

YOLO (You Only Look Once) real-time object detection algorithm, which is one of the most effective object detection algorithms that also encompasses many of the most innovative ideas coming out of the computer vision research community. Object detection is a critical capability of autonomous vehicle technology. It is an area of computer vision that's exploding and working so much better than just a few years ago. At the end of this article, we will see a couple of recent updates to YOLO by the original researchers of this important technique. YOLO is a clever convolutional neural network (CNN) for doing object detection in real-time. The algorithm applies a single neural network to the full image, and then divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities. Fig.2 shows that the analysis of YOLO model.

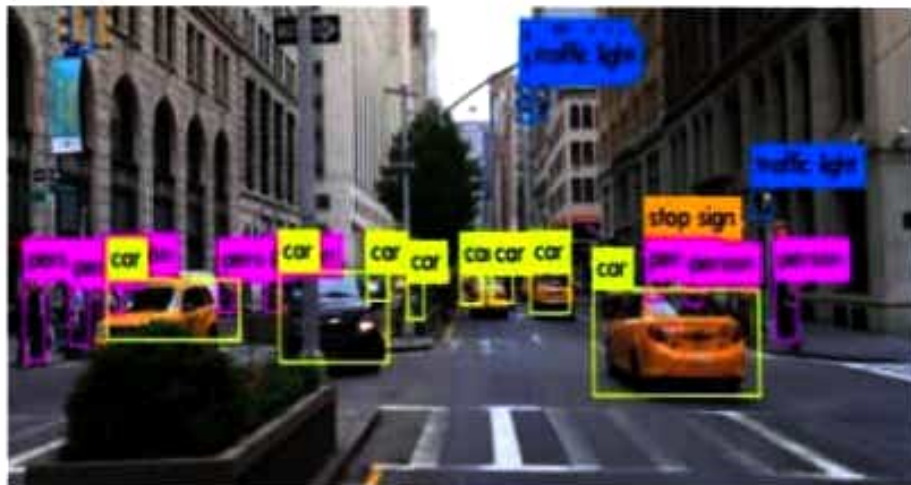


Fig.2 Analysis of YOLO Model

YOLO achieves high accuracy while also being able to run in real-time. The algorithm “only looks once” at the image in the sense that it requires only one forward propagation pass through the neural network to make predictions. After non-max suppression, it then outputs recognized objects together with the bounding boxes. With YOLO, a single CNN simultaneously predicts multiple bounding boxes and class probabilities for those boxes. YOLO trains on full images and directly optimizes detection performance. Fig.3 shows the object detection.



Fig.3 Object detection

Object detection

A. Convolution Neural Networks (CNN):

CNN is widely used neural network architecture for computer vision related tasks. Advantage of CNN is that it automatically performs feature extraction on images i.e. important features are detected by the network itself. CNN is made up of three important components called Convolutional Layer, Pooling layer, fully connected Layer. Considering a gray scale image of size 32×32 would have 1024 nodes in multi-layer approach. This process of flattening pixels loses spatial positions of the image.

B. Region-based Convolutional Neural Networks (R-CNN):

The Region-based Convolutional Network method (RCNN) is a combination of region proposals with Convolution Neural Networks (CNNs). R-CNN helps in localising objects with a deep network and training a high-capacity model with only a small quantity of annotated detection data. It achieves excellent object detection accuracy by using a deep ConvNet to classify object proposals. R-CNN has the capability to scale to thousands of object classes without resorting to approximate techniques, including hashing. The fig 4. shows that Regional based convolutional neural network.

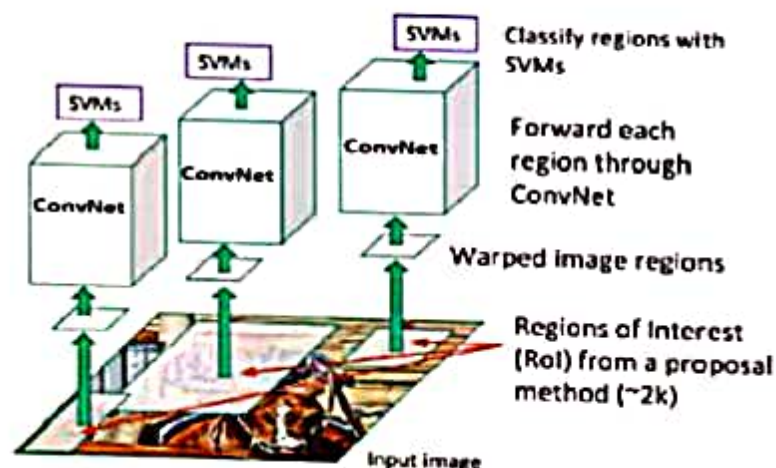


Fig.4 Regional based CNN

C. Single Shot MultiBox Detector (SSD):

Single Shot Detector (SSD) is a method for detecting objects in images using a single deep neural network. The Single Shot Detector network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes. Fig 5. shows Single Shot MultiBox Detector.

