

YOLO

Real Time Object Detection Algorithm

Presentation by
Mahesh Bhume

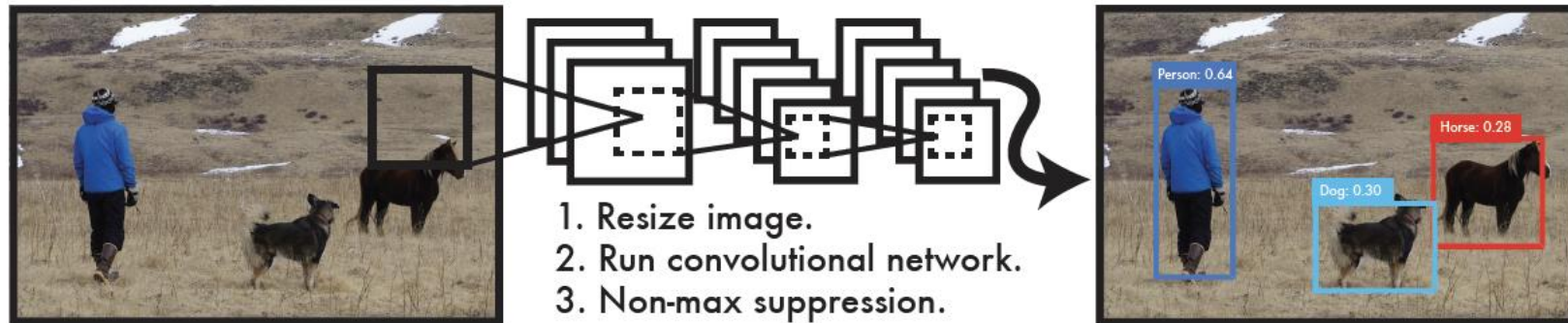
[Joseph Redmon](#), [Santosh Divvala](#), [Ross Girshick](#), [Ali Farhadi](#), “You Only Look Once: Unified, Real-Time Object Detection,” 2015 Cornell university on Computer Vision and Pattern Recognition, [arXiv:1506.02640](#)

Introduction

- Earlier the pipeline of traditional object detection models were not using deep learning due to lack of GPU and computational resources.
- As we humans can glance at an image and instantly know what objects are in the image, where they are, and how they interact. The human visual system is fast and accurate, allowing us to perform complex tasks like driving with little conscious thought. Fast, accurate algorithms for object detection would allow computers to drive cars without specialized sensors, enable assistive devices to convey real-time scene information to human users, and unlock the potential for general purpose, responsive robotic systems.
- Earlier detection systems used classifier to detect an object, these systems take a classifier for that object and evaluate it at various locations and scales in a test image. Systems like deformable parts models (DPM) use a sliding window approach where the classifier is run at evenly spaced locations over the entire image.
- More recent approaches like R-CNN use region proposal methods to first generate potential bounding boxes in an image and then run a classifier on these proposed boxes. After classification, post-processing is used to refine the bounding boxes, eliminate duplicate detections, and rescore the boxes based on other objects in the scene
- These complex pipelines are slow and hard to optimize because each individual component must be trained separately
- Whereas YOLO algorithm has reframe object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities. Using this model, you only look once (YOLO) at an image to predict what objects are present and where they are.
- YOLO is a one shot detectors, meaning that it only does one pass on the images to output all the detections. The obvious advantage with this method is the speed up in the computation and the increase in the number of frame being processed by second (45 FPS) and trade off is accuracy.

