1)what do REGION PROPOSALS entail.?

Ans : A Region Proposal Network or RPN, is a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. The RPN is trained end-to-end to generate high-quality region proposals.Region proposal algorithms seek to replace the traditional image pyramid and sliding window approach. These algorithms: Accept an input image. Over-segment it by applying a superpixel clustering algorithm.

The output of a region proposal network (RPN) is a bunch of boxes/proposals that will be passed to a classifier and regressor to eventually check the occurrence of objects. RPN predicts the possibility of an anchor being background or foreground, and refine the anchor.

2)what do you mean by NON-MAXIMUM SUPPRESSION (NMS)?

Ans : Non Maximum Suppression (NMS) is a technique used in numerous computer vision tasks. It is a class of algorithms to select one entity eg bounding boxes out of many overlapping entities. We can choose the selection criteria to arrive at the desired results.Non max suppression is a technique used mainly in object detection that aims at selecting the best bounding box out of a set of overlapping boxes

3)what exactly is MAP ?

Ans : MAP is a mean average precision.

The MAP compares the ground-truth bounding box to the detected box and returns a score. The higher the score, the more accurate the model is in its detections.& MAP is a popular evaluation metric used for object detection i.e. localisation and classification. Localization determines the location of an instance e.g. bounding box coordinates and classification tells you what it is e.g. a dog or cat. Image classification and localization.

4) what is a frames per second(FPS)?

Ans : frames per second is a

defines how fast your object detection model process your video and generates the desired output. The first step for any custom object detection is to grab images for labeling.& Frames per second (FPS) is a unit that measures display device performance in video captures and playback and video games. FPS is used to measure frame rate -- the number of images consecutively displayed each second -- and is a common metric used in video capture and playback when discussing video quality.

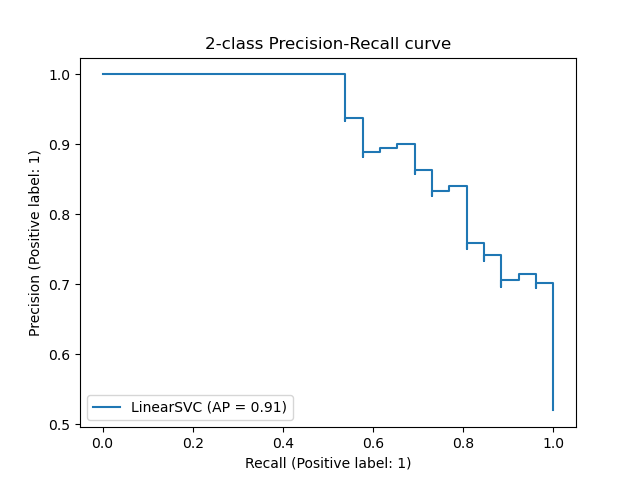
5)what is an IOU( intersection over union)?

Ans : Intersection over Union a value used in object detection to measure the overlap of a predicted versus actual bounding box for an object. The closer the predicted bounding box values are to the actual bounding box values the greater the intersection & the greater the IoU value.& its also detect an evaluation metric used to measure the accuracy of an object detector on a particular dataset.

6)Describe the Precision Recall Curve.

Ans :

The precision-recall curve shows the tradeoff between precision and recall for different threshold. A high area under the curve represents both high recall and high precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate.a PR curve is simply a graph with Precision values on the y-axis and Recall values on the x-axis. In other words, the PR curve contains TP/(TP+FN) on the y-axis and TP/(TP+FP) on the x-axis. It is important to note that Precision is also called the Positive Predictive Value (PPV).



7)what is term "selective search"?

Ans : Selective Search is a region proposal algorithm used in object detection. It is designed to be fast with a very high recall. It is based on computing hierarchical grouping of similar regions based on color, texture, size and shape compatibility.

8)Describe the R-CNN models four components.

Ans : Components of a Convolutional Neural Network. Convolutional networks are composed of an input layer, an output layer, and one or more hidden layers. A convolutional network is different than a regular neural network in that the neurons in its layers are arranged in three dimensions (width, height, and depth dimensions) The different layers of a CNN. There are four types of layers for a convolutional neural network: the convolutional layer, the pooling layer, the ReLU correction layer and the fully-connected layer.CNN is a type of deep learning model for processing data that has a grid pattern, such as images, which is inspired by the organization of animal visual cortex [13, 14] and designed to automatically and adaptively learn spatial hierarchies of features, from low- to high-level patterns.

9)what exactly is the localization module ?

Ans : Image localization is a spin-off of regular CNN vision algorithms. These algorithms predict classes with discrete numbers. In object localization, the algorithm predicts a set of 4 continuous numbers, namely, x coordinate, y coordinate, height, and width, to draw a bounding box around an object of interest.

10)what are the R-CNN disadvantages ?

Ans : Training is a multi-stage pipeline -End-to-end joint training.

Training is expensive in space and time. - Convolutional layer sharing. Classification in memory.

Test-time detection is slow - Single scale testing, SVD fc layer.