KONKAN GYANPEETH COLLEGE OF ENGINEERING, KARJAT

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

Object Detection and Localization

Guide:

Professor A. D. Palsodkar

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

PAPER TITLE	AUTHOR	DESCRIPTION
• 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.

- Object Detection Algorithm based on improved YOLO v3

Liquan Zhao, Shuaiyang Li

- The ‘You Only Look Once’ v3 (YOLOv3) method is among the most widely used deep learning-based object detection methods.

- It uses the k-means cluster method to estimate the initial width and height of the predicted bounding boxes.

- You Only Look Once: Unified, Real-Time Object Detection

Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi)

- Authors present YOLO, an approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, they frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities.

- A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation

- YOLO9000: Better, Faster, Stronger

Joseph Redmon, Ali Farhadi

- In this authors introduce YOLO9000, a state-of-the-art, real-time object detection system that can detect over 9000 object categories. They propose various improvements to the YOLO detection method, both novel and drawn from prior work.

- Finally, they propose a method to jointly train on object detection and classification.

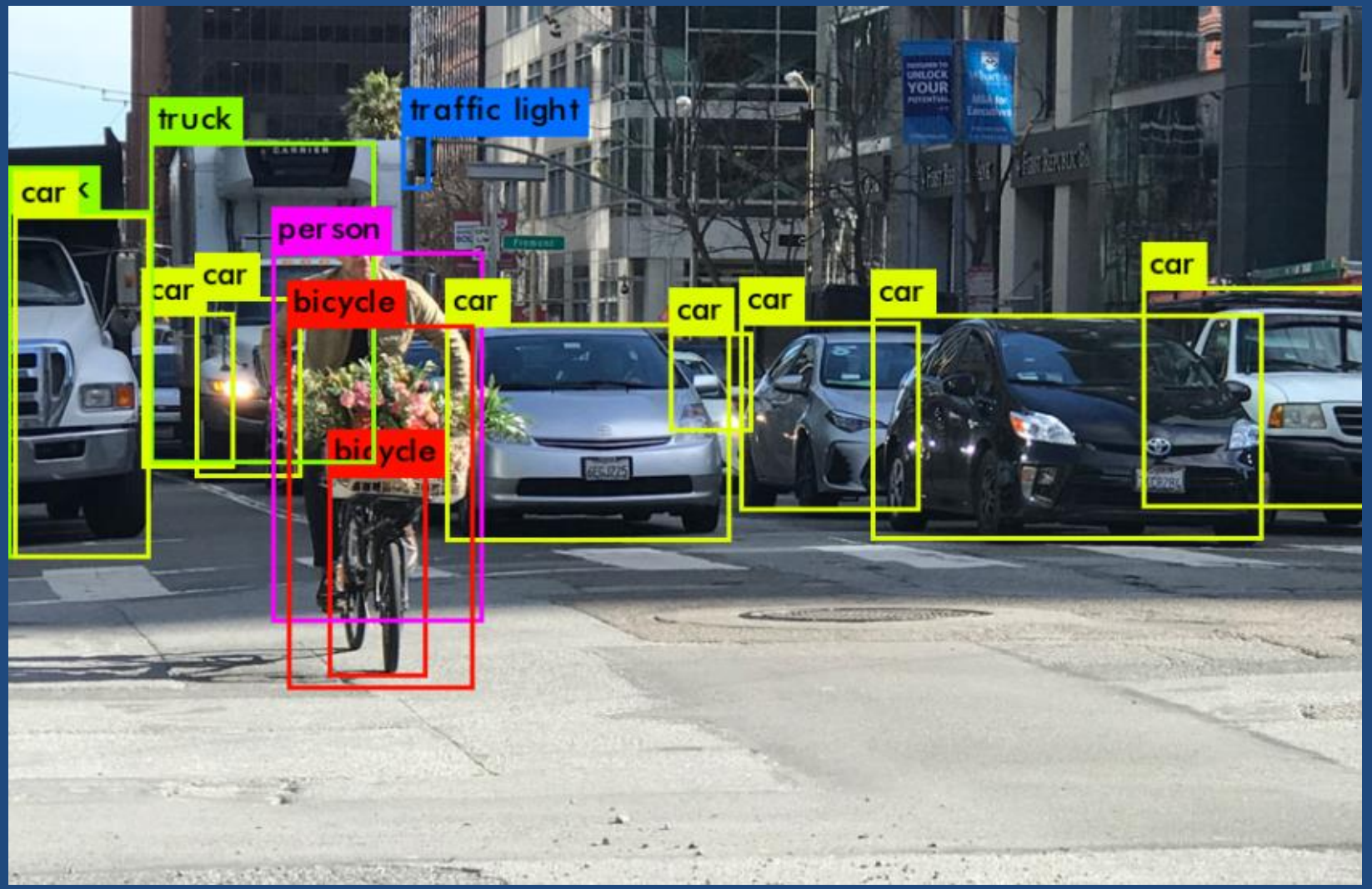
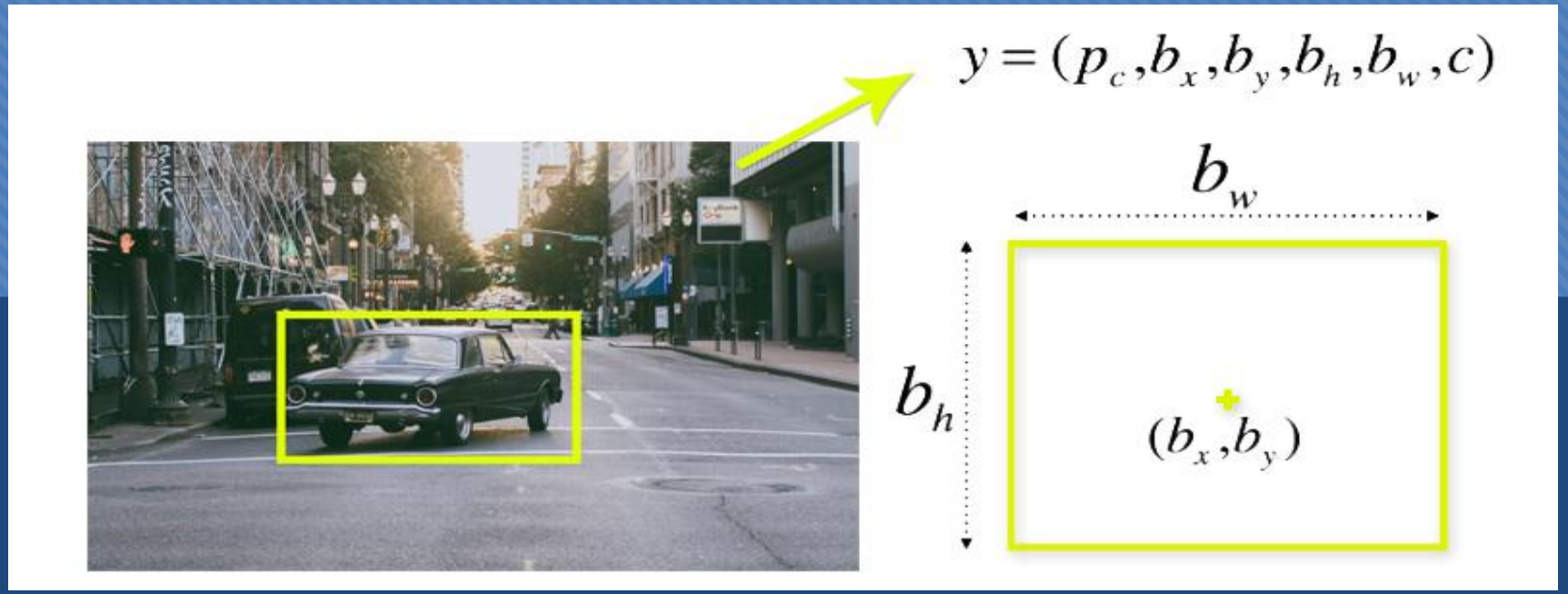


Fig: YOLO in action

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.).

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.
- Identifying the type of object in an image and also exact location of the object inside image.
- To analyze scenes in an image.