TYPES OF OBJECT DETECTION ALGORITHMS

- Object detection algorithms can be divided into two categories: classification-based algorithms and regression-based algorithms.
- **Classification based algorithms (sliding window/region proposal-based):** Some of the region proposal based techniques are SPP-NET , R-FCN , FPN , Mask R-CNN , R-CNN , Fast R-CNN and Faster R- CNN.
- There are two approaches for generating the region proposal
- 1)**Sliding window** algorithm: The regions are generated for each pixel location and then scaled it.
- 2)**Selective search** algorithm: would generate region proposals which overlap with each other and contain the object partially using image segmentation technique.
- These region proposal algorithms examine an input image and then identify where a potential object could be, this is implemented in three stages
- **Selection of region** that is of interest (ROI) in the image.
- **Features extraction** (Generation of feature vector): The features are extracted from each region proposal by leveraging the pre-trained Convolution Neural network.
- **Classification of region:** Each region is classified as positive region or negative region based on the existence or non-existence of the object.
- Examples would be : Haar features and histogram of oriented gradients
- **Regression-based algorithms:**
- Instead of selecting and singling out regions of interest in an image, they predict classes and their relevant bounding boxes for the whole image in one run.
- So advantage with this system is, the objects are located with the bounding boxes as well as their class is predicted in the same run.
- Complex pipeline is not necessary for regression-based algorithms.

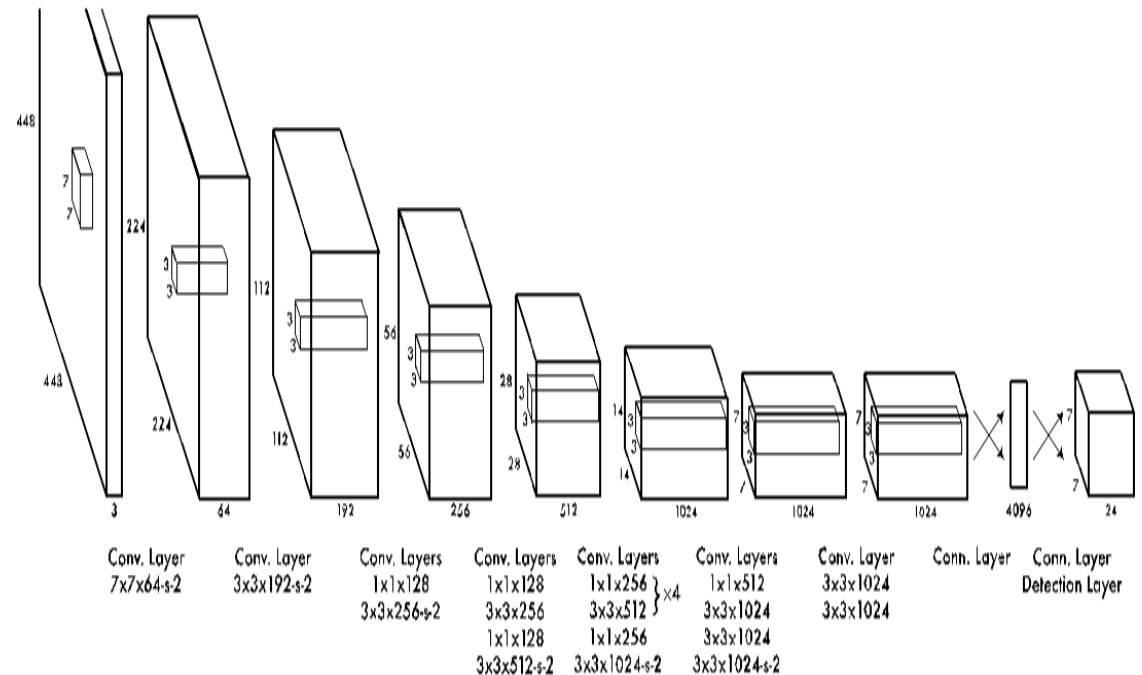


- It passes the image through the CNN algorithm only once to get the output, thus the name.
- YOLO aim to predict a class of an object and the bounding box specifying object location.

YOU ONLY LOOK ONCE (YOLO) ALGORITHM

Architecture of the YOLO algorithm

- The architecture of the network is quite simple, it is a series of convolutional layers followed by fully connected layers.
- The main idea is to have a grid of boxes to cover all the image being processed. The last layer contains all the boxes, coordinates and classes. This way you can cover the whole image with a pre-defined set of boxes.
- The diagram shows the layers of the network.

