**Assignment\_12**

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.**

1. Describe two Fast R-CNN loss functions.
2. Describe the DISABILITIES OF FAST R-CNN
3. 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: The RoI Pooling layer is just a type of max-pooling, where the pool size is dependent on the input size. Doing this ensures that the output is always of the same size. This layer is used because the fully-connected layer always expects the same input size, but input regions to the FC layer may have different sizes.**

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

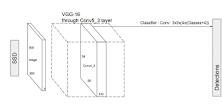
**Ans: Fully Convolutional Networks, or FCNs, are an architecture used mainly for semantic segmentation. They employ solely locally connected layers, such as convolution, pooling and upsampling. Avoiding the use of dense layers means less parameters (making the networks faster to train).**

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. 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: If there are multiple object to localize on an image, we use multiple object detection. As like the object localization, neural network creates 7 output vectors, but grid by grid. One image is divided by a grid 4 by 4 or 16 by 16 etc.. This example has 4 x 4 grid.**

11. What are dilated (or atrous) convolutions?

**Ans: The idea of Dilated Convolution is come from the wavelet decomposition . It is also called “atrous convolution”, “algorithme à trous” and “hole algorithm”. Thus, any ideas from the past are still useful if we can turn them into the deep learning framework**.