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

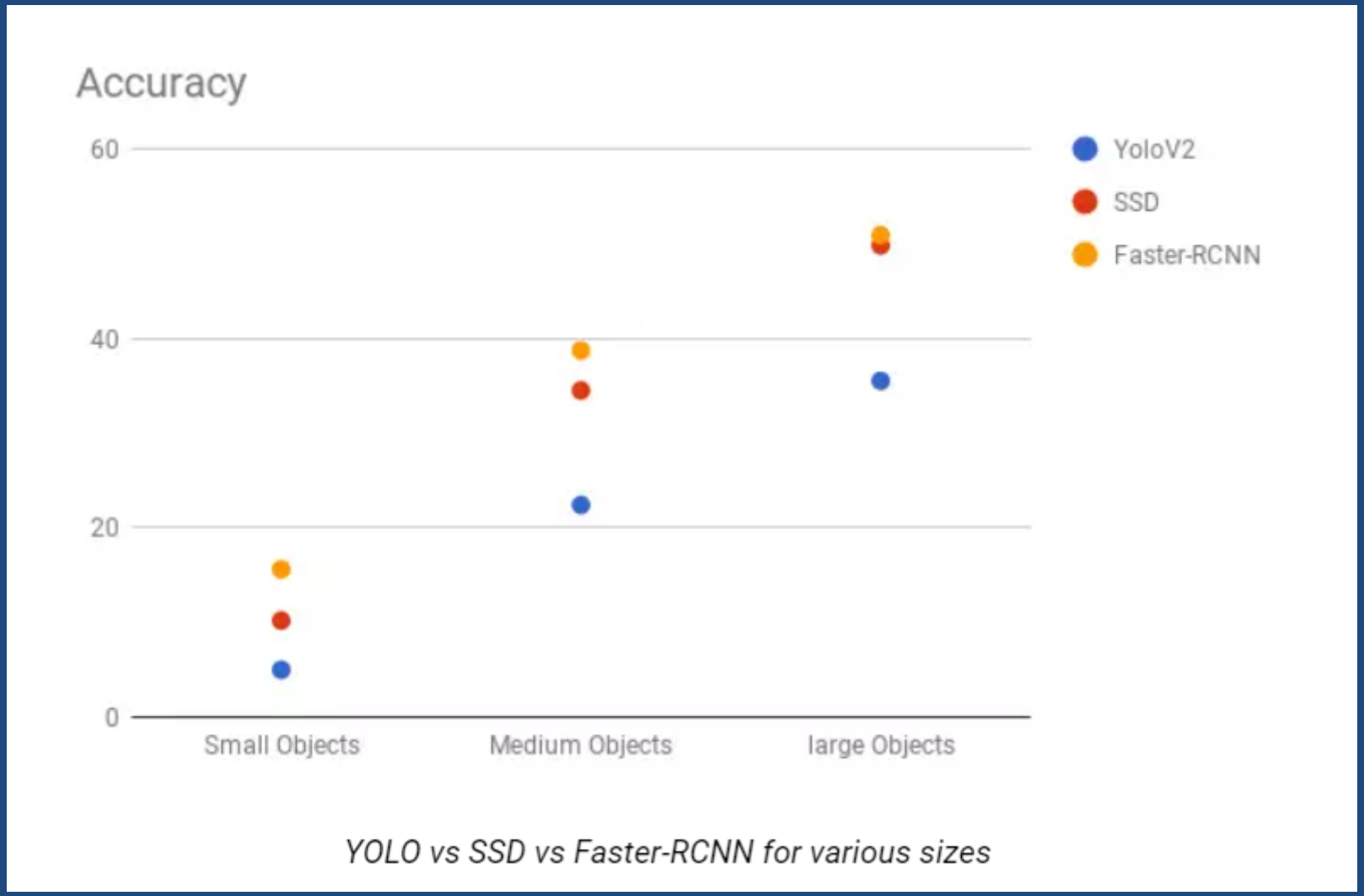
Problem Statement:

