1. Describe the Quick R-CNN architecture.

2. Describe two Fast R-CNN loss functions.

3. Describe the DISABILITIES OF FAST R-CNN

4. Describe how the area proposal network works.

5. Describe how the RoI pooling layer works.

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

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

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

9. HOW DOES THE SSD NETWORK PREDICT?

10. Explain Multi Scale Detections?

11. What are dilated (or atrous) convolutions?

Answer:

1. The Quick R-CNN architecture is an improvement over the previous R-CNN models that uses a single-stage network for object detection. It includes a single network that processes the whole image and predicts object classes and bounding box coordinates. The network first processes the input image through a convolutional neural network (CNN) to obtain a convolutional feature map. Then, the RoI pooling layer is applied to each region proposal to extract a fixed-length feature vector. These feature vectors are then fed into a sequence of fully connected (fc) layers for classification and regression to predict the object class and bounding box coordinates, respectively.
2. The two loss functions used in Fast R-CNN are:

* Classification loss: Measures the difference between predicted class probabilities and the true class labels. This is computed using the softmax function and cross-entropy loss.
* Localization loss: Measures the difference between the predicted bounding box coordinates and the ground-truth bounding box coordinates. This is computed using the smooth L1 loss.

1. The main disadvantages of Fast R-CNN are:

* It still requires region proposals, which can be slow.
* The RoI pooling layer is not differentiable, making end-to-end training difficult.
* It still uses separate models for classification and object detection, which can lead to suboptimal performance.

1. The region proposal network (RPN) is a network that is used in the Faster R-CNN architecture to generate region proposals directly from the input image. It works by sliding a small network over the convolutional feature map of the input image and predicting region proposals at each sliding window position. The RPN predicts two scores for each proposal: a score for how likely the proposal contains an object and a score for how accurate the proposal is. The proposals with high scores are then passed on to the next stage for further processing.
2. The RoI pooling layer is a layer that is used in object detection networks to extract fixed-length features from region proposals of varying sizes. It works by dividing the region proposal into a fixed number of sub-windows and then max-pooling the features within each sub-window to obtain a fixed-length feature vector. This allows the network to process region proposals of different sizes and aspect ratios without the need for resizing or cropping.
3. Fully convolutional networks (FCNs) are a type of neural network that is used for dense pixel-wise prediction tasks such as semantic segmentation. Unlike traditional CNNs, which use fully connected layers for classification, FCNs use only convolutional and pooling layers. This allows the network to take input of arbitrary size and output a dense prediction map of the same size.
4. Anchor boxes are pre-defined bounding boxes of different sizes and aspect ratios that are used in object detection networks such as SSD and YOLO. The network predicts the offset and confidence score for each anchor box, which is then used to generate the final bounding box prediction. By using multiple anchor boxes of different sizes and aspect ratios, the network is able to detect objects of different sizes and shapes.
5. The Single-shot Detector (SSD) architecture is a one-stage object detection network that uses a single network to predict object classes and bounding boxes. It works by using a series of convolutional layers with decreasing spatial resolution to detect objects at different scales. The network also predicts the offset and size of anchor boxes of different aspect ratios and scales. The final object detection is obtained by combining the predictions from multiple layers of the network.
6. The SSD network predicts object classes and bounding boxes by applying a series of convolutional layers to the input image. The network also predicts the offset and size of anchor boxes of different aspect ratios and scales. The final object detection is obtained by combining the predictions from multiple layers of the network.
7. Multi Scale Detections is a technique used in object detection to detect objects at different scales within an image. In this approach, the same detection network is applied to feature maps generated from images at different scales. This allows the network to detect objects of different sizes within the same image.
8. Dilated (or atrous) convolutions are a type of convolutional layer that can increase the receptive field of a convolutional neural network without adding additional parameters or computational cost. Instead of using a standard convolutional kernel that scans each pixel of the input, dilated convolutions use a kernel with gaps between the pixels. This increases the receptive field of the layer, allowing it to process a larger area of the input image with the same number of parameters and computational cost. Dilated convolutions are often used in semantic segmentation tasks, where the network needs to process the entire image to predict a segmentation mask.