KONKAN GYANPEETH COLLEGE OF ENGINEERING, KARJAT

Affiliated to University of Mumbai, Approved by AICTE, New Delhi.

Object Localization

Group members:

Deep Dama (07)
Durwankur Gursale (17)
Anuja Jadhav (22)

Contents:

Abstract:

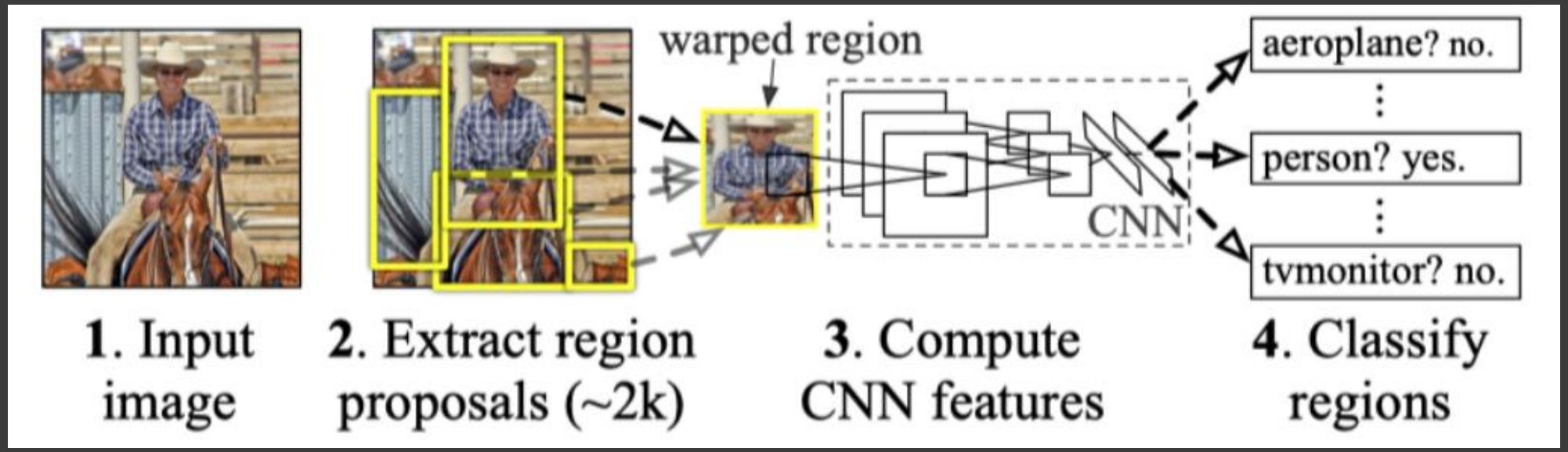
- *Object localization* refers to identifying the location of one or more objects in an image and drawing a bounding box around their extent. *Image classification* involves predicting the class of one object in an image. *Object detection* combines these two tasks and localizes and classifies one or more objects in an image. Object detection is one of the areas of computer vision that is maturing very rapidly. Today, there is a plethora of pre-trained models for object detection (YOLO, RCNN, Fast RCNN, Mask RCNN, Multi-box etc.). So, it only takes a small amount of effort to detect most of the objects in a video or in an image.

Introduction:

- In object localization you identify the object and locate where exactly it is present within the image. Simply put, object localization aims to locate the main object in an image.
- Location of the object is depicted using the bounding box. A bounding box (usually shortened to Bbox) is an area defined by two longitudes and two latitudes.

Literature Survey:

SR. NO.	PAPER TITLE	AUTHOR(PUB.YEAR)	DESCRIPTION
1.	Region-based Convolutional Networks for Accurate Object Detection and Segmentation	Ross Girshick, Jeff Donahue, Student Member, IEEE, Trevor Darrell, Member, IEEE, and Jitendra Malik, Fellow, IEEE	• R-CNN is a region based approach that led to a wave of research in object detection • It is a two-stage framework, i.e, region proposal stage, and region classification and refinement stage. • R-CNN extracts regions of interest from an input image by using a technique called selective search.


warped region
CNN
aeroplane? no.
person? yes.
tvmonitor? no.
1. Input image
2. Extract region proposals (~2k)
3. Compute CNN features
4. Classify regions

2.	Faster R-CNN	Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun	• Faster R-CNN is faster than previous versions of R-CNN. • It introduced a technique called anchor box. Anchor boxes are pre-defined prior boxes with different aspect ratios and sizes but share the same central location.

