

OBJECT DETECTION WITH HIGH-VISION

BY

FAHIMA YESMIN

ID: 171-15-8994

AND

ALAMGIR MD. SUWEB

ID: 171-15-9170

This Report is presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised by

Mr. Raja Tariqul Hasan Tusher

Senior Lecturer

Department of CSE

Daffodil International University

Co-Supervised by

Mr. Raja Tariqul Hasan Tusher

Senior Lecturer

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2021

APPROVAL

This Project/internship titled “Object Detection With High Vision”, submitted by Fahima yesmin and Alamgir md.Suweb, ID: 171-15-8994 and 171-15-9170 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 28-01-2021.

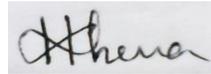
BOARD OF EXAMINERS



Dr. Touhid Bhuiyan
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Most. Hasna Hena
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Nusrat Jahan
Senior Lecturer

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Shamim H Ripon
Professor

Department of Computer Science and Engineering
East West University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr. Raja Tariqul Hasan Tusher, Senior Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma

Supervised By



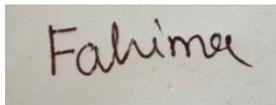
Mr. Raja Tariqul Hasan Tusher

Senior Lecturer

Department of CSE

Daffodil International University

Submitted by:

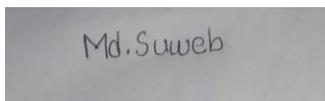


Fahima Yesmin

ID: 171-15-8994

Department of CSE

Daffodil International University



Alamgir Md. Suweb

ID: 171-15-9170

Department of CSE

Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project successfully.

The real spirit of achieving a goal is through the way of excellence and austere discipline. We would have never succeeded in completing our task without the cooperation, encouragement and help provided to us by various personalities.

We would like to express our deepest appreciation to our supervisor Mr. Raja Tariqul Hasan Tusher (Senior Lecturer), supervisor and Mr. Raja Tariqul Hasan Tusher co-supervisor, BSc in CSE Program, Daffodil International University. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project. We are especially grateful to our honorable teachers.

We would like to express our heartiest gratitude to Dr. Touhid Bhuiyan Head, Department of CSE, for his kind help to finish our project and we would also like to admit with much appreciation the crucial role of the staff of Daffodil International University (DIU), who gave me the permission to access all kind of library materials and equipment to gain knowledge and to clear out our understandings. We have to appreciate the guidance given by the other supervisors and lecturers who has helped us to clear our understanding and created a concern and importance of completing the project report carefully with maintain good knowledge and quality.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Machine learning using High Vision with detection is the branch of the science of computers and software systems which can recognize and get to know the images and scenes very well. Computer Vision is consisting of various things like as image recognition, image generation, object detection, image super-resolution and much more. Object detection is being widely used for detecting face, vehicle detection, pedestrian counting, web images, security systems and self-driving cars. In this project, we are working with Python Programming Language and it uses Machine Learning Libraries like TensorFlow, Keras, OpenCV and Numpy on backend. It implements the State-Of-The-Art Algorithms like YOLOv4, TinyYOLOv4, CDCL, MaskRCNN, and DeepLabv3+ to recognize objects. This project can detect the various objects in the input video and in the real time as well and can also provide an output video having the objects highlighted.

TABLE OF CONTENTS

CONTENTS	PAGE NO
Approval	i
Declaration	ii
Acknowledgement	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	1 - 5
1.1 Introduction	1
1.2 Motivation	2
1.3 Aims and objectives	2
1.3.1 Aims of the project	3
1.3.2 Objectives of the project	3
1.4 Feasibility study	3
1.5 Expected Outcome	5
1.6 Report Layout	5
CHAPTER 2: BACKGROUND	6 - 8
2.1 Related works	6
2.2 Comparitive study	6
2.3 Scope of project	7
2.4 Challanges	7
2.4.1 What do we think ?	7
2.4.2 Time	8