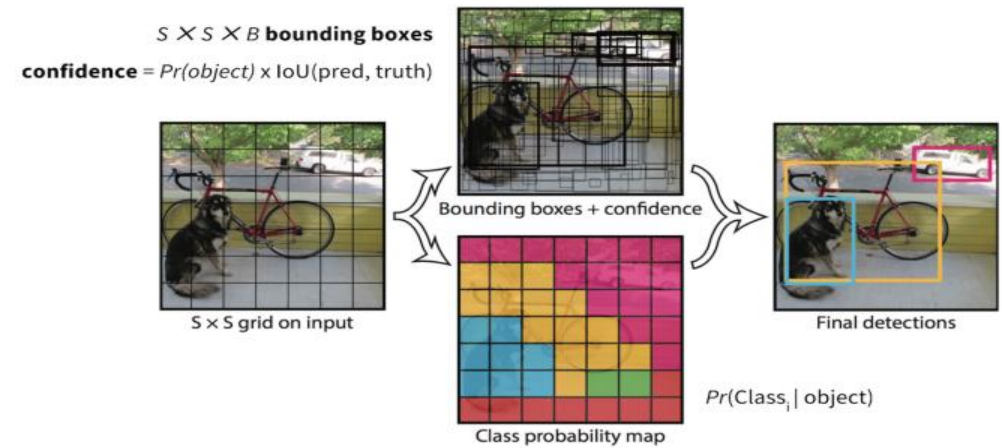


How Yolo algorithm works

- It works using three techniques:

1. Residual blocks: The image is divided into various grids, each grid has a dimension of $S \times S$.

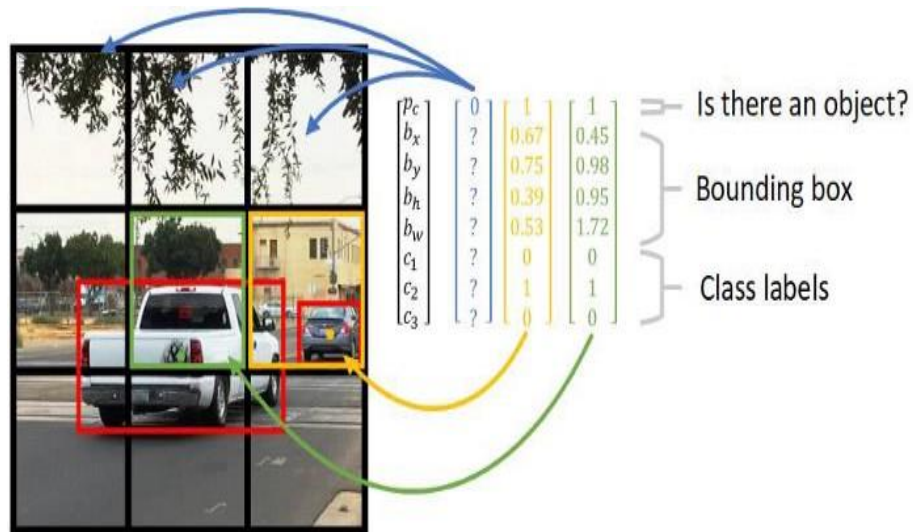
- The image shows how an input image is divided into grids and there are many grid cells of equal dimension. Every grid cell will detect objects that appear within them and if an object center appears within a certain grid cell, then this cell will be responsible for detecting it.



