PhD

FoveaBox Beyond Anchor-based Object Detector ax1904.03797

2019-12-16 11:28:36 AM

Good literature overview of automatic anchor box learning techniques

Supposedly free of anchor boxes but still predicts multiple boxes for each spatial location in the feature map which means that the main problem of gross over prediction compared to the ground truth still remains

Only problem it seems to address is that the likely aspect ratios of the boxes do not have to be chosen apriori which also allegedly makes it more invariant to distortions like image elongations

Multiple scales are still used through the feature pyramid network thingy

Network architecture is made up of a feature extraction backbone followed by a couple of subnetworks – one for classifying each pixel and one to predict the corresponding bounding box coordinates

Some sort of scale binning based assignment of ground truth objects to different scales of the feature pyramid to allow each level to be responsible for only certain scales of objects – seems to be riddled with heuristics

name is supposed to be inspired by the fovea of the human eyes in that there would be a local attention mechanism which is however implemented using a mess of heuristics based on some sort of shrinking of ground truth box before assigning labels

Process of selecting the final boxes for evaluation is also riddled with heuristics including a pyramid level specific thresholding to select just the 1000 maximum scoring boxes followed by 0.5 threshold NMS to select only 100 in all though it is not at all clear how such a process would be useful for practical detection where usually less than 5 objects will be there in any given frame