1. Describe the Quick R-CNN architecture.

**Ans : The Fast R-CNN consists of a CNN (usually pre-trained on the ImageNet classification task) with its final pooling layer replaced by an “ROI pooling” layer and its final FC layer is replaced by two branches — a (K + 1) category softmax layer branch and a category-specific bounding box regression branch.**

2. Describe two Fast R-CNN loss functions.

**Ans : u-True class label, u∈0,1,…,K; by convention, the catch-all background class has u=0.**

**v-True bounding box v=(vx,vy,vw,vh).**

3. Describe the DISABILITIES OF FAST R-CNN

**Ans : One drawback of Faster R-CNN is that the RPN is trained where all anchors in the mini-batch, of size 256, are extracted from a single image. Because all samples from a single image may be correlated (i.e. their features are similar), the network may take a lot of time until reaching convergence.**

4. Describe how the area proposal network works.

**Ans : 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. In nutshell , RPN predicts the possibility of an anchor being background or foreground, and refine the anchor.**

5. Describe how the RoI pooling layer works.

**Ans : ROI pooling solves the problem of fixed image size requirement for object detection network. RO I pooling produces the fixed-size feature maps from non-uniform inputs by doing max-pooling on the inputs. The number of output channels is equal to the number of input channels for this layer.**

6. What are fully convolutional networks and how do they work? (FCNs)

**Ans : A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.**

7. What are anchor boxes and how do you use them?

**Ans : Anchor boxes are a set of predefined bounding boxes of a certain height and width. These boxes are defined to capture the scale and aspect ratio of specific object classes you want to detect and are typically chosen based on object sizes in your training datasets.**

8. Describe the Single-shot Detector's architecture (SSD)

**Ans : SSD is a single-shot detector. It has no delegated region proposal network and predicts the boundary boxes and the classes directly from feature maps in one single pass. To improve accuracy, SSD introduces: small convolutional filters to predict object classes and offsets to default boundary boxes.**

9. HOW DOES THE SSD NETWORK PREDICT?

**Ans : SSD uses a matching phase while training, to match the appropriate anchor box with the bounding boxes of each ground truth object within an image. This property is used for training the network and for predicting the detected objects and their locations once the network has been trained.**

10. Explain Multi Scale Detections?

**Ans : A unified deep neural network, denoted the multi-scale CNN (MS-CNN), is proposed for fast multi-scale object detection. In the proposal sub-network, detection is performed at multiple output layers, so that receptive fields match objects of different scales.**

11. What are dilated (or atrous) convolutions?

**Ans : Dilated convolutions or atrous convolutions, previously described for wavelet analysis without signal decimation, expands window size without increasing the number of weights by inserting zero-values into convolution kernels.**