**ABSTRACT**

Efficient and accurate object detection has been an important topic in the advancement of computer vision systems. With the advent of deep learning techniques, the accuracy for object detection has increased drastically. The project aims to incorporate state-of-the-art technique for object detection with the goal of achieving high accuracy with a real-time performance. A major challenge in many of the object detection systems is the dependency on other computer vision techniques for helping the deep learning based approach, which leads to slow and non-optimal performance. In this project, we use a completely deep learning based approach to solve the problem of object detection in an end-to-end fashion.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Sr.no.** | **Topics** | **Page no.** |
| **1** | **Introduction** | **3** |
| 1.1 | Problem Statement | 3 |
| **2** | **Related Work** | **5** |
| 2.1 | Bounding Box | 5 |
| 2.2 | Classification + Regression | 5 |
| 2.3 | Unified Method | 6 |
| **3** | **Hardware/software implementations** | **7** |
| **4** | **Implementation works details** | **8** |
| 4.1 | Real-life application | 8 |
| 4.2 | Object detection workflow | 9 |
| **5** | **Chapter source code** | **10** |
| **6** | **Output screens** | **11** |
| **7** | **Individual contribution** | **12** |
| **8** | **Conclusion** | **13** |
| 8.1 | Limitations | 13 |
| 8.2 | Future work | 14 |
| **9** | **Bibliography** | **16** |
| **10** | **Annexures** | **17** |
|  | - Reference  - Plagiarism | 17  18 |

**CHAPTER1: INTRODUCTION**

* 1. **Problem Statement**:

Many problems in computer vision were saturating on their accuracy before a decade. However, with the rise of deep learning techniques, the accuracy of these problems drastically improved. One of the major problem was that of image classification, which is defined as predicting the class of the image. A slightly complicated problem is that of image localization, where the image contains a single object and the system should predict the class of the location of the object in the image (a bounding box around the object). The more complicated problem (this project), of object detection involves both classification and localization. In this case, the input to the system will be a image, and the output will be a bounding box corresponding to all the objects in the image, along with the class of object in each box. An overview of all these problems is depicted in Fig. 1

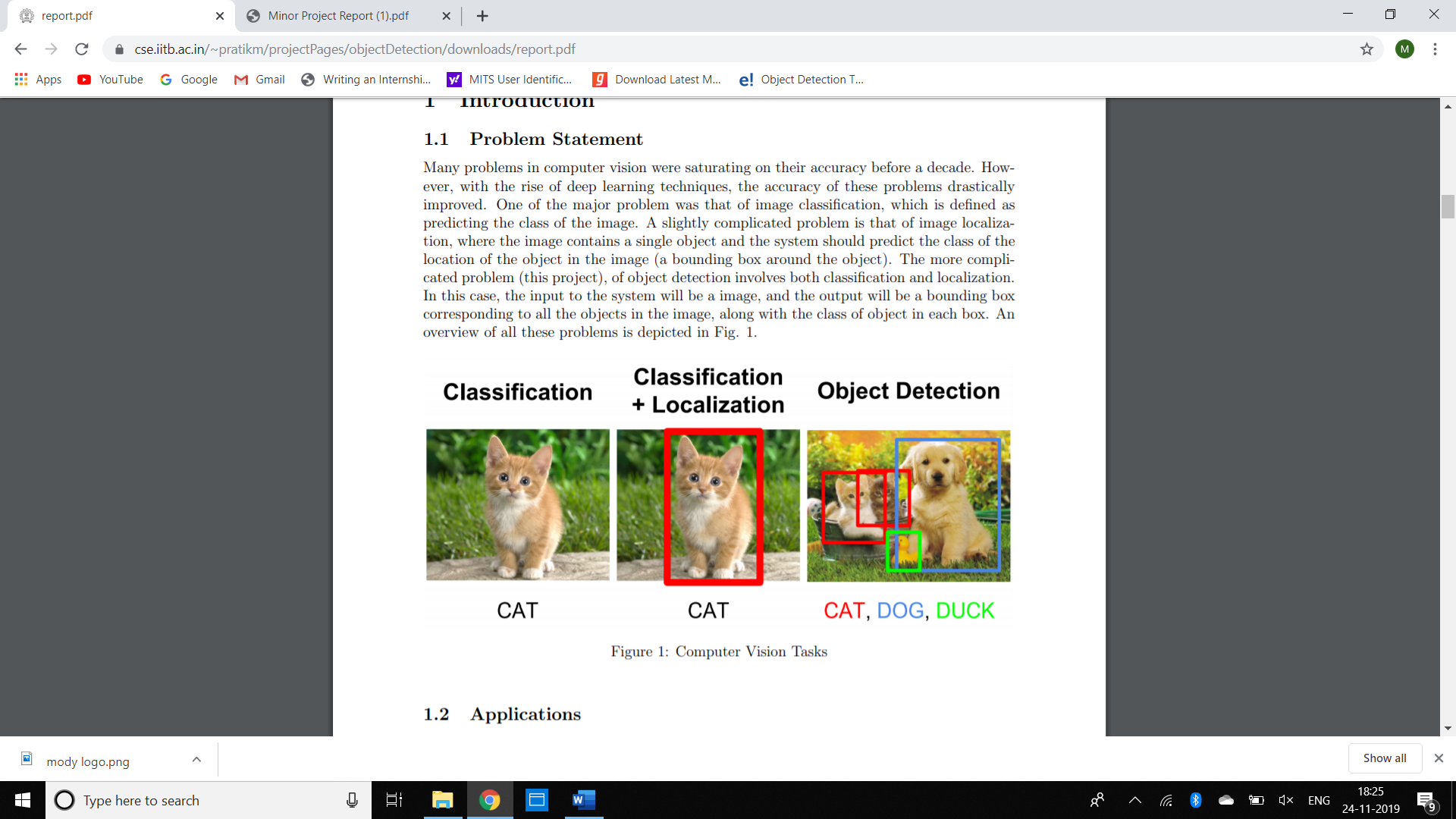


Figure 1: Computer Vision Task



Figure 2: Applications of object detections

**CHAPTER 2: RELATED WORK**

The major concepts involved during the implementation of the project have been discussed below.

**2.1 Bounding Box**

The bounding box is a rectangle drawn on the image which tightly fits the object in the image. A bounding box exists for every instance of every object in the image. For the box, 4 numbers (center x, center y, width, height) are predicted. This can be trained using a distance measure between predicted and ground truth bounding box. The distance measure is a jaccard distance which computes intersection over union between the predicted and ground truth boxes as shown in Fig. 4.

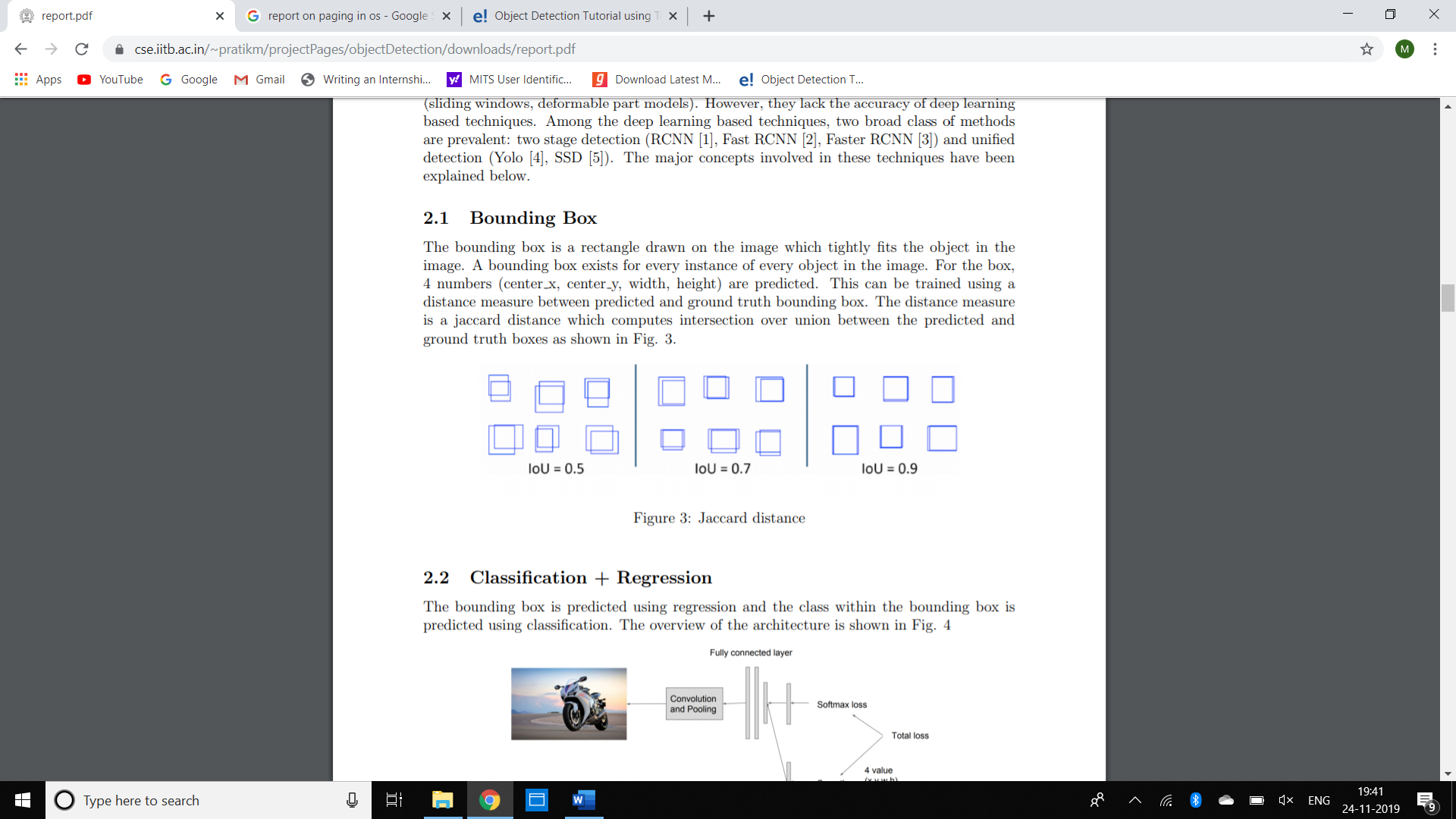


Figure 4: Jaccard Distance

**2.2 Classification + Regression**

The bounding box is predicted using regression and the class within the bounding box is predicted using classification. The overview of the architecture is shown in Fig. 5

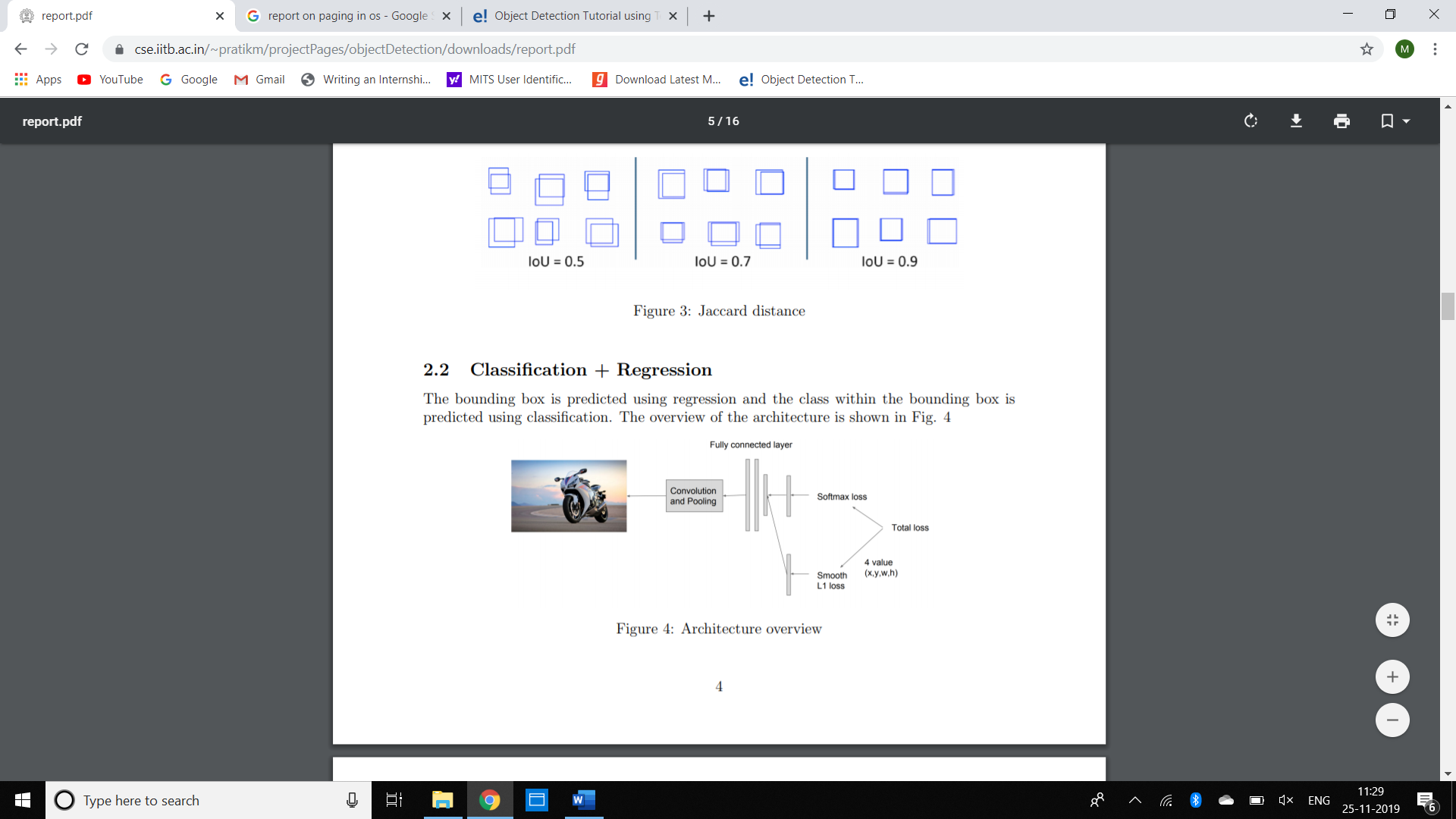


Figure 5: Architectural overview

**2.3Unified Method**

This method produces a pre-define a set of boxes to lookfor objects. Using convolutional feature maps from later layers of the network, run anothernetwork over these feature maps to predict class scores and bounding box offsets. The broadidea is depicted in Fig. 6. The steps are mentioned below:

1. Train a CNN with regression and classification objective.

2. Gather activation from later layers to infer classification and location with a fullyconnected or convolutional layers.

3. During training, use jaccard distance to relate predictions with the ground truth.

4. During inference, use non-maxima suppression to filter multiple boxes around the same object.

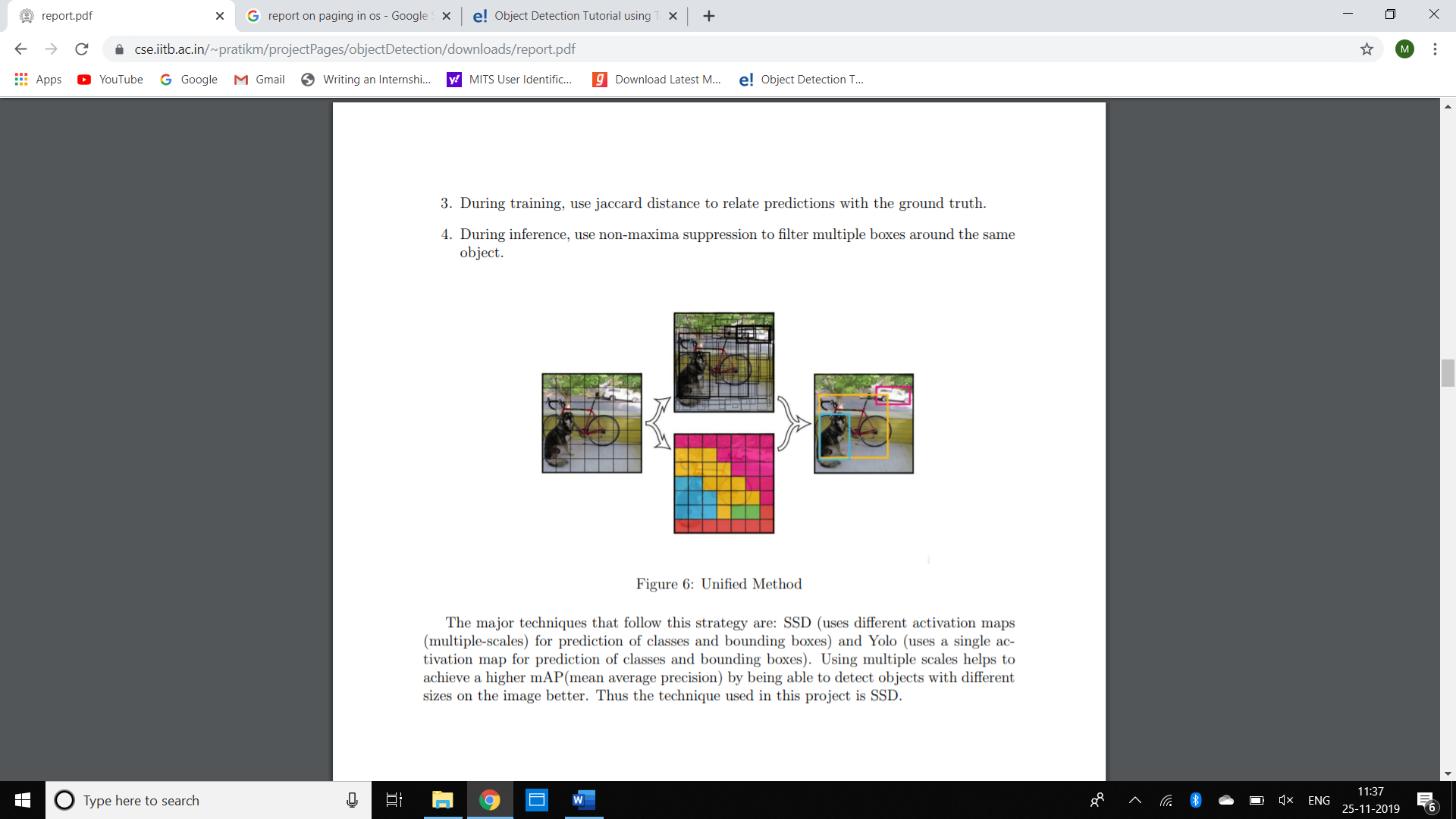


Figure 6: Unified Method

The major techniques that follow this strategy are: SSD (uses different activation maps (multiple-scales) for prediction of classes and bounding boxes) and Yolo (uses a single activation map for prediction of classes and bounding boxes). Using multiple scales helps to achieve a higher map(mean average precision) by being able to detect objects with different sizes on the image better. Thus the technique used in this project isSD.

**CHAPTER-3 HARDWARE/SOFTWARE IMPEMENTATIONS**

Hardware Used:

Windows operating system 64bit

Software Used:

1. Jupiter Notebook
2. Tensorflow
3. Anaconda
4. Opencv library

**CHAPTER 4 IMPLEMENTATION WORKS DETAILS**

The project is implemented in python 3. Tensorflow was used for training the deep network and OpenCV was used for image pre-processing.

The system specifications on which the model is trained and evaluated are mentioned as follows: CPU - Intel Core i5-8265U 3.4 GHz, RAM –8 Gb, GPU - Nvidia Titan Xp.

**Dataset:** For the project purpose COCO dataset is used. It is publicly available pre-trained dataset .It is a large-scale object detection, segmentation, and captioning dataset.

**4**.**1 Real-life application**

Self-Driving Cars: Self-driving cars (also known as autonomous cars) are vehicles that are capable of moving by themselves with little or no human guidance. Now, in order for a car to decide its next step, i.e. either to move forward or to apply breaks, or to turn, it must know the location of all the objects around it. Using Object Detection techniques, the car can detect objects like other cars, pedestrians, traffic signals, etc

**Face detection:** Popular applications include face detection and people counting.

**People Counting**: Object detection can be also used for people counting, it is used for analysing store performance or crowd statistics during festivals.

**Vehicle detection**

Similarly when the object is a vehicle such as a bicycle or car, object detection with tracking can prove effective in estimating the speed of the object. The type of ship entering a port can be determined by object detection(depending on shape, size etc). This system for detecting ships are currently in development in some European countries.

**Manufacturing Industry**

Object detection is also used in industrial processes to identify products. Say you want your machine to only detect circular objects. Hough circle detection transform can be used for detection.

**Online images**

Apart from these object detection can be used for classifying images found online. Obscene images are usually filtered out using object detection.

**Security**

In the future we might be able to use object detection to identify anomalies in a scene such as bombs or explosives (by making use of a quadcopter).

**4.2 Object Detection Workflow**

Different Object Detection has different approach but they all are based on same principle.

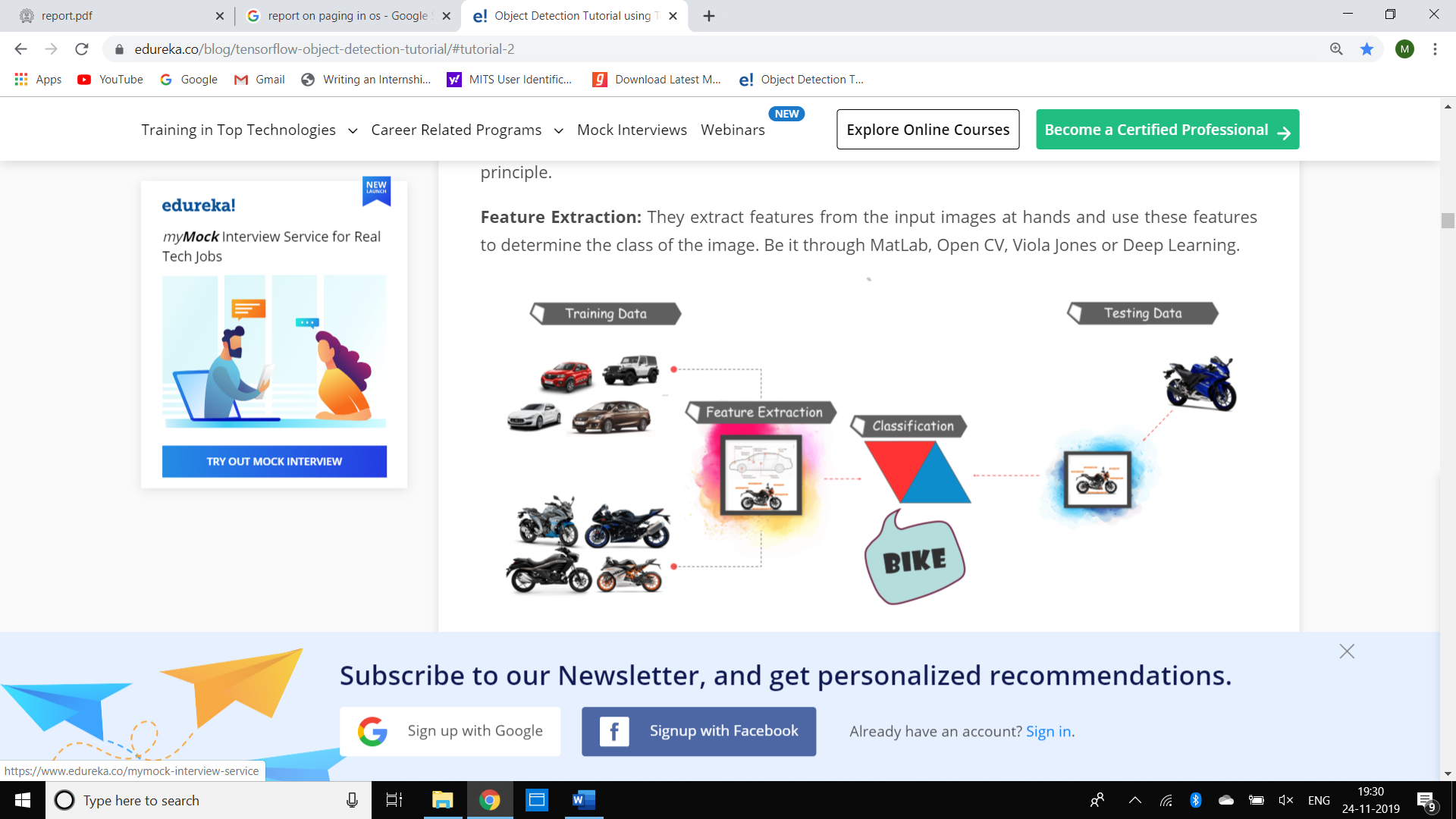


Figure 3: Object Detection Workflow

**Feature Extraction:**They extract features from the input images at hands and use these features to determine the class of the image. Be it through MatLab, Open CV, Viola Jones or Deep Learning.

**CHAPTER5-SOURCE CODE**

import cv2

import matplotlib.pyplot as plt

font=cv2.FONT\_HERSHEY\_PLAIN

video=cv2.VideoCapture(0)

count=0

while True:

count+=1

ret,frame=video.read()

frame=cv2.resize(frame,None, fx=2,fy=2)

count+=1 path='C:\\Users\\Priti\\Desktop\\goedu\\python\\project\\object\\cnn1\\img'+str(count)+'.png'

#frame\_RGB=cv2.cvtColor(frame,cv2.COLOR\_BGR2RGB)

if count%50==0:

frame=test1(frame)

cv2.imwrite(path,frame)

#will save in current folder

cv2.imread(path)

cv2.imshow('live video',frame)

#boxes,confidences,class\_ids=test(path)

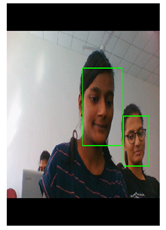
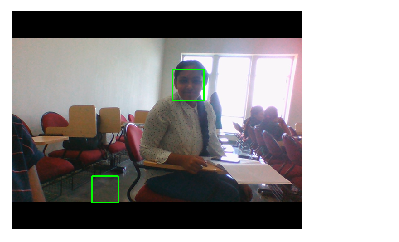
elif cv2.waitKey(20)&0xFF==27:

break

video.release()

**CHAPTER 6: OUTPUT SCREENS**

****

****

**CHAPTER 7: INDIVIDUAL CONTRIBUTION**

**CHAPTER 8: CONCLUSION**

Object detection and tracking is one of the challenging research tasks of computer vision aimed at detecting moving objects from video sequence. It is followed by predicting the path of moving object for the duration of its presence in video frame sequences. This study has provided a comprehensive review of the state-of-the-artmethods on object detection and tracking with focus on soft computing basedapproaches. The project has made the following primary contributions

* An exhaustive review of the literature related to conventional and softcomputing based approaches for detection and tracking.
* The study has analyzed various research algorithms, challenges, datasets andapplications.
  1. **Limitations**

**1.Dual priorities: object classification and localization**

* 1. The major challenge faced by real-time object detection is not only to classify the image but also to determine the object’s position also known as localization of object. The researchers might uses multi-task loss functions to penalize both misclassifications and localization errors.

1. **Speed for real time detection**
   1. The only task of live object detection is not only to correctly localize and detect important objects but they also need to be incredibly fast at prediction time to meet the real-time demands of video processing.
2. **Multiple spatial scales and aspect ratio**
   1. There are many applications of object detection in which items of interest may appear in a wide range of sizes and aspect ratios. Developers implements special techniques to ensure detecting algorithms are able to capture objects at multiple scales and views.
3. **Limited Data**
   1. A very limited amount of annotated data is available which proves to be another substantial hurdle. Object detection datasets typically contain ground truth examples for about a dozen to a hundred classes of objects, while image classification datasets can include upwards of 100,000 classes. Furthermore, crowdsourcing often produces image classification tags for free (for example, by parsing the text of user-provided photo captions). Gathering ground truth labels along with accurate bounding boxes for object detection, however, remains incredibly tedious work.
4. **Class Imbalance**
   1. Class imbalance proves to be an issue for most classification problems, and object detection is no exception. Consider a typical photograph. More likely than not, the photograph contains a few main objects and the remainder of the image is filled with background.

**8.2 Future Work**

Addition of a temporally consistent network would enable smooth detectionand more optimal than per-frame detection.Despite the achievements, the proposed approach can still be

improved and few directions have been listed here.

1. Occlusions lead to partial detection in videos due to high density of objects and

low angle of camera for monitoring video sequence. In the experiments, it has

been found that when occlusion occurs, object tracking suffers significantly. It

shows that proposed approach can be improved using more sophisticated and

robust object tracking method. An effective occlusion handling approach can be

targeted for further improvement of results.

2. Effective handling of issues like dynamic background and unexpected object

motion can also become subject of future work. Solutions based on deep leaning

techniques can be a potential candidate for integration in the proposed technique.

3. Increase in performance in presence of camera jitter and dynamic background is

also highly desired in multiple object tracking applications.

4. Feed from CCD cameras is of very poor quality and further advancement can be

in utilizing proposed approach on real time videos captured from CCD cameras.

5. The reliable detection of shadow is a challenging task as shadows have same

magnitude and movement pattern similar to foreground objects. The existence of

shadows causes distortion in shape of the object, object loss, merging of the

object. The proposed approach needs to be improved in this context.

6. Furthermore, hybrid techniques may be investigated for further improvements in

results and overcoming failures in complex environments.

The investigation process on soft computing based approaches for object detectionand tracking in videos is a going on process and will continue to evolve in near future.These techniques will be the pioneers for the development of robust and sophisticatedapplications in the future also.

**BIBLIOGRAPHY**

[1] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchiesfor accurate object detection and semantic segmentation. In The IEEE Conference onComputer Vision and Pattern Recognition (CVPR), 2014.

[2] Ross Girshick. Fast R-CNN. In International Conference on Computer Vision (ICCV),2015.

[3] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards realtime object detection with region proposal networks. In Advances in Neural InformationProcessing Systems (NIPS), 2015.

[4] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once:Unified, real-time object detection. In The IEEE Conference on Computer Vision andPattern Recognition (CVPR), 2016.

[5] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, ChengYang Fu, and Alexander C. Berg. SSD: Single shot multibox detector. In ECCV, 2016.

[6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scaleimage recognition. arXiv preprint arXiv:1409.1556, 2014.