2020-10-28 3:06:46 PM

Uses multiple points (9 in the paper) instead of just the four bounding box corners to define a better fitting object region

multistage regression idea from cascade RCNN is used to iteratively refine the points

Some sort of deformable convolution is used to obtain a feature map for a given set of points

The loss itself still seems to be limited to the bounding box corners since that is all that is present in the ground truth and these are generated using some sort of heuristics based best fit Bounding box method which is, however, differentiable to allow end to end pipeline involving all the points

yolov1 like Center points based representation to cover the space of bounding boxes instead of anchor boxes

Lots of stuff borrowed from retinanet including FPN backbone and focal loss

In essence, it seems to be a cross between yolov1 and retinanet with deformable convolution

Performance seems only slightly better than existing stuff Including retina net and yolov1 with newer backbone and other tricks and might well be within the range of hand tuning improvements