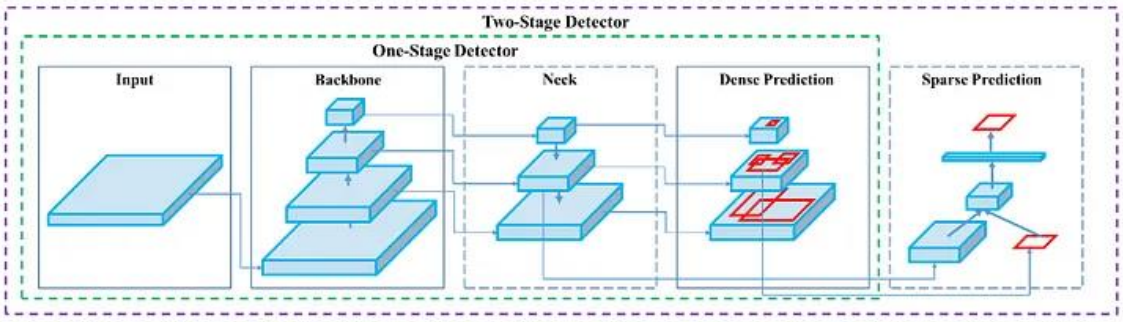
Presentation

Architecture:

CNN-based Object Detectors are primarily applicable for recommendation systems. YOLO (**Y**ou **O**nly **L**ook **O**nce) models are used for Object detection with high performance. YOLO divides an image into a grid system, and each grid detects objects within itself. They can be used for real-time object detection based on the data streams. They require very few computational resources.

To understand how Yolov5 improved the performance and its architecture, let us go through the following high-level Object detection architecture:



Input: { Image, Patches, Image Pyramid, ... }

Backbone: { VGG16 [68], ResNet-50 [26], ResNeXt-101 [86], Darknet53 [63], ... }

Neck: { FPN [44], PANet [49], Bi-FPN [77], ... }

Head:

Dense Prediction: { RPN [64], YOLO [61, 62, 63], SSD [50], RetinaNet [45], FCOS [78], ... }

Sparse Prediction: { Faster R-CNN [64], R-FCN [9], ... }

General Object Detector will have a **backbone** for pre-training it and a **head** to predict classes and bounding boxes. The ***Backbones*** can be running on GPU or CPU platforms. The ***Head*** can be either **one-stage** (e.g., YOLO, SSD, RetinaNet) for **Dense prediction** or **two-stage** (e.g., Faster R-CNN) for **the Sparse prediction** object detector. Recent Object detectors have some layers (***Neck***) to collect feature maps, and it is between the backbone and the Head.

- **Accurracy metrics:**

AP (Average precision) is a popular metric in measuring the accuracy of object detectors like Faster R-CNN, SSD, etc. Average precision computes the average precision value for recall value over 0 to 1

The mAP compares the ground-truth bounding box to the detected box and returns a score. The higher the score, the more accurate the model is in its detections.

IoU (Intersection over union)

IoU measures the overlap between 2 boundaries. We use that to measure how much our predicted boundary overlaps with the ground truth (the real object boundary). In some datasets, we predefine an IoU threshold (say 0.5) in classifying whether the prediction is a true positive or a false positive.

The IoU is calculated by dividing the area of intersection between the 2 boxes by the area of their union. The higher the IoU, the better the prediction.

Usually, the object detection models are evaluated with different IoU thresholds where each threshold may give different predictions from the other thresholds.