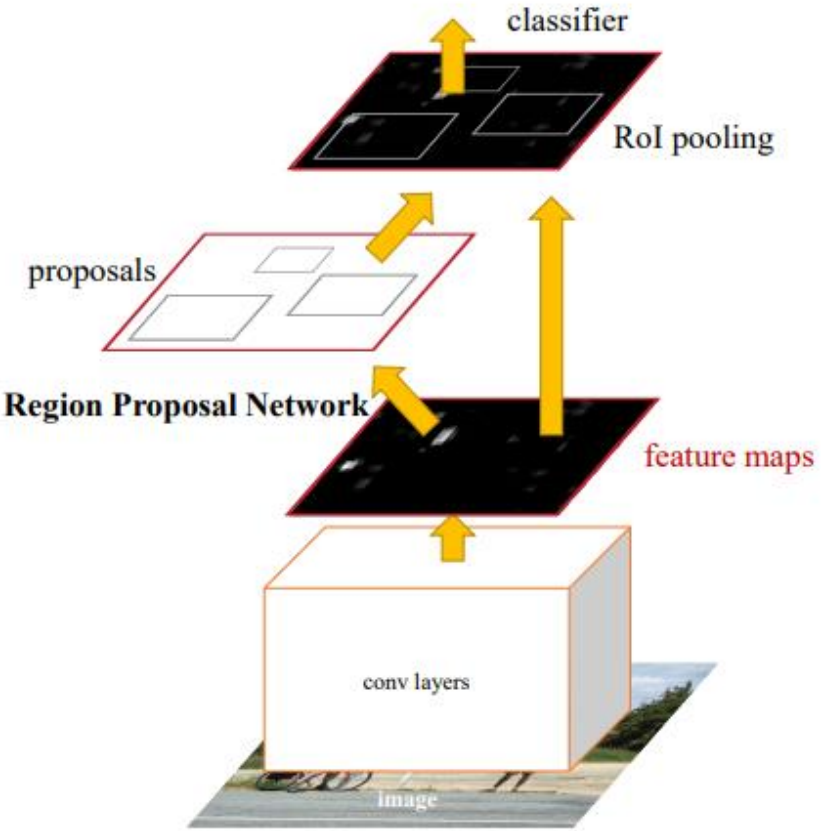
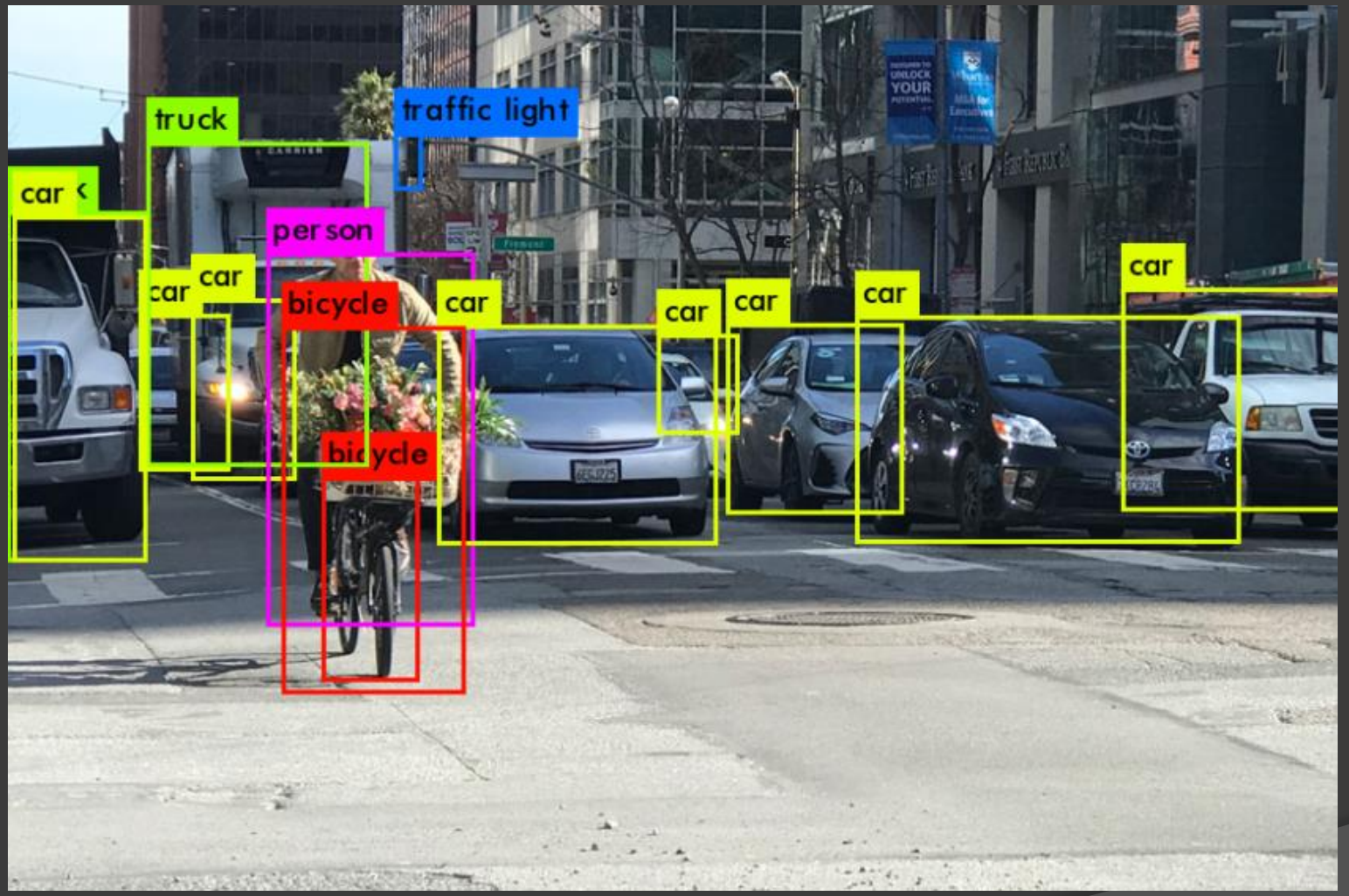
