Assignment3C

YOLO

(You Only Look Once)

YOLO (You Only Look Once) is a state-of-the-art object detection algorithm that uses a neural network to detect objects in an image. The YOLO architecture is designed to be fast and accurate, allowing it to process images in real-time.
The YOLO architecture consists of a series of convolutional layers followed by fully connected layers. The convolutional layers are used to extract features from the input image, while the fully connected layers are used to make predictions about the objects in the image. YOLO uses a single convolutional network to predict both the class probabilities and the bounding boxes of the objects in the image.
YOLO divides an image into a grid and predicts bounding boxes and class probabilities for each grid cell.One of the unique features of YOLOv2 and its subsequent models is the use of anchor boxes. Anchor boxes are predefined boxes of different sizes and aspect ratios that are used to predict the bounding boxes of objects in the image. Instead of predicting the coordinates of the bounding box directly, YOLO predicts the offsets between the anchor box and the true bounding box. This allows YOLO to handle objects of different sizes and aspect ratios.

The anchor boxes are represented as vectors, with each vector containing the width and height of the box,the aspect ratio,probability of the object classes and confidence score.
An anchor box B can be represented as:
[x1,y1,x2,y2,w1,h1,.......pc1,pc2,pc3...]
The vectors are learned during training, allowing the model to adapt to the specific characteristics of the objects in the dataset.
The YOLO algorithm uses a unified detection approach, which means that it predicts the class probabilities and bounding boxes of all objects in the image simultaneously through one single network and training. This is in contrast to other object detection algorithms that use a sliding window approach or 2-step ,3-step training as in RCNNs which can be slower and less accurate.

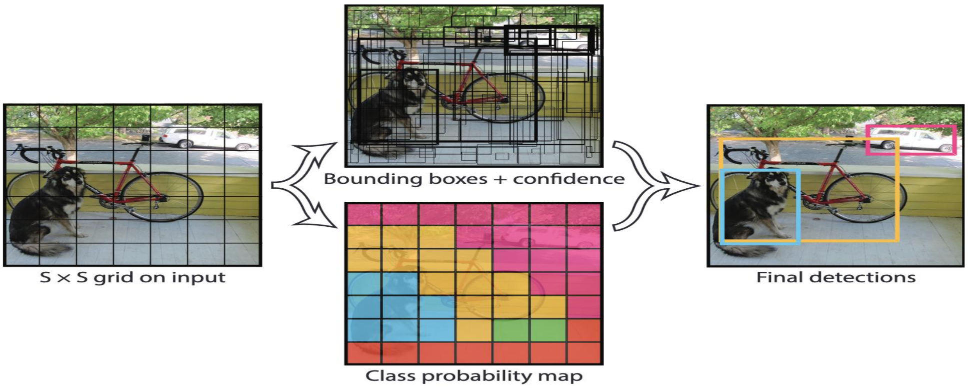

Fig: YOLO algorithm

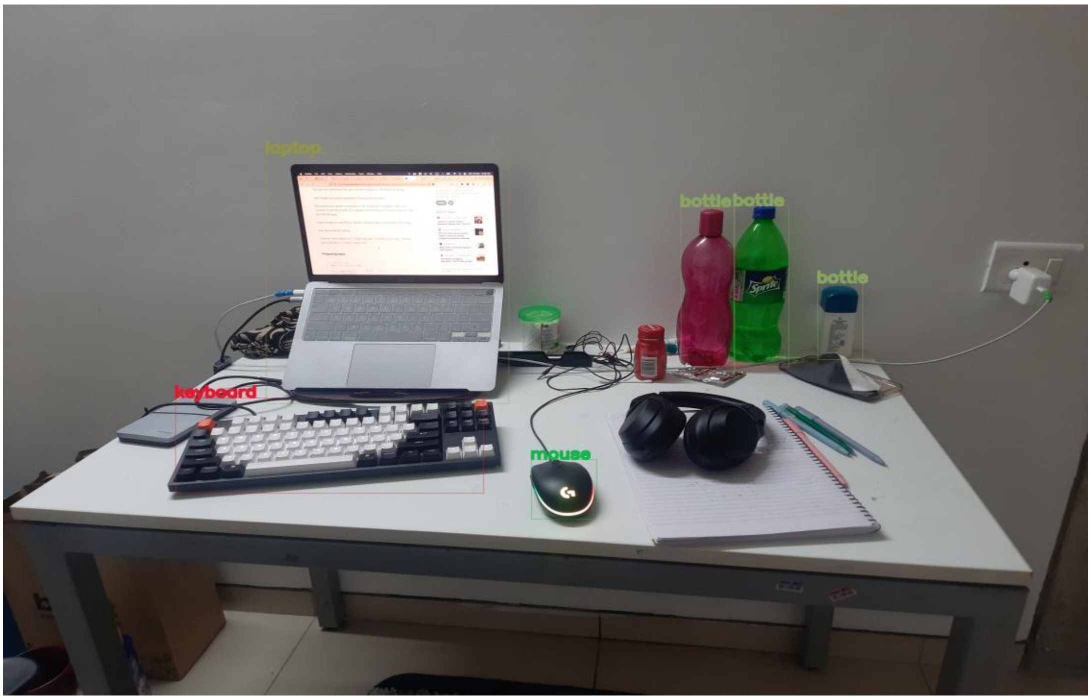

Fig: Object detection with YOLO

The algorithm has evolved over time and has been updated in different versions. Here are some of the differences between YOLO v1, and v2:

