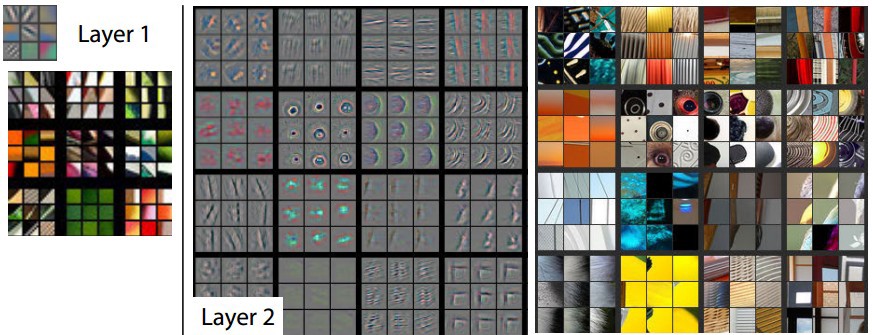
**Faster R-CNN (Region based Convolutional Neural Network):**

Feature Extraction:

VGG-16 is used as CNN for R-CNN. VGG is used to extract features of the image.

Each convolutional layer creates abstractions based on the previous information. The first layers usually learn edges, the second finds patterns in edges in order to activate for more complex shapes and so forth. Eventually we end up with a convolutional feature map which has spatial dimensions much smaller than the original image, but greater depth.

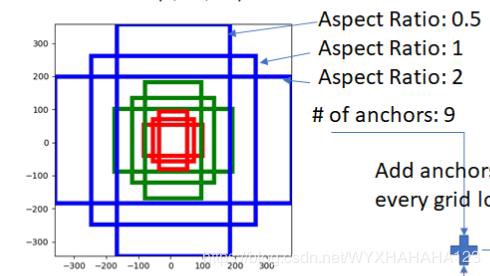


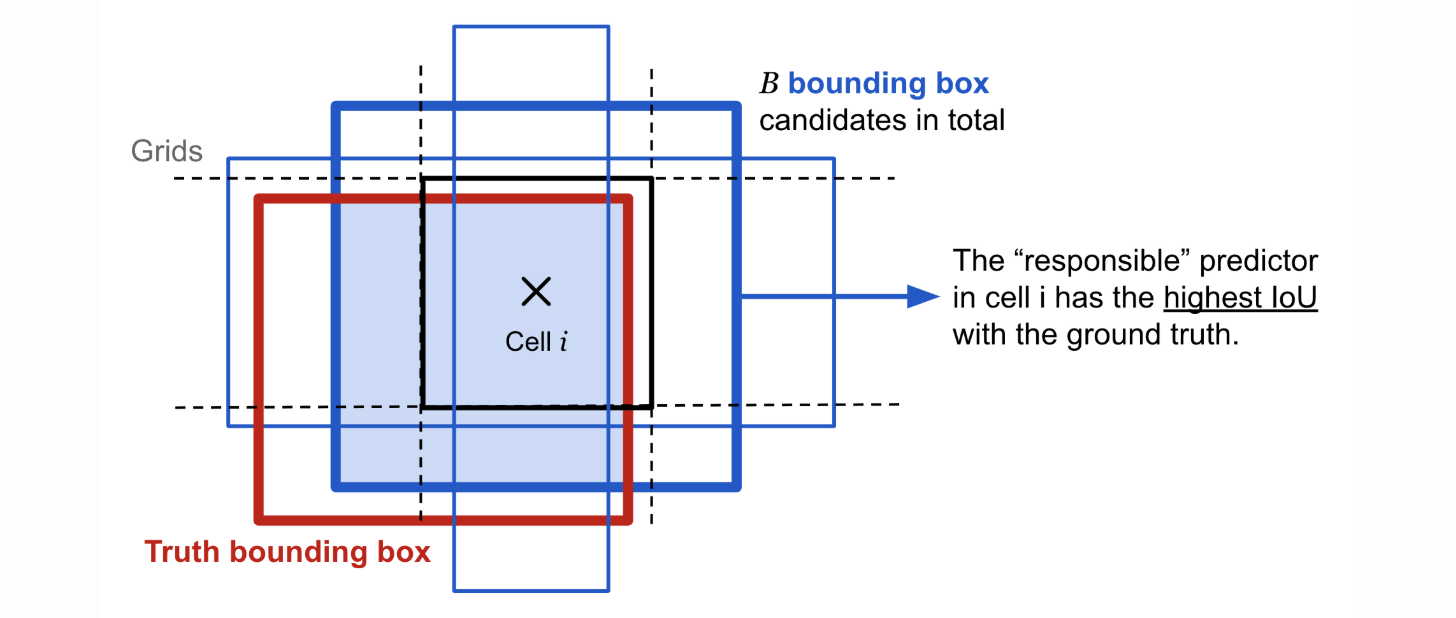
Anchors:

Anchor boxes with 9 different aspect ratios is run over the convolutional feature map in the last layer where all the features are really high level and matured. This anchor coordinates are then mapped to the coordinate in the original image.

RPN (Region Proposal Network) takes all the reference boxes (anchors) and outputs a set of good proposals for objects.

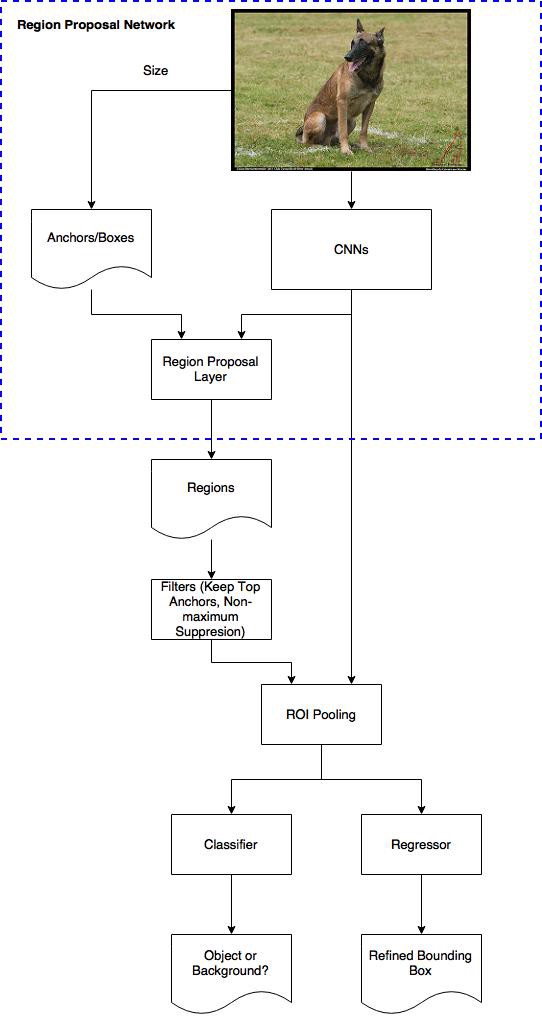
RPN does two different type of predictions: the binary classification and the bounding box regression adjustment. We turn to using the anchors with the biggest IoU to the ground truth boxes.





ROI:

After the RPN step, we have a bunch of object proposals with no class assigned to them. Our next problem to solve is how to take these bounding boxes and classify them into our desired categories. The location of object proposals are cropped in convolutional feature map and then each crop is resize to a fixed sized 14×14×*convdepth* using interpolation (usually bilinear). After cropping, max pooling with a 2x2 kernel is used to get a final 7×7×*convdepth* feature map for each proposal.



<https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2>