1. What do REGION PROPOSALS entail?

A1. Region proposals are a technique used in object detection to generate potential bounding boxes around objects in an image. This helps to reduce the search space for the object detection algorithm by only considering a smaller set of regions that are likely to contain objects.

The region proposal algorithm typically involves generating a large set of candidate regions using techniques such as sliding windows, selective search, or edge boxes. Each region is evaluated using a scoring function that takes into account features such as size, shape, color, texture, and context. The regions with the highest scores are selected as proposals and passed to the object detection algorithm for further processing.

Region proposals have been widely used in state-of-the-art object detection algorithms such as Faster R-CNN and Mask R-CNN. By reducing the search space, these algorithms are able to achieve high accuracy while maintaining reasonable computational efficiency.

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

A2.   
Non-Maximum Suppression (NMS) is a technique used to eliminate redundant detections in object detection models. When an object detection model generates multiple bounding boxes for the same object, NMS is used to remove all but the most accurate bounding box.

The NMS algorithm works by first sorting the bounding boxes based on their confidence scores. The bounding box with the highest confidence score is considered the most accurate. The algorithm then iterates through the sorted list of bounding boxes and compares each bounding box with the ones that follow it. If the intersection over union (IoU) of two bounding boxes exceeds a certain threshold, typically 0.5, then the bounding box with the lower confidence score is discarded.

The NMS algorithm helps to reduce the number of redundant bounding boxes generated by object detection models, resulting in more accurate object detection and localization.

3. What exactly is mAP?

A3. mAP stands for mean Average Precision, which is a common performance metric used in object detection tasks. It is used to evaluate the accuracy of the model in predicting the location and class of objects in an image.

The average precision (AP) is calculated for each class of object in the predicted output. It is then averaged across all the classes to obtain the mAP. AP is calculated as the area under the precision-recall curve, where precision is the ratio of true positives to the total number of predicted positives and recall is the ratio of true positives to the total number of actual positives in the dataset. The precision-recall curve is generated by plotting precision against recall for different classification thresholds.

The mAP provides a single score that represents the overall accuracy of the model in detecting objects in an image. It is a widely used metric in object detection tasks, and higher mAP scores indicate better performance.

4. What is a frames per second (FPS)?

A4.   
Frames per second (FPS) is a measure of the number of frames or images displayed per second in a video or animation. It is a common metric used to describe the performance of computer vision models, particularly in applications such as object detection and tracking, where real-time performance is critical. Higher FPS indicates smoother and more fluid motion, while lower FPS may result in choppiness or lag.

5. What is an IOU (INTERSECTION OVER UNION)?

A5. Intersection over Union (IoU) is a metric used to evaluate the performance of an object detection algorithm. It measures the overlap between the predicted bounding box and the ground truth bounding box of an object.

The IoU is computed as the ratio of the intersection between the predicted and ground truth bounding boxes to the union of the two boxes. Mathematically, it can be expressed as:

IoU = (Area of Intersection)/(Area of Union)

A value of 1 indicates that the predicted and ground truth bounding boxes perfectly overlap, while a value of 0 indicates no overlap at all. In practice, a threshold value for the IoU is set to determine whether a prediction is considered a true positive or a false positive.

6. Describe the PRECISION-RECALL CURVE (PR CURVE)

A6. The precision-recall curve (PR curve) is a graphical representation of the performance of a binary classification model that predicts a probability score for each positive class. The PR curve is a plot of the precision and recall values as the threshold for the classification decision changes.

Precision is defined as the number of true positives divided by the number of true positives plus false positives, while recall (also known as sensitivity) is defined as the number of true positives divided by the number of true positives plus false negatives.

The PR curve plots the recall on the x-axis and the precision on the y-axis, with each point on the curve representing a different threshold value. The area under the PR curve (AUC-PR) is a common metric used to compare the performance of different classification models. A higher AUC-PR indicates better performance, as it means the model is able to achieve high precision at high recall values.

The PR curve is useful in situations where the distribution of positive and negative classes is imbalanced, as it focuses on the performance of the model on the positive class.

7. What is the term "selective search"?

A7. Selective Search is a popular region proposal algorithm used in object detection and image segmentation tasks. It is a bottom-up approach that identifies and groups together similar regions of an image based on their color, texture, and shape. The algorithm generates a large set of region proposals that are used as input to a machine learning model for further processing. Selective Search is known for its accuracy in detecting objects in complex scenes and is widely used in state-of-the-art object detection systems.

8. Describe the R-CNN model's four components.

A8. The R-CNN (Region-based Convolutional Neural Network) model consists of four main components:

1. **Selective search**: This component generates a large number of region proposals by analyzing the input image to identify regions that may contain objects. It uses a hierarchical grouping algorithm to group similar pixels into regions at different scales.
2. **CNN feature extractor**: Each region proposal is fed into a pre-trained CNN to extract a fixed-length feature vector that describes the content of the region. The pre-trained CNN may be a popular model such as VGG or ResNet that has been trained on a large dataset such as ImageNet.
3. **Region-specific CNN**: The feature vector is then fed into a region-specific CNN that is trained to classify the object in the proposed region. This CNN consists of one or more fully connected layers that map the fixed-length feature vector to a vector of class scores.
4. **Bounding box regression**: In addition to classifying the object in the proposed region, the model also predicts the coordinates of the bounding box that tightly encloses the object. This is done by training a regression model to predict the offset between the proposed bounding box and the true bounding box of the object.

Finally, the output of the R-CNN model is a set of object detections, each consisting of a bounding box and a predicted class label. The model is trained end-to-end by jointly optimizing the classification and regression losses over a large dataset of labeled images.

9. What exactly is the Localization Module?

A9. The Localization Module is a component of the R-CNN object detection model that predicts the object's position in the image. The Localization Module takes an object proposal generated by the Region Proposal Network (RPN) and predicts the coordinates of the object's bounding box. This is done using a regression algorithm that takes features from the feature maps produced by the CNN backbone and outputs four numbers that represent the predicted x and y coordinates of the box's center, as well as its width and height.

The Localization Module is trained using a loss function that penalizes incorrect predictions of the bounding box coordinates. During training, the ground truth bounding box coordinates are compared to the predicted coordinates, and the loss is backpropagated through the network to update the weights. By predicting the object's location more accurately, the Localization Module helps to improve the overall object detection accuracy of the R-CNN model.

10. What are the R-CNN DISADVANTAGES?

A10. The R-CNN model has several disadvantages, including:

1. Slow training and testing times: R-CNN has a slow training and testing time compared to other object detection models. This is because it requires training a separate SVM for each object category, which is a time-consuming process.
2. High memory usage: R-CNN uses a lot of memory, which can make it difficult to train and test on low-end hardware.
3. Inability to share features: R-CNN cannot share features across object proposals, which can lead to redundant computation and slow performance.
4. Difficulty in real-time object detection: R-CNN is not well-suited for real-time object detection because of its slow training and testing times.
5. Complex architecture: R-CNN has a complex architecture that can be difficult to understand and implement, especially for those who are new to deep learning.
6. Sensitivity to object proposals: R-CNN is highly sensitive to the quality of the object proposals generated by selective search, which can lead to missed detections or false positives.