PhD

Faster R-CNN Towards Real-Time Object Detection with Region Proposal Networks tpami17 ax16_1

2018/11/05 12:38 PM

2018/12/07 1:30 PM

region proposal network is implemented as a small network which is used in a sliding window manner, that is the patch corresponding to a small 3 x 3 window is extracted from a convolutional feature map and this is used as input to the network which then outputs for by K and 2 by K numbers for the box regression and classifier subnetworks respectively;

the point to note is that but it is the same network that is used at all the spatial locations so that the parameters are all shared

The region proposal in the box classification networks share quite a few up there fully convolutional layers [5 to 13] and it is output feature map of the last shared layer which becomes the input for the sliding window thing;

the K proposed boxes are parameterized relative to K reference boxes or anchors that are defined for each spatial location;

it is not quite clear how but somehow this is supposed to make the proposals translation in variant;

An anchor is assigned a positive label if either it has a maximum overlap with the ground to box or its overlap with any ground truth box > 0.7;
it is assigned the negative label if its overlap is < 0.3 with all of the ground truth boxes;
boxes that satisfy neither condition are not included in the loss function;

One thing that is not at all clear is as to how the correspondence is obtained between the 3 x 3 sliding windows that are defined in the space of the feature map and the anchor boxes as well as a ground truth boxes that are defined in the original image space

As far as I can see, the loss function must be defined for each forward pass of the small convolutional network on the patch extracted from a 3 x 3 sliding window but to do this, there must be a way to relate the sliding window location in the feature map to the corresponding location in the order in the limited space where the anchor’s and the ground truth boxes are defined;

The shared training is carried out using a 4 step alternating process – train the RPN, train the detector, use the detector weights to initialize the RPN weights and retrain only the layers unique to it, train only the detector layers while keeping the shared layers fixed

Out of the total 20,000 proposed boxes only about 6000 remain after removing cross boundary boxes and non-maximum suppression is applied to these to retain only about 2000 boxes for the final training