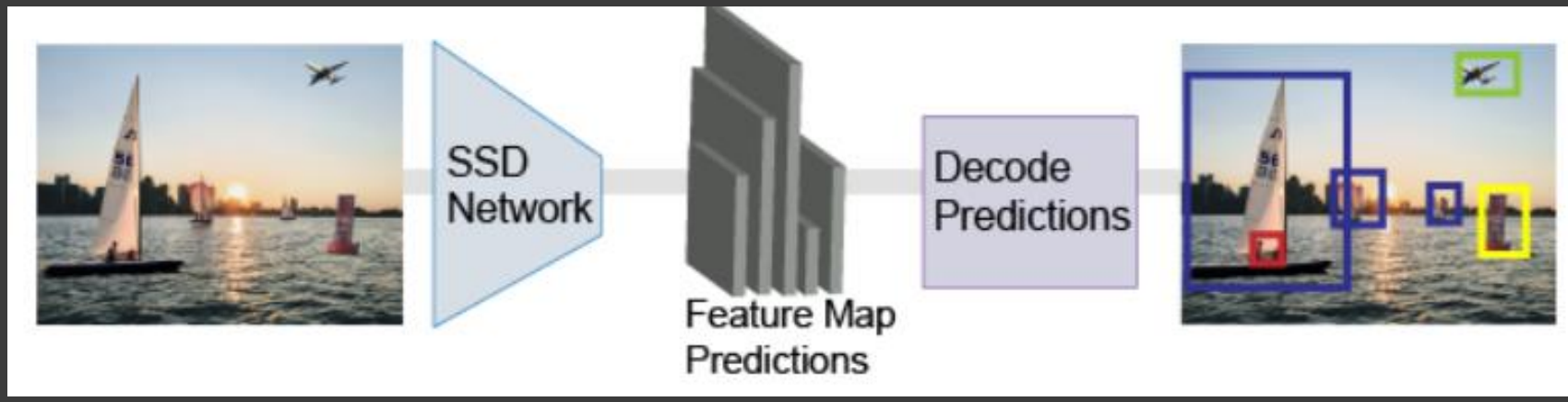


