|  |  |
| --- | --- |
| 1 | Describe the Quick R-CNN architecture. |
| Ans | Quick R-CNN is a variation of the Faster R-CNN architecture for object detection. It is an improved version of the R-CNN architecture which uses a single network to predict both the object's location and its class. Quick R-CNN uses a region proposal network (RPN) to generate candidate bounding boxes of potential objects in an image. These bounding boxes are then fed into a separate network to classify the objects and refine their locations. The RPN works by scanning a predefined set of anchor boxes across the image, and predicting a score for each one indicating how likely it is to contain an object of interest. The network then outputs a set of bounding boxes which are used as input to the second network. This second network is a convolutional neural network (CNN) which is responsible for predicting the class label of the object inside the bounding box, as well as refining the location of the box. The Quick R-CNN architecture is faster than the original R-CNN architecture due to the fact that it uses a single network to both generate and classify the bounding boxes. |
| 2 | Describe two Fast R-CNN loss functions. |
| Ans | The Fast R-CNN loss function consists of two parts: the localization loss and the classification loss.  **Localization loss:** This loss function is used to measure how well the model localizes the objects in the image, and is usually a form of a regression loss such as Smooth L1 Loss or a variant of the Huber Loss.  **Classification loss:** This loss function is used to measure how well the model can classify the objects in the image, and is usually a form of a cross-entropy loss such as softmax or sigmoid cross-entropy.  **Regression Loss:** This loss is used to train the model for bounding box regression. It is a smooth L1 loss between the predicted bounding box coordinates and the ground truth coordinates. |
| 3 | Describe the DISABILITIES OF FAST R-CNN |
| Ans | * Slow inference speed: Fast R-CNN is relatively slower compared to other detectors like YOLO and SSD, due to its two-stage approach. * Difficulty with overlapping objects: Fast R-CNN is not very good at handling overlapping objects in an image, which can lead to incorrect bounding box predictions. * Sensitive to hyperparameters: Fast R-CNN is very sensitive to the selection of hyperparameters, and if the wrong set of hyperparameters is used, it can lead to sub-optimal performance. * Not suitable for real-time applications: Fast R-CNN is not suitable for real-time applications due to its slower inference speed. |
| 4 | Describe how the area proposal network works. |
| Ans | The Area Proposal Network (APN) is a convolutional neural network architecture used to detect objects in an image. The APN consists of two major parts: the proposal network and the region classifier. The proposal network is used to generate the object proposals, which are regions of an image that may contain objects. These proposals are then passed to the region classifier, which is used to classify the object in each proposal. The region classifier uses a fully-connected layer to output the object class and a bounding box for the object. The APN is trained using a supervised learning approach which uses labeled images to learn the object classes and their bounding boxes. |
| 5 | Describe how the RoI pooling layer works. |
| Ans | The RoI pooling layer is a type of layer that is used in a convolutional neural network to reduce the spatial dimensions of an input. The RoI pooling layer works by dividing the input into regions of interest (RoIs) and then applying a max pooling operation across each RoI. This allows the network to focus on the most important features in each region while reducing the computational complexity of the overall network. The RoI pooling layer is especially useful for tasks such as object detection and semantic segmentation. |
| 6 | What are fully convolutional networks and how do they work? (FCNs) |
| Ans | Fully convolutional networks (FCNs) are a type of convolutional neural network (CNN) that are used for image segmentation, object detection, and image recognition. They are based on a series of convolutional layers that are connected with upsampling layers to enable the network to learn higher-level features of the input data. The convolutional layers extract features from the input data and the upsampling layers are used to increase the resolution of the input data. The upsampling layers are also used to merge features from different layers of the network. The end result is an output that is a segmentation of the input data into the desired classes. FCNs can be used to detect and segment objects in images and videos, as well as identify and classify objects in images and videos. |
| 7 | What are anchor boxes and how do you use them? |
| Ans | Anchor boxes are a type of bounding box used in object detection algorithms. They are fixed boxes of various sizes and aspect ratios placed over an image. The goal is to create anchor boxes that can be used to detect different classes of objects in a given image. Anchor boxes are used in conjunction with a sliding window approach, where an algorithm moves the anchor boxes over the image and makes a prediction at each location. If the predictions meet a certain criteria, the bounding box is accepted and used to detect the target object. |
| 8 | Describe the Single-shot Detectors architecture (SSD) |
| Ans | Single-shot Detectors (SSD) is a type of object detection architecture based on a deep neural network. It is a one-stage object detection architecture that uses a single deep neural network to predict object locations and class probabilities simultaneously. SSD uses anchor boxes of different aspect ratios and scales to detect objects, and applies non-maximum suppression (NMS) to remove multiple predicted boxes for the same object. Unlike two-stage detectors, SSD does not require a region proposal network (RPN) and instead directly predicts object locations and class probabilities for each anchor box. SSD is fast and accurate, making it suitable for real-time applications. |
| 9 | HOW DOES THE SSD NETWORK PREDICT? |
| Ans | The SSD Network uses deep learning to predict the most likely outcome of a given situation. Specifically, it utilizes convolutional neural networks to analyze the data and identify patterns. These patterns are then used to make predictions about the future. |
| 10 | Explain Multi Scale Detections? |
| Ans | Multi-scale detection is a computer vision technique used to detect objects in an image or video at multiple scales. This technique is often used in object detection and recognition systems to improve accuracy and reduce false positives. Multi-scale detections are useful for detecting objects at different scales, such as small objects in the foreground and large objects in the background. This technique can also be used for tracking objects over time, as the object's size may change. Multi-scale detection can also be used to identify objects that are occluded or partially visible. |
| 11 | What are dilated (or atrous) convolutions? |
| Ans | Dilated (or atrous) convolutions are convolutional kernels that have adjustable gaps between the kernel’s elements. This allows the kernel to cover a wider range of input values and makes the network more robust. The gaps between the kernel’s elements can be increased or decreased to control the receptive field size. Dilated convolutions have been shown to improve accuracy and efficiency of deep learning networks. |