3. a) This operation is called non-maximum suppression and removes boxes with an ~~overlap~~ IoU higher than a set threshold for another box with higher confidence score.

3. b) False, the deeper layers have lower resolutions and are used to detect larger objects. The smaller objects are detected by higher resolution feature maps earlier in SSD.

3. c) They use different bounding box aspect ratios at the same spatial location to cover a lot of different object types at the location. For example a car and a person have different aspect ratios and using different aspect ratios allow one bounding box to approach the true boundary ~~for~~ the object no matter what shape the object has.

3. d) The main difference between SSD and YOLO is the use of multi-scale feature maps and convolutional predictors.
SSD uses feature maps of different sizes to detect objects of varying size, as well as convolutional filters which produce category scores or shape offsets.
YOLO uses a single scale feature map and a fully connected layer for predictions.

3. e) For this feature map we have

$$H \cdot W \cdot 6 = 38 \cdot 38 \cdot 6 = \underline{\underline{8664}}$$

anchor boxes.

3. f) In total we have

$$6 \cdot \left(\sum_{i=1}^{6} H_i \cdot W_i \right) = 6\left(38^2 + 19^2 + 10^2 + 5^2 + 3^2 + 1^2\right) = \underline{\underline{11640}}$$

anchor boxes for the entire network.