Figure 2: Faster R-CNN is a single, unified network for object detection. The RPN module serves as the 'attention' of this unified network.

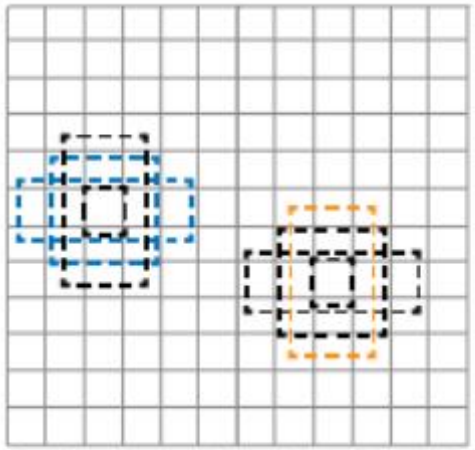
3.	Real-Time Object Detection with Yolo	Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam	• YOLO is a Regression based algorithm • In this, we won't select the interested regions from the image. Instead, we predict the classes and bounding boxes of the whole image at a single run of the algorithm and detect multiple objects using a single neural network. • YOLO algorithm is fast as compared to other classification algorithms.


truck
traffic light
car
person
car
car
bicycle
car
car
car
car
car
bicycle

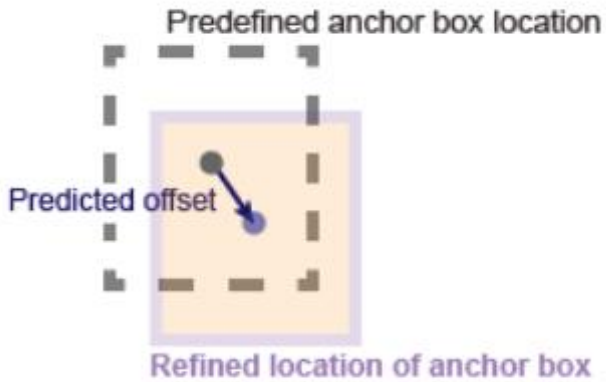
4.	SSD: Single Shot MultiBox Detector	Wei Liu , Dragomir Anguelov, Dumitru Erhan , Christian Szegedy , Scott Reed , Cheng-Yang Fu , Alexander C. Berg	• SSD specially targets to small objects. • One of its ways to detect small objects is using the anchor box. • Introducing anchor box not only increased the amount of object to detect for each cell, but also helped the network to better differentiate overlapping small objects



