1. Describe the Quick R-CNN architecture.

Ans. Faster R-CNN is a single-stage model that is trained end-to-end. It uses a novel region proposal network (RPN) for generating region proposals, which save time compared to traditional algorithms like Selective Search. It uses the ROI Pooling layer to extract a fixed-length feature vector from each region proposal.

2. Describe two Fast R-CNN loss functions.

Ans. RPN Loss Function - The first term is the classification loss over 2 classes (There is object or not). The second term is the regression loss of bounding boxes only when there is object (i.e. p_i* =1). Thus, RPN network is to pre-check which location contains object.

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. 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.

5. Describe how the RoI pooling layer works.

Ans. ROI pooling solves the problem of fixed image size requirement for object detection network. ROI 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. Fully convolutional (deep neural) networks, or FCNs, are commonly used for computer vision tasks, such as semantic segmentation, super-resolution, etc. One of their best properties is that they are applicable to inputs of any size, for example images of different sizes.

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

Ans. An object detector that uses anchor boxes can process an entire image at once, making real-time object detection systems possible. Because a convolutional neural network (CNN) can process an input image in a convolutional manner, a spatial location in the input can be related to a spatial location in the output.

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

Ans. 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. Essentially, the anchor box with the highest degree of overlap with an object is responsible for predicting that object's class and its location.

10. Explain Multi Scale Detections?

Ans. Multiscale edge detectors typically smooth the signal at various scales and determine the sharp variation of points from their first- or second-order derivatives.

11. What are dilated (or atrous) convolutions?

Ans. Dilated convolutions, also known as atrous convolutions, have been widely explored in deep convolutional neural networks (DCNNs) for various tasks like semantic image segmentation, object detection, audio generation, video modeling, and machine translation.