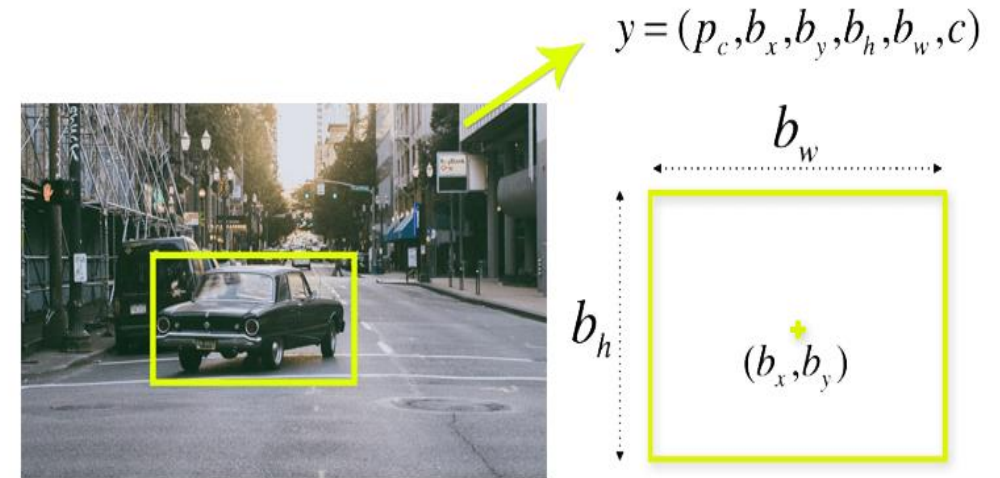
2. Bounding box regression

Each bounding box can be described using four descriptors:

- Centre of a bounding box (b_x, b_y)
- Width (b_w)
- Height (b_h)
- The value corresponding to a class of an object, it is represented by the letter c .

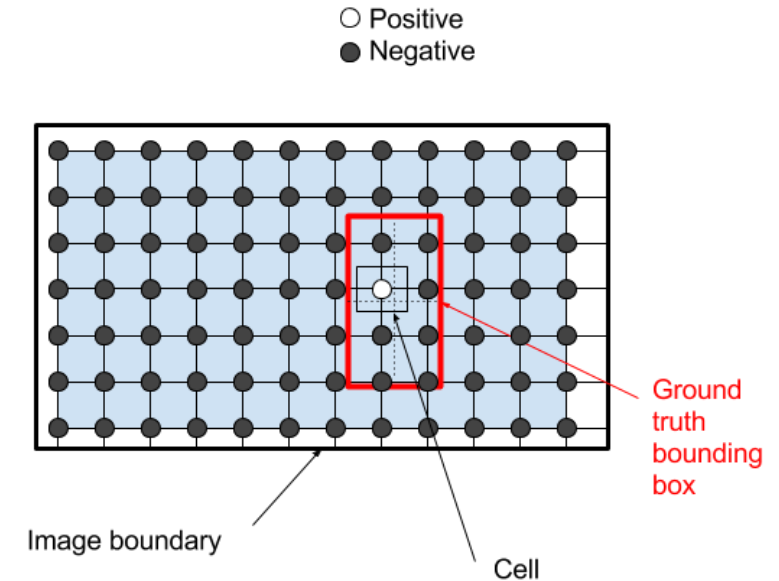


- In addition, we have to predict the p_c value, which is the probability that there is an object in the bounding box.
- The following image shows an example of a bounding box. The bounding box has been represented by a yellow outline.
- The image represents the probability of an object appearing in the bounding box.



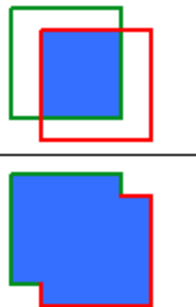
How it find center of the Bounding Box(BB)

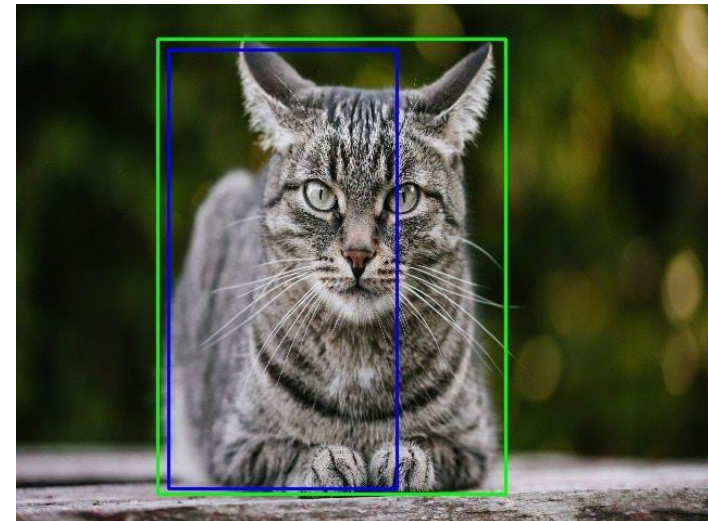
- It uses Positives and negatives cells
- A position on the grid, that is the closest position to the center of one of the ground truth bounding boxes, is positive
- It uses Anchor Box principle when there are multiple objects overlapping each other.



