmentation alone improves mAP by 8.8% for SSD.
- Having multiple output layers at different resolutions is also very critical for the performance.