Ground truth image and
bounding boxes

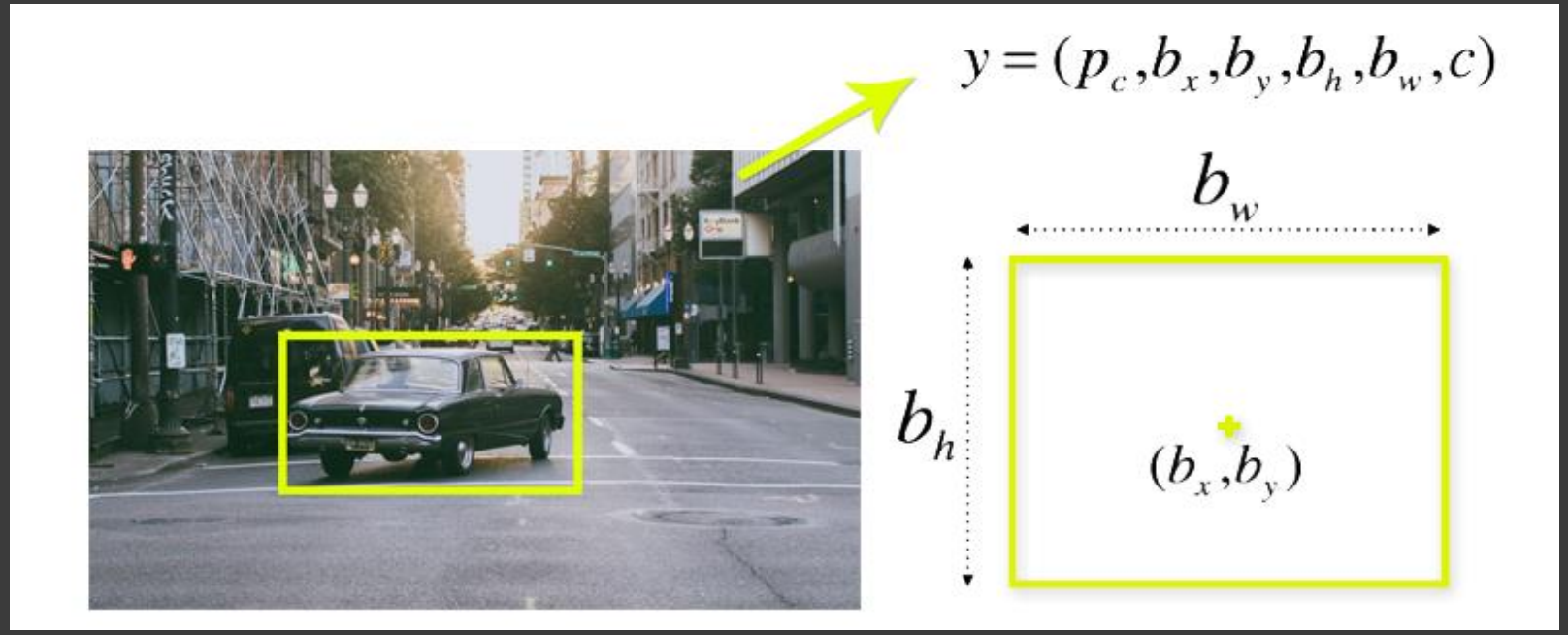


Anchor boxes at each predefined
location in each feature map



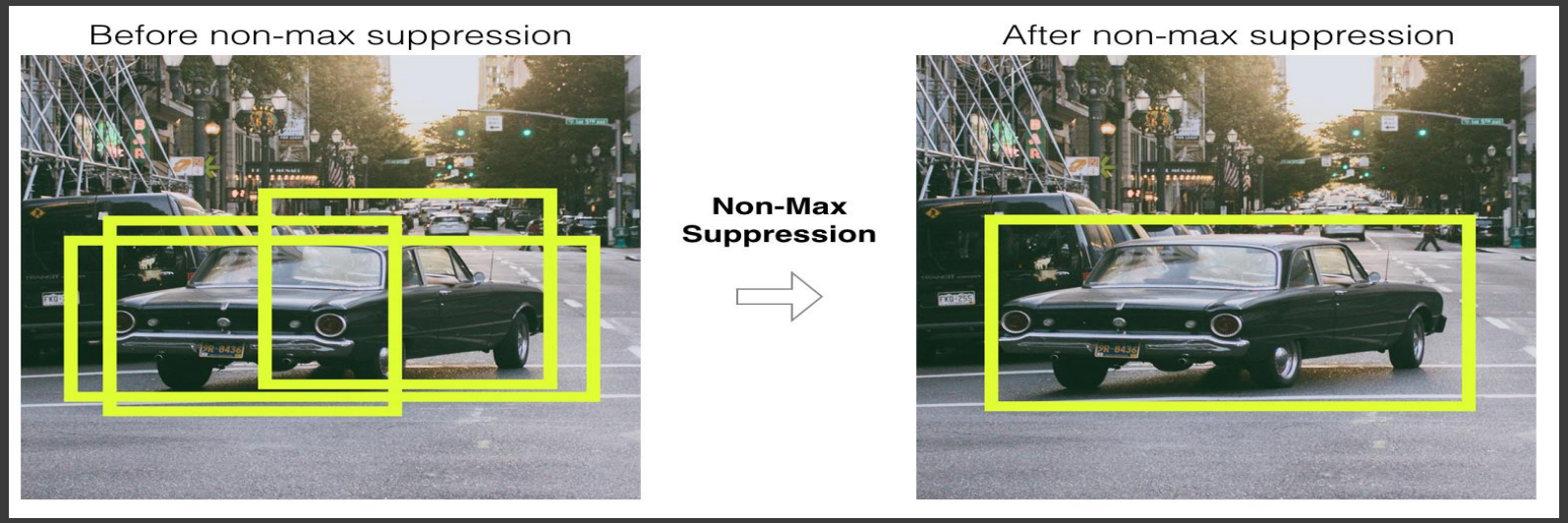
Regression based algorithms

- Algorithms based on regression – they predict classes and bounding boxes for the whole image **in one run of the algorithm**. The two best known examples from this group are the **YOLO (You Only Look Once)** family algorithms and SSD (Single Shot Multibox Detector).
- They are commonly used for real-time object detection as, in general, they trade a bit of accuracy for large improvements in speed.



- To understand the YOLO algorithm, it is necessary to establish what is actually being predicted. Ultimately, we aim to predict a class of an object and the bounding box specifying object location. Each bounding box can be described using four descriptors:

1. center of a bounding box (**bx by**)
2. width (**bw**)
3. height (**bh**)
4. value **c** is corresponding to a class of an object (such as: car, traffic lights, etc.).



- Most of the cells and bounding boxes will not contain an object.
- Therefore, we predict the value pc, which serves to remove boxes with low object probability and bounding boxes with the highest shared area in a process called **non-max suppression**.

Objectives:

- To detect all instances of objects from a known class, such as people, cars or faces in an image.
- Object detection systems construct a model for an object class from a set of training examples.
- To analyze scenes in an image or video

Scope:

- The scope of this project is to detect all instances of objects from a known class such as people cars or faces in an image.
- Once an object instance has been detected (e.g., a face), it is be possible to obtain further information, including: to recognize the specific instance (e.g., to identify the subject's face), to track the object over an image sequence (e.g., to track the face in a video), and to extract further information about the object (e.g., to determine the subject's gender)
- The system developed in this project is such that it will add a bounding box to locate an object in an image once it is detected

