PhD

YOLO9000 Better, Faster, Stronger ax16_12

2018/12/09 9:38 AM

Seems to be mostly about using a bunch of tricks from pre-existing works to improve upon the original yolo and fix its 2 main shortcomings – poor localization and low recall;

tricks include:

- batch normalization [2% map increase]
- fine tuning the classification network itself using high-resolution images [4% map increase]
- anchor boxes[0.3% decrease in map but 7% increase in recall]
- using IOU clustering to get the anchor boxes instead of hand picking them[6.2% map increase compared to RCNN with 9 clusters]
- Parameterizing bounding boxes relative to the center of the grid with the predicted values constrained to be between 0 and 1 [5% map increased]
- Using higher resolution features by using a passthrough layer to combined features from the penultimate layer to those from the last one by first reshaping the former from 26 x 26 x 512 into 13 x 13 x 2048 and then concatenating them;
- Performing multiscale training by randomly changing the shape of the input image to vary from 320 x 320 to 608 x 608;

It also performs joint classification and detection training by using some kind of hierarchical classification that takes advantage of the structure that exists between the classes in ImageNet to construct something called a word tree;
it back propagates on the detection specific elements of the loss when the corresponding data is available but only on the classification loss when it is not;