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

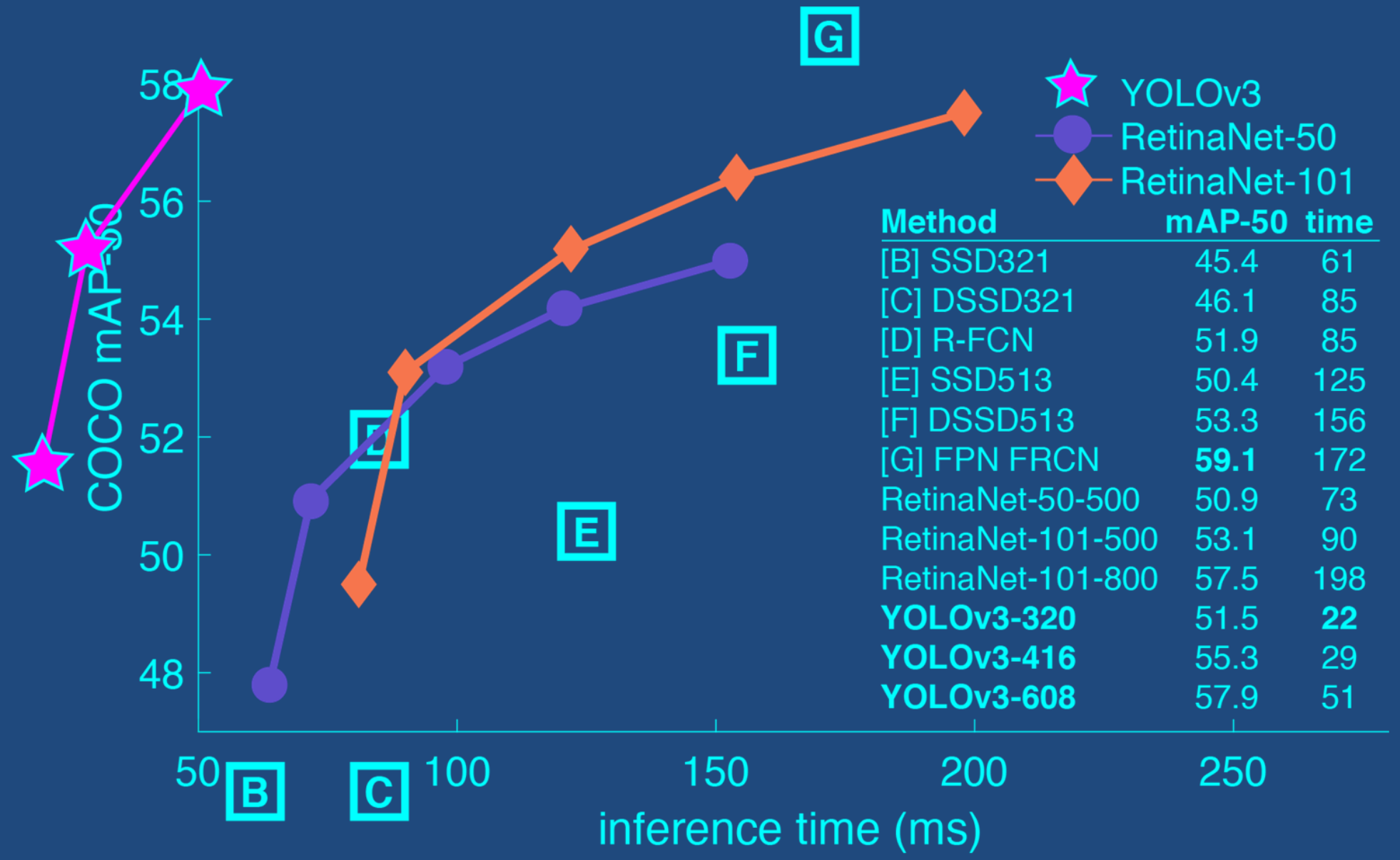
Comparison between different models



YOLO vs SSD vs Faster-RCNN for various sizes

YOLO

- The "You Only Look Once," or YOLO, family of models are a series of end-toend deep learning models designed for fast object detection, developed by Joseph Redmon, et al. and first described in the 2015 paper titled "You Only Look Once: Unified, Real-Time Object Detection."
- The approach involves a single deep convolutional neural network (originally a version of GoogLeNet, later updated and called DarkNet based on VGG) that splits the input into a grid of cells and each cell directly predicts a bounding box and object classification.
- YOLO can work well for multiple objects where each object is associated with one grid cell. But in the case of overlap, in which one grid cell actually contains the centre points of two different objects, we can use something called anchor boxes to allow one grid cell to detect multiple objects.