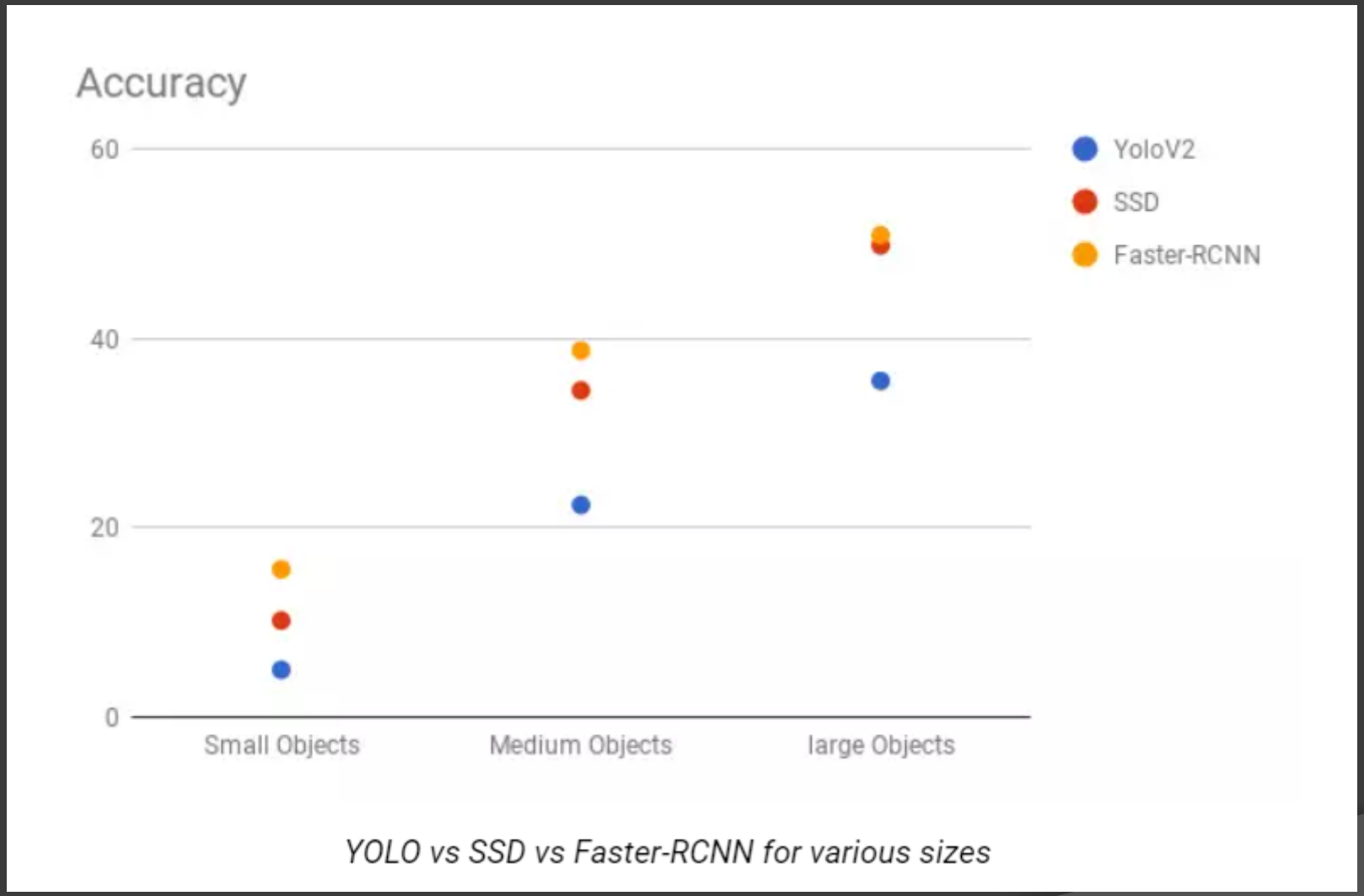
Problem Statement:

To build a system that will detect all instances of objects from a known class such as people cars or faces in an image.