3. Intersection over union (IOU)

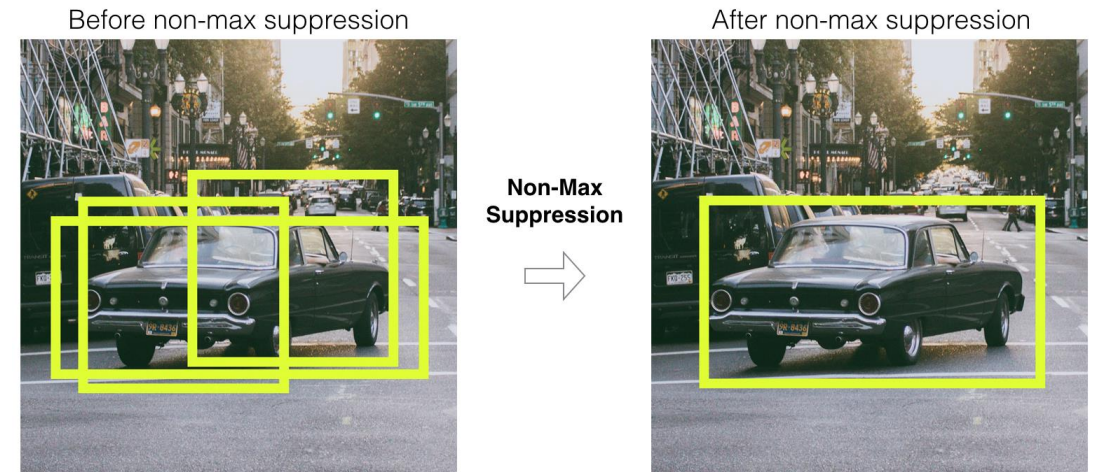
- IOU is a popular metric to measure localization accuracy and calculate localization errors in object detection models.
 - It describes how boxes overlap.
 - To calculate the IOU we use following figure
 - First we calculate the total area covered by the two bounding boxes—also known as the Union.
 - The intersection divided by the Union, gives us the ratio of the overlap to the total area, providing a good estimate of how close the bounding box is to the original prediction.
- YOLO uses IOU to provide an output box that surrounds the objects perfectly.
 - Each grid cell is responsible for predicting the bounding boxes and their confidence scores.
 - In the image, there are two bounding boxes, one in green and the other one in blue. The blue box is the predicted box while the green box is the real box. YOLO ensures that the two bounding boxes are equal.

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} = \frac{\text{Intersection}}{\text{Union}}$$