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

Flowchart

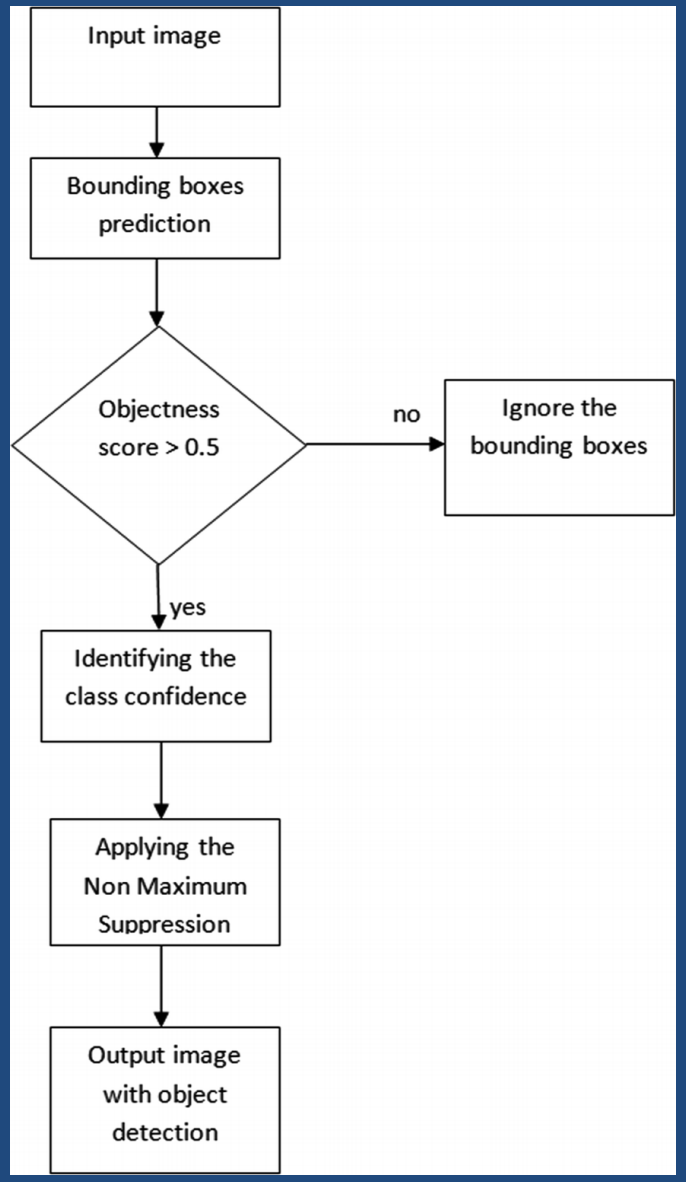


Fig: flowchart

Darknet

- Darknet is mainly for object detection, & have different architecture, features than other deep learning frameworks.
- It is faster than many other NN architectures & approaches like FasterRCNN, etc.
- Darknet Architecture is a specialized framework, & is very useful because of its faster rate & accuracy.