Architecture: The detection layer of YOLOv1 contains 24 convolutional layers and 2 fully connected layers. YOLOv2's architecture , the Darknet-19 has 19 convolutional layers and 5 max-pooling layers.YOLOv2 removes all fully connected layers and uses anchor boxes to predict bounding boxes. One pooling layer is removed to increase the resolution of output.

Batch Normalization: YOLOv1 only had convolutional layers in the feature extraction network, YOLOv2 introduced Batch normalization which improves the convergence of the model.

Accuracy: YOLO v1 was the fastest of the available detection models when it was first published but it lacked in accuracy a lot. It suffered from localizing smaller objects and objects very close to one another. YOLO v2 the 2nd version of the model is faster and also more accurate than its predecessor.

Anchor boxes: YOLO v2 introduced the concept of anchor boxes, which are predefined bounding boxes of different sizes and aspect ratios that are used to detect objects of different shapes and sizes. YOLOv2 removed the fully connected layers of YOLO and replaced them with anchor boxes to predict bounding boxes.

Resolution: YOLOv2 is trained on 416x416 images as compared to 448x448 of YOLOv1. This decreased resolution and the addition of anchor boxes has increased the mAP(mean average precision) from 63.4 at 45fps to 76.8 at 67fps. YOLOv2 trained on 288x288 images gives mAP of 69 at 91fps. At 544x544 it gives 78.6 mAP at 40fps.

The most recent official release of YOLO is YOLOv7 which gives a good amount of accuracy over 286fps.

To summarize, we can say that although YOLO is a state-of-the-art model for object detection, it still suffers from accuracy challenges compared to RCNNs and other models. It also suffers form challenges of Occlusion and detection of small objects.But when it comes to speed, it is the fastest model available. That’s a trade-off of speed over accuracy that YOLO offers us and it is the most preferred technique in most of the object detection tasks especially in real-time scenarios.

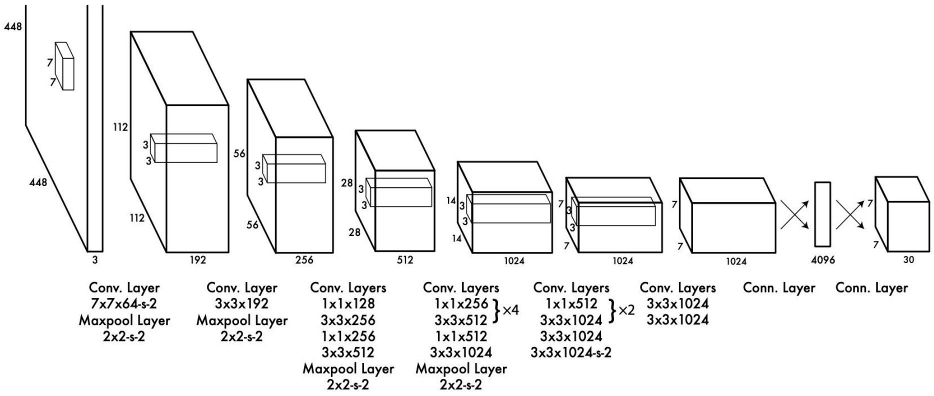

Fig: Architecture of YOLOv1

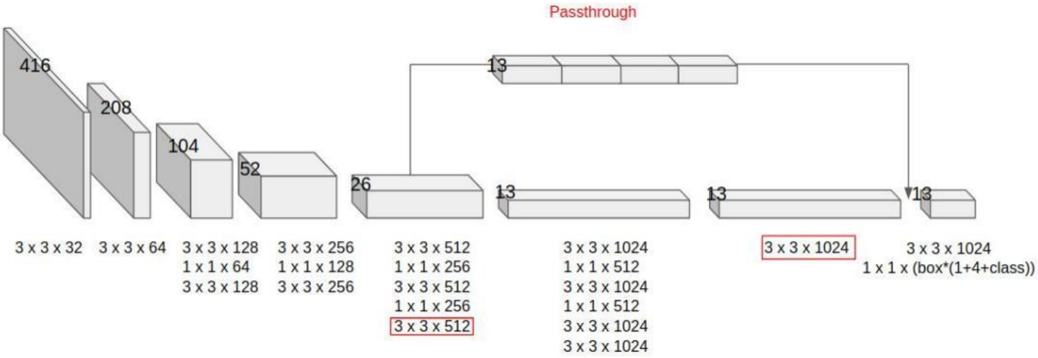

Fig: Architecture of YOLOv2