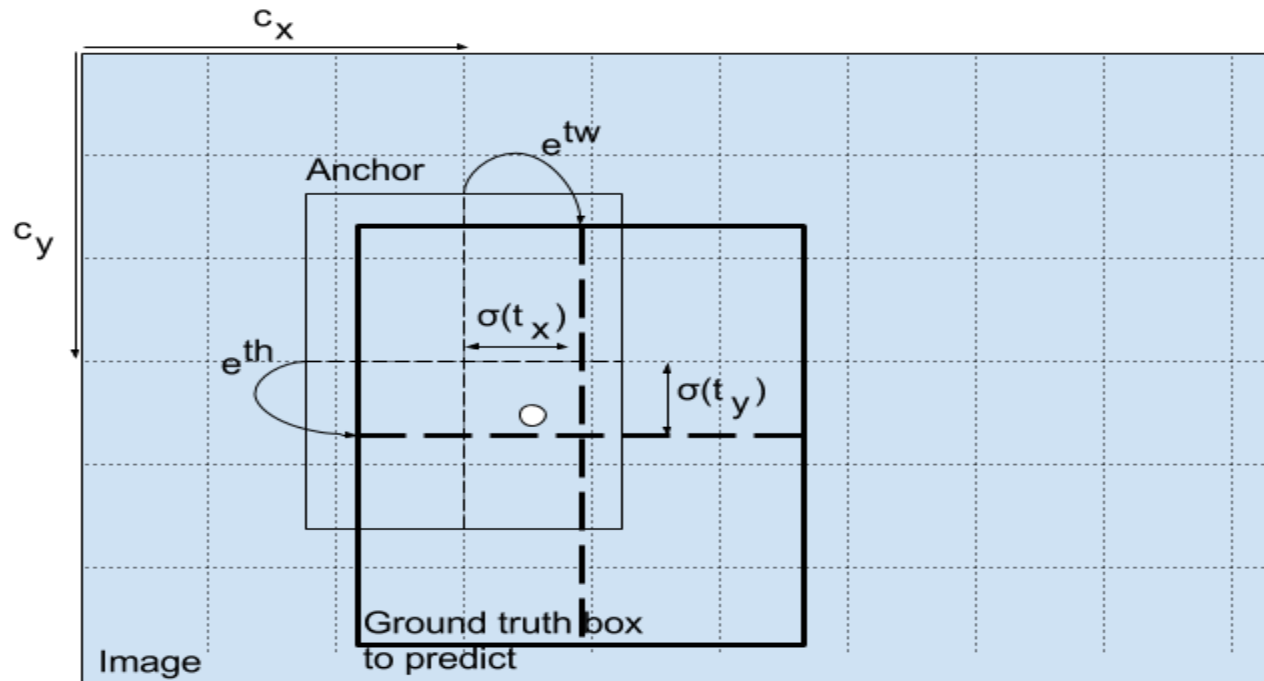
How YOLO deal with duplicate predictions

- YOLO makes use of Non Maximal Suppression to deal with this issue.
- In Non Maximal Suppression, YOLO suppresses all bounding boxes that have lower probability scores.
- YOLO achieves this by first looking at the probability scores associated with each decision and taking the largest one.
- This step is repeated till the final bounding boxes are obtained.
- This is achieved by value p_c which is Y , which serves to remove boxes with low object probability and bounding boxes with the highest shared area in a process called non-max suppression.

