3 a)

**non-maximum suppression (nms)**

**3 b)**

**False -> Deeper layers -> lower resolution -> bigger objects.**

**The statement is false, as deeper layers have lower resolution, and is therefore able to detect bigger objects.**

**3 c)**

**The use different bounding box aspect ratios to get different shapes in the bounding boxes, for instance some with height bigger than width, or the opposite. This is to be able to look for different kinds of shapes, as for instance a car and a pedestrian usually fit into different bounding box shapes. The aspects assures that default bounding boxes of different shapes are utilized.**

**3 d)**

**3 e)**

**Main difference(s) between YOLO and SSD.**

**YOLO uses a FCN layer to classify after the initial feature extraction, and uses a single scale feature map in the end to classify both class and bounding box, while SSD uses multiple layers of convolution afer the initial feature extraction, and classifies some bounding boxes with classes for each convolution layer. This results in many more predictions than in the YOLO, and also greater accuracy.**

**3 f).**

**We have one anchor location per input value.**

**This gives us 38x38 anchor locations, and 6 anchors boxes per location, meaning 8664 anchor boxes.**

**3 g)**

**Same as last task, the sum of the input size times number of boxes created per location.**

**(38\*38+19\*19+10\*10+5\*5+3\*3+1\*1)\*6 = 11640 anchor boxes.**