2019/05/16 10:38:40 AM

- extends Faster RCNN by replacing the ROI evaluation network by an FCN so all output computations can be shared to provide further speedups

- this raises the issue of simultaneously achieving location invariance, so that an object might be detected irrespective of where it is located within the image, and location sensitivity so that it is localized accurately

- this is resolved by using special purpose feature maps called **position sensitive score maps** where each ROI is divided into a regular grid of *k x k* subregions or bins and each is represented for each class by a single feature map so that there are a total of *k^2*(C+1)* feature maps for C classes

- each such feature map is designed to be activated when a specific part of an object of that class is encountered – the part corresponding to a particular ROI bin

- ROI pooling for each bin and category is then performed only over the corresponding feature map
    - Average pooling is performed in the paper though max pooling can be done as well

- this gives an independent score for each of the bins which are combined by a voting strategy to classify the overall ROI

    - Simple averaging over the *k^2* scores is performed to give a C+1 dimensional vector which is then subjected to soft max to obtain the final classification for the ROI

- similar strategy is used for bounding box regression as well by having an additional sibling 4*k^2 feature maps

- each of these two sets of feature maps is produced by a single convolutional layer of the appropriate dimensionality - *k^2*(C+1)*-d and *k^2*4*-d

- supposed to be 2.5 – 20 times faster than Faster RCNN but doesn't seem to be so, at least in the TF object detection API