

2020-07-19 10:01:03 PM

- Builds on the center net detector by adding the offset of each detection from its corresponding detection in the previous frame to each point output and then using minimum distance as a simple metric for associating boxes in consecutive frames
- No global context or long term occlusion handling
- A heat map of the predicted boxes – each of which is simply a point – is used as an additional input
- apparently some very aggressive data augmentation is needed to prevent the network from learning the trivial solution of simply never moving objects around to minimize the training error
 - in fact it is so aggressive that even the resultant hallucinated motion from static images is supposedly enough to train it to an extent