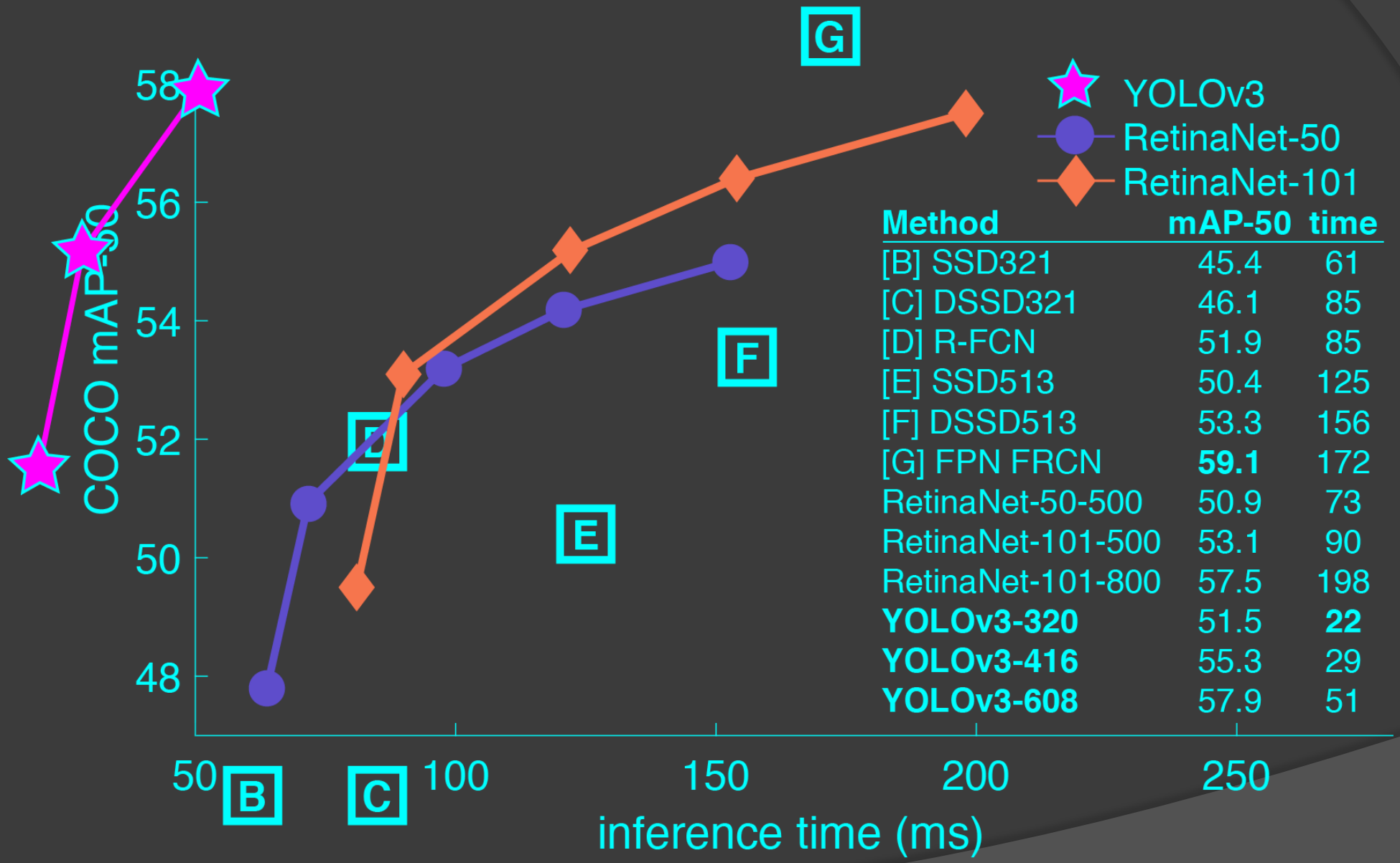
Sub-problem:

- To detect objects from several different classes
- To classify multiple objects from a single image.
- To create a bounding box for the images detected

Comparison between different models



YOLO vs SSD vs Faster-RCNN for various sizes



Method	mAP-50	time
[B] SSD321	45.4	61
[C] DSSD321	46.1	85
[D] R-FCN	51.9	85
[E] SSD513	50.4	125
[F] DSSD513	53.3	156
[G] FPN FRCN	**59.1**	172
RetinaNet-50-500	50.9	73
RetinaNet-101-500	53.1	90
RetinaNet-101-800	57.5	198
YOLOv3-320	51.5	**22**
YOLOv3-416	55.3	29
YOLOv3-608	57.9	51

Most suitable model for the project: YOLO

- YOLOv3 is fast
- In mAP measured at .5 IOU YOLOv3 is on par with Focal Loss but about 4x faster.
- Moreover, you can easily tradeoff between speed and accuracy simply by changing the size of the model, no retraining required!

Conclusion:

- We need to train a neural network based on a images dataset.
- Various filters have to be used to identify , classify objects.
- Bounded-Box must be used to localize object. Ie. Provide object location.
- Our objective will be to detect and locate objects using suitable methodologies.

References:

- Region-based Convolutional Networks for Accurate Object Detection and Segmentation Ross Girshick, Jeff Donahue, Student Member, IEEE, Trevor
- Submitted on 8 Jun 2015 (v1), last revised 9 May 2016 (this version, v5)] YOLO v1 Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi
- [Submitted on 8 Dec 2015 (v1), last revised 29 Dec 2016 (this version, v5)] SSD Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott
- Reed, Cheng-Yang Fu, Alexander C. Berg[Submitted on 4 Jun 2015 (v1), last revised 6 Jan 2016 (this version, v3)] Faster R-CNN Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun