Object Recognition and Localisation

Project Report

Supervisor: Dr Amit Kumar

INDIAN INSTITUTE OF
INFORMATION TECHNOLOGY
KOTA

Team Members:

Aditya Shrivastava | Ram Ratan | Saurabh Mina

2019KUCP1038 | 2019KUCP1039 | 2019KUCP1085

Abstract

Computer Vision is the branch of the science of computers and software systems which can recognise as well as understand images and scenes. Computer Vision is consists of various aspects such as image recognition, object detection, image generation, image super-resolution and many more. Object detection is widely used for face detection, vehicle detection, pedestrian counting, web images, security systems and self-driving cars. In this project, we are using highly accurate object detection-algorithms and methods such Mask-RCNN and fast yet highly accurate ones like SSD. In this project we have developed two approaches to localise a point in the image frame of reference to the real world coordinates system by using a simple mathematical approach and further calculated the errors associated with both the methods to find the implementation more suitable for our use case.

Supervisor: Dr Amit Kumar

Team Members:

Aditya Shrivastava	2019KUCP1038
Ram Ratan	2019KUCP1039
Saurabh Mina	2019KUCP1085

Introduction

1.1 Objective

The goal of “object detection” is to find the location of an object in a given picture accurately and mark the object with the appropriate category. To be precise, the problem that object detection seeks to solve involves determining where the object is, and what it is. However, solving this problem is not easy. Unlike the human eye, a computer processes images in two dimensions. Furthermore, we want to localise the said object with high accuracy.

1.2 Approach

The approach used incorporates computer vision and machine learning concepts along with mathematical concepts to calculate the 3D coordinates of any given point on the 2D image plane.

1.3 Applications

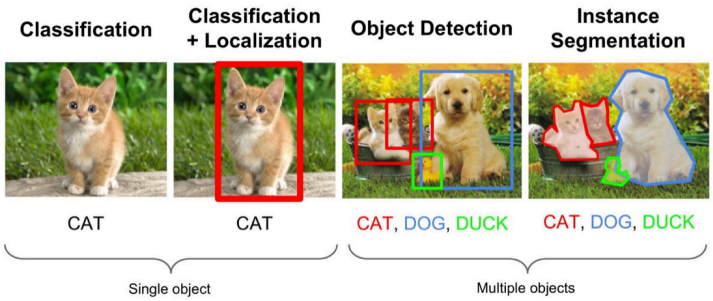
1.3.1 Prosthetic Arm Grasping

To make an intelligent prosthetic arm capable of grasping user defined objects we must estimate the real world position of said object in real time with a low margin of error, our proposed method can full fill this task.

1.3.2 Self Driving Cars

Self-driving cars are the Future, there’s no doubt in that. But the working behind it is very tricky as it combines a variety of techniques to perceive their surroundings, including radar, laser light, GPS, and computer vision to avoid collisions a method must be developed which can detect and localise objects with fast speeds and low error

Object Detection



2.1 Computer Vision Concepts

Object Detection is the process of finding and recognising real-world object instances such as car, bike, TV, flowers, and humans out of an images or videos. An object detection technique lets you understand the details of an image or a video as it allows for the recognition, localisation, and detection of multiple objects within an image

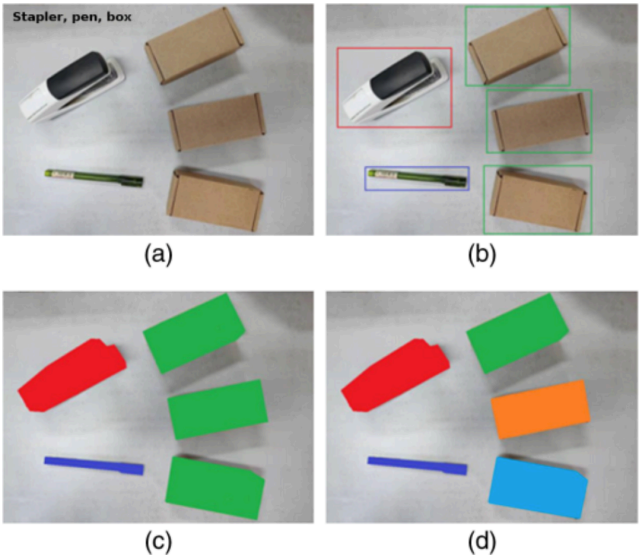
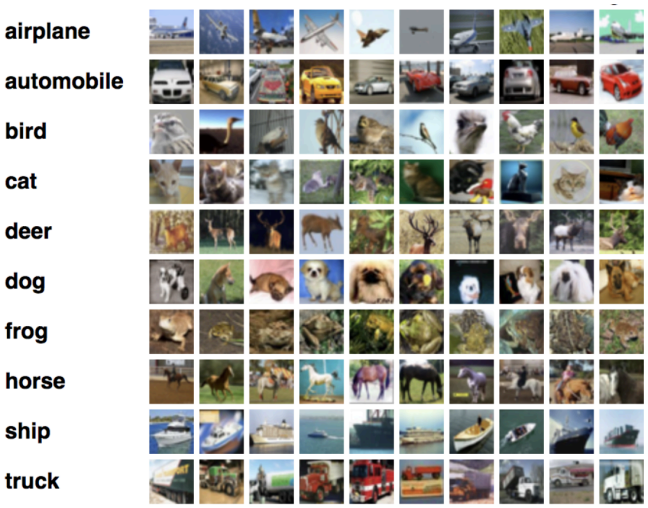


Diagram of the four basic tasks of computer vision:
(a) image classification,
(b) object detection,
(c) semantic segmentation, and
(d) instance segmentation.

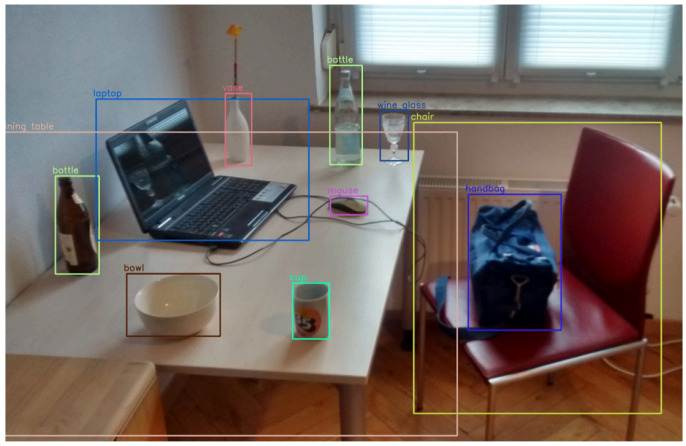
2.1.1 Image Classification:

Image Classification (often referred to as Image Recognition) is the task of associating one (single-label classification) or more (multi-label classification) labels to a given image.



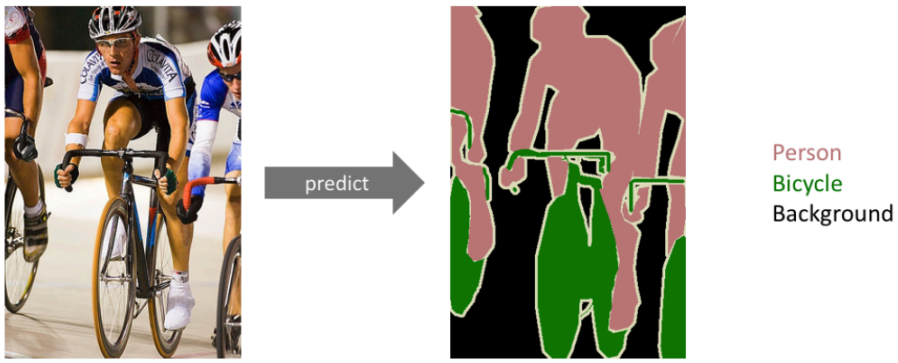
2.1.2 Object detection :

Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class in digital images and videos.



2.1.3 Semantic segmentation:

Semantic segmentation is a computer vision task in which we label specific regions of an image according to what's being shown.

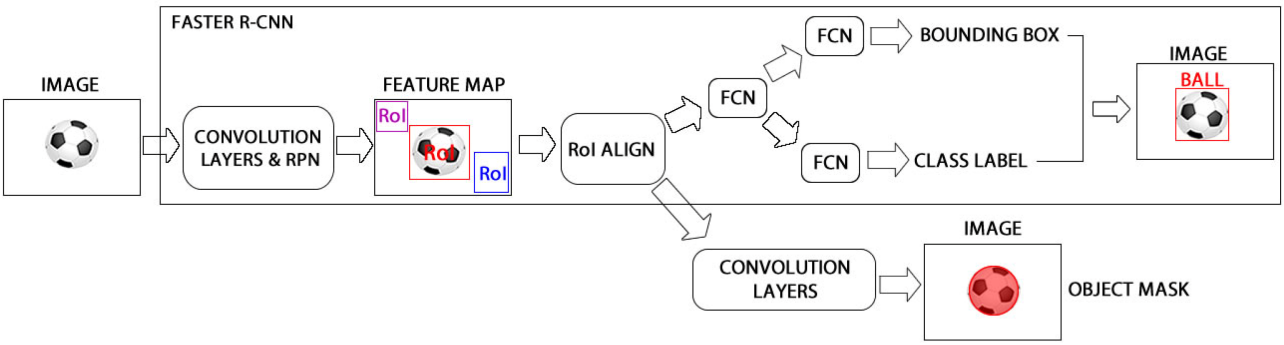


2.1.4 Instance Segmentation:

Instance segmentation is a computer vision task for detecting and localising an object in an image. Instance segmentation is a natural sequence of semantic segmentation, and it is also one of the biggest challenges compared to other segmentation techniques.

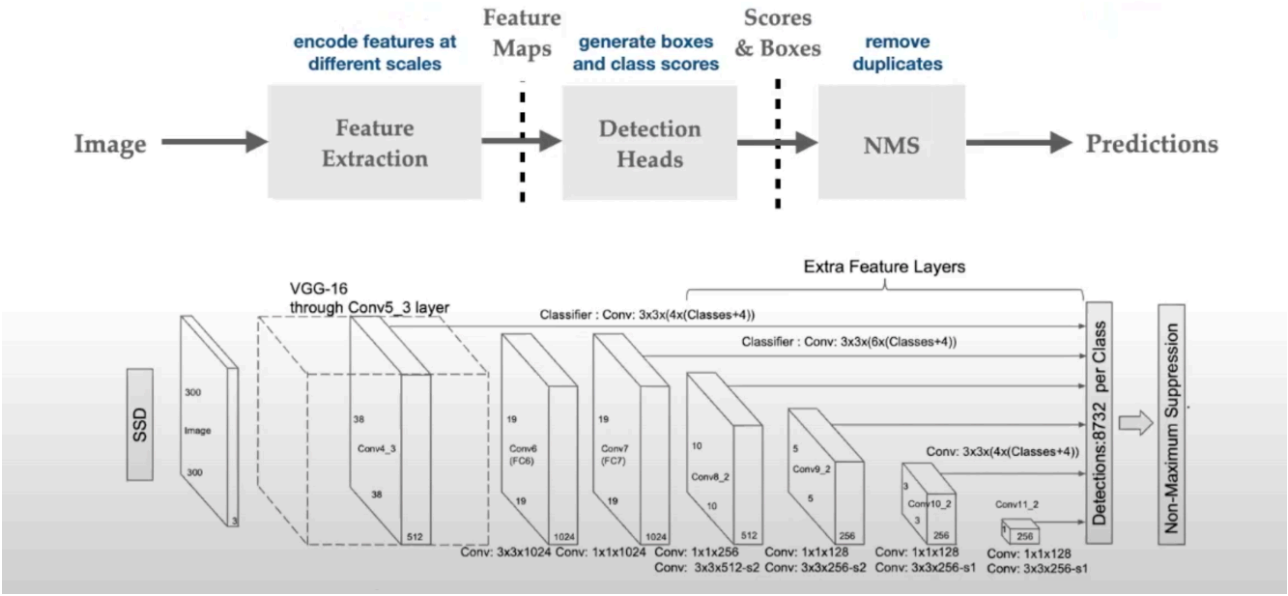
Object Detection Methods

3.1 Mask RCNN



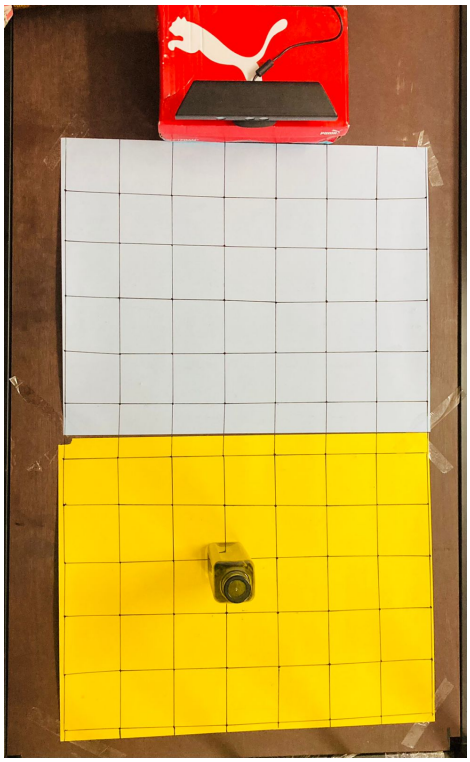
The Mask R-CNN architecture, which is composed of convolution layers, region proposal networks (RPNs), and fully connected networks (FCNs). The Faster R-CNN performs the region proposal selection; the region of interest (RoI) ALIGN sets up all RoIs to the same shape; the FCNs make the object labelling and the bound box; and the convolution layers perform the pixel determination of each object (object mask).

3.2 SSD (Single Shot Detector)

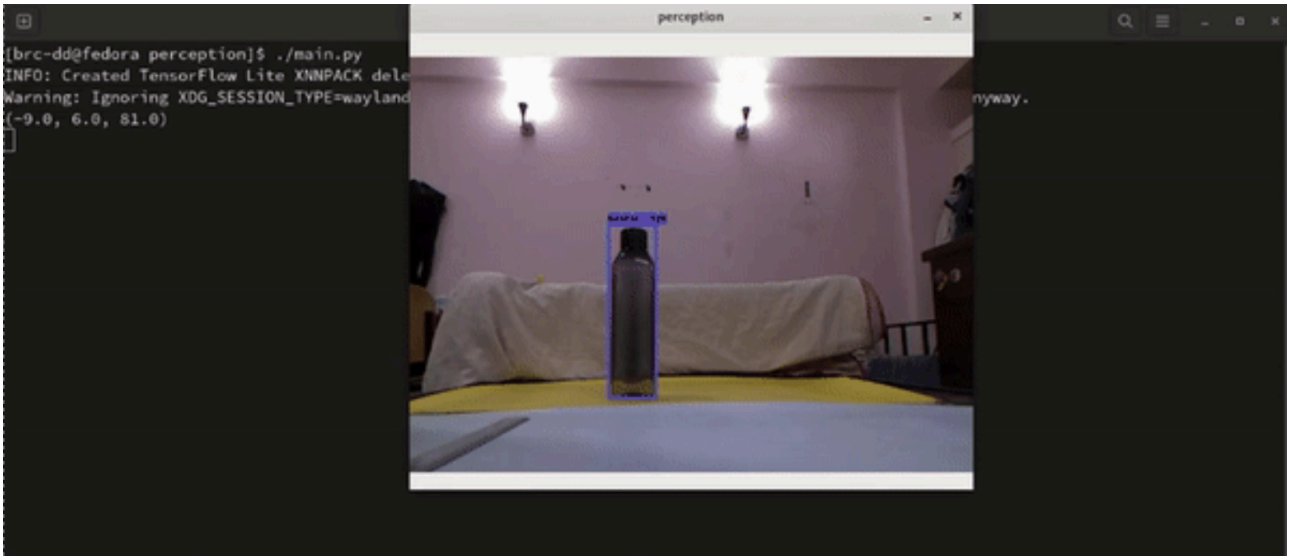


The SSD model works as follows, each input image is divided into grids of various sizes and at each grid, the detection is performed for different classes and different aspect ratios. And non maximum suppression is applied to get the final detection from the set of overlapping detections. This is the basic idea behind the SSD model.

4 Setup



5 Output



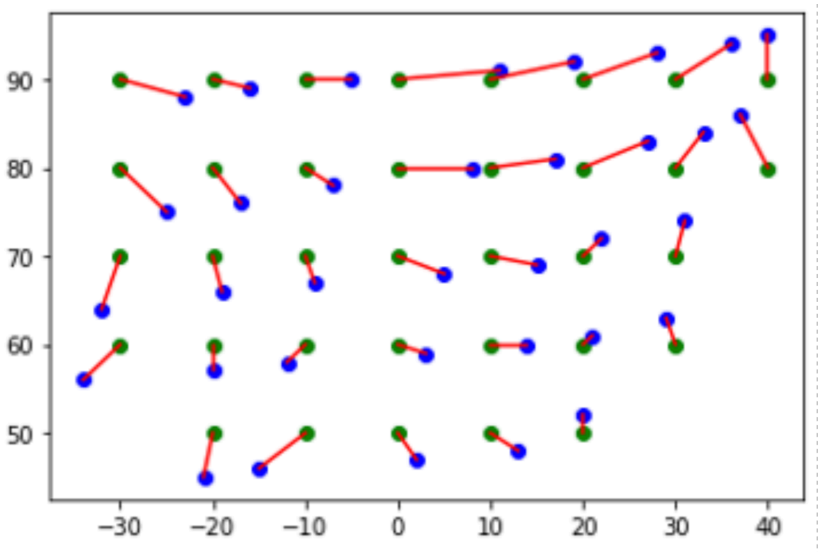
6 Results

6.1 Mask RCNN

Traverse the mask calculate the coords for each point take average over the whole bottle to get final results

Error = +-7cm

Depth / Z - coordinate



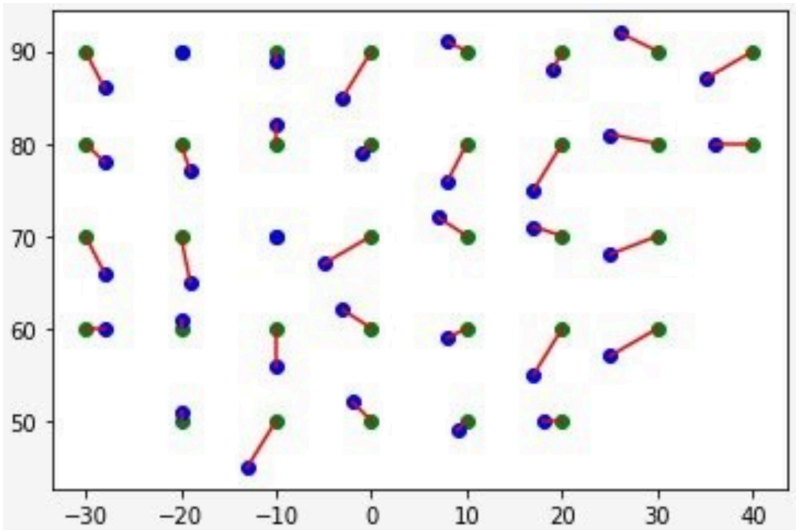
X - coordinate

6.2 SSD

Calculate coords of each point then calculate the median of the values to get results

Error = +-3cm

Depth / Z - coordinate



X - coordinate

Future Scope

The main useful applications of object detection: Vehicle's Plates recognition, self-driving cars, Tracking objects, face recognition, medical imaging, object counting, object extraction from an image or video, person detection.

We have had great progress in the field, processing a single image used to take 20 seconds per image and today it takes less than 20 milliseconds.

The possibility of applications of object detection systems to robotic excavation when venturing into previously unexplored territory, such as the deep sea or other planets, in which the detection systems will have to learn new object classes on the job.

Conclusion

Our purposed object recognition and localisation algorithm is completely implemented using SSD mobile net. We chose SSD over Mask RCNN as We get less error using SSD method as compared to Mask RCNN method.
For our future implementation are planning to implement grasping for PROSTHETIC ARM.

References

- K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.
- Liu, Wei & Anguelov, Dragomir & Erhan, Dumitru & Szegedy, Christian & Reed, Scott & Fu, Cheng-Yang & Berg, Alexander. (2016). SSD: Single Shot MultiBox Detector. 9905. 21-37. 10.1007/978-3-319-46448-0_2.
- C. KAYMAK and A. UCAR, "Implementation of Object Detection and Recognition Algorithms on a Robotic Arm Platform Using Raspberry Pi," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018, pp. 1-8, doi: 10.1109/IDAP.2018.8620916.
- C. -H. Chen, H. -P. Huang and S. -Y. Lo, "Stereo-based 3D localization for grasping known objects with a robotic arm system," 2011 9th World Congress on Intelligent Control and Automation, 2011, pp. 309-314, doi: 10.1109/WCICA.2011.5970749.