Implementation

Fig: Homepage for the webapp

Fig: Uploading an image for detecting different classes of objects in it

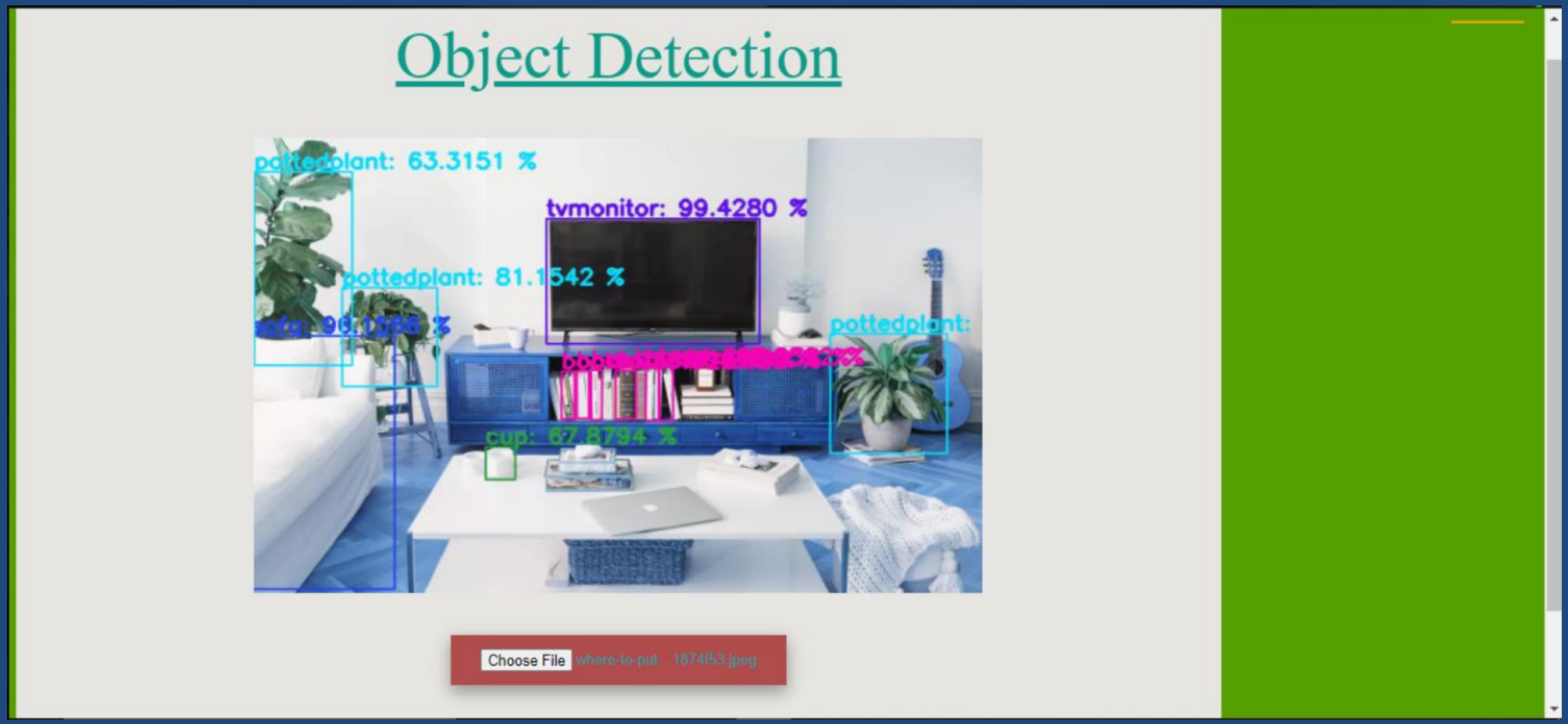

Object Detection
pottedplant: 63.3151 %
tvmonitor: 99.4280 %
pottedplant: 81.1542 %
pottedplant:
cup: 67.8794 %
Choose File
where-to-put...1874l53.jpeg

Fig: Output

Conclusion:

- Object detection models are capable of tracking multiple people at once, in real-time, as they move through a given scene or across video frames. From retail stores to industrial factory floors, this kind of granular tracking could provide invaluable insights into security, worker performance and safety, retail foot traffic, and more.
- After researching through various papers related to Object Detection and Localization, we took a basic object detection model and improved its accuracy to detect images from different classes and produce a bounding box around it. Comparing between several models, we decided to use YOLO because of its speed and convenience for real time object detection.

References:

- Object Detection Algorithm based on improved YOLO v3 (Author: Liquan Zhao, Shuaiyang Li)
- Submitted on 8 Jun 2015 (v1), last revised 9 May 2016 (this version, v5)] YOLO v1 Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi
- You Only Look Once: Unified, Real-Time Object Detection (Author: Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi)
- 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
- YOLO9000: Better, Faster, Stronger (Author: Joseph Redmon, Ali Farhadi)