

Object Detection model like YOLOV4, Faster RCNN, SSD has 3 phases:

1. Feature Extraction/Backbone model:

- Values are extracted by concept of CNN based backbone networks (convolution + pooling). Here we have used models like ResNet (for Faster RCNN, SSD), Darknet (for YOLOV4).
- The value like 50, 101 signifies the depth i.e. layers used in the model.

2. Feature Recombination/Neck/aggregation model:

- Extracted features are recombined. Here it is done via Feature Pyramid Network (FPN) for Faster RCNN, SSD, PANet in case of YOLOV4.

3. Prediction:

- Predicting the bounding boxes and then classifying the vehicle.

Note:

- Faster RCNN was used due to better accuracy out of RCNN, Fast RCNN, Faster RCNN.
- Single Shot detector (SSD), YOLO, Faster R-CNN takes only one shot to detect (SSD concept) multiple objects present in an image using multiple bounding boxes. It is faster (less computational time). It divides the image into grid and grids are used to make the bounding / anchor box.
- Multiple Shot detector like R-CNN takes sliding window multiple times on the image. It is slower than SSD.
- Here we have used the complete model like YOLOV4 and used it (Transfer Learning).
- Loss function varies according to the algorithm and thus not chosen as a comparator.