2020-06-14 11:57:27 AM

Seems to be built on top of corner net with the main contribution being that the post prediction point merging stage as been done away with to be replaced by direct regression of the bounding box height and width

The network outputs a simple key point heat map and is trained on a Gaussian weighted mask generated from the centroids of all the objects in the training image

Some sort of encoder decoder architecture is used similar to corner net with 4 different backbone architectures having been experimented with

Accuracy is not necessarily better than other detectors but is seem to be compatible and apparently is much faster

NMS as also apparently been done away with although no explanation is given as to how the 100 top peaks that are chosen from the heat map are then reduced to the actual boxes

regression outputs that accompany each pixel of the heat map can apparently be used for predicting other stuff such as 3D detection as well as human pose estimation and seem to be able to work quite well for all of these cases as well

Training of these outputs is only done for the pixel corresponding to the centroid of each actual object rather than all the pixels