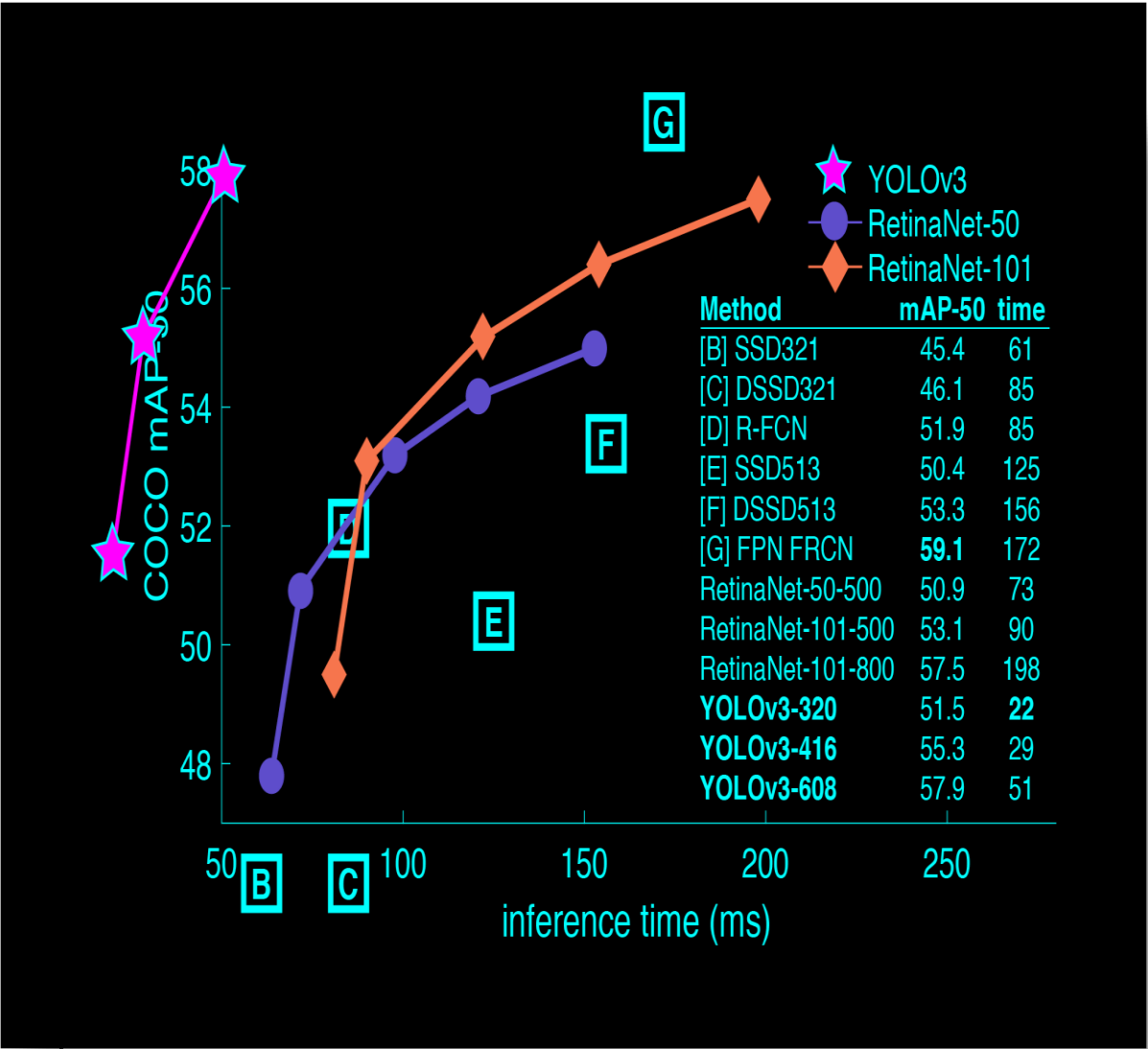


A regressor rather than a classifier

- For every positive position, the network predicts a **regression** on the bounding box precise position and dimension
- these predictions are relative to the grid position and anchor size (instead of the full image) as in the Faster-RCNN models for better performance.
- where (cx, cy) are the grid cell coordinates and (pw, ph) the anchor dimensions.



Data Base Used COCO



Comparison with other detectors

References:

- [1] [Joseph Redmon](#), [Santosh Divvala](#), [Ross Girshick](#), [Ali Farhadi](#), "You Only Look Once: Unified, Real-Time Object Detection," 2015 Cornell university on Computer Vision and Pattern Recognition, [arXiv:1506.02640](#)
- [2] Handalage, Upulie & Kuganandamurthy, Lakshini. (2021). Real-Time Object Detection Using YOLO: A Review. 10.13140/RG.2.2.24367.66723.
- [3] Rachita Byahatti , Dr. S. V. Viraktamath , Madhuri Yavagal, 2021, Object Detection and Classification using YOLOv3, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 02 (February 2021)
- [4] Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam, RealTime Object Detection with Yolo, International Journal of Engineering and Advanced Technology (JEAT) ISSN: 2249 – 8958, Volume-8, Issue-3S, February 2019
- [5] Liu, L., Ouyang, W., Wang, X. *et al.* Deep Learning for Generic Object Detection: A Survey. *Int J Comput Vis* 128, 261–318 (2020). <https://doi.org/10.1007/s11263-019-01247-4>
- [6] Zhanchao Huang, Jianlin Wang, "DC-SPP-YOLO: Dense Connection and Spatial Pyramid Pooling Based YOLO for Object Detection", College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
- [7] [Petr Hurtik](#), [Vojtech Molek](#), [Jan Hula](#), [Marek Vajgl](#), [Pavel Vlasanek](#), [Tomas Nejezchleba](#), "Poly-YOLO: higher speed, more precise detection and instance segmentation for YOLOv3", arXiv:2005.13243

Link to the past:

Reinhardt W. (1979) Figure-Ground Discrimination by the Visual System of the Fly. In: Haken H. (eds) Pattern Formation by Dynamic Systems and Pattern Recognition. Springer Series in Synergetics, vol 5. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-67480-8_10
Chapter : Representation and Processing of Associations Using Vector Space Operations Pages 199-207