CHAPTER 3: REQUIREMENT SPECIFICATIONS	9 - 12
3.1 System requirements	9
3.1.1 Jupyter Notebook	9
3.1.2 Spyder IDE	9
3.1.3 PyCharm IDE	10
3.2 Steps to be followed	10
CHAPTER 4: ALGORITHMS	13 - 22
4.1 State of art	13
4.2 YOLO (You Only Look Once)	13
4.3 CDCL (Cross –Domain Complementary Learning)	20
4.4 Mask RCNN	20
4.5 Deep lab V3 plus	21
CHAPTER 5: IMPLEMENTATION AND TESTING	23 - 27
5.1 Testing Methodology	23
5.1.1 Functional Testing	23
5.1.2 Unit Testing	23
5.2 Output Samples	23
CHAPTER 6: CONCLUSION	28 - 29
6.1 Conclusion	28
6.1.1 Future Enhancement	28
REFERENCES	30

LIST OF TABLES

TABLES

Table 3.1: System Requirements

9

LIST OF FIGURES

FIGURES

Figure 4.2.1: Example 1

15

Figure 4.2.2: Example 2

16

Figure 4.2.3: Bounding Box

18

Figure 4.2.4: Area of IOU

18

Figure 4.2.5: Non max Suppression

19

Figure 4.4.1: Mask R-CNN

21

Figure 4.5.1: Architecture of Deep V3

22

Figure 5.2: Human pose detection

24

Figure 5.2.1: Cars Detection

25

Figure 5.2.2: Instance segmentation detection

25

Figure 5.2.3: Body parts detection

26

Figure 5.2.4: Human pose detection -2

26

Figure 5.2.5: Semantic segmentation detection

27

CHAPTER 1

INTRODUCTION

1.1 Introduction

A high Level python Library “High-Vision” can be defined as a Python Package which can be used to perform multiple complex tasks like Object Detection, Instance Segmentation, Semantic Segmentation, Human Pose Estimation and Human Part Segmentation. As the future of AI is changing by every minute there is lots of research in the field of Machine Learning, Deep Learning and Computer Vision and almost by every month researchers propose new methods and ideas to overcome the limitations of previous methods.

There are many people using libraries like TensorFlow, Keras, and PyTorch to implement these complex tasks, but that requires a lot of code to implement from scratch. So, to overcome this problem some developers have made high level libraries that can perform tasks like Object Detection and Image Segmentation, but the limitation of those libraries is that it can only perform one task or some libraries can perform two tasks, an example of this is a popular python library “ImageAI” which is a high level library, it requires just 6-9 lines of code, and it is used by thousands of people but this is limited to Object Detection and Image Prediction, it is built using TensorFlow and Keras on backend. The limitations of ImageAI are that it cannot perform tasks other than Object Detection and Image Prediction and it is not compatible with TensorFlow 2.

Normally, when someone want to use even a pre-implemented model like RetinaNet, YOLO etc. to perform Object Detection, Image Segmentation etc. they still have to write various lines of code to use that model in real life scenarios. It takes a considerable amount of time to write the code which in some cases can be a hectic task. So, to overcome these limitations we have implemented a python based High Level Library to perform:

- Object Detection
- Instance Segmentation
- Semantic Segmentation
- Human Pose Estimation

- Human Parts Segmentation

“High-Vision” can perform such complicated tasks with very simple 4-5 lines of code. Almost everyone can use this library even if they are not from Machine Learning background they can perform these tasks with no effort at all.

For Students and Developers it will be very easy to use Artificial Intelligence and they can even extend this “High-Vision” library to perform more complex tasks and projects like Vehicle Number Plate Detection and Recognition, Virtual Clothes Try On System etc. By using this exciting implementation, more people will be looking to use Artificial Intelligence to make other exciting projects. This Library will help Students and Developers to be more motivated towards AI, and it will enhance their learning and developing skills.

1.2 Motivation

We believe that High-Vision will be the easiest library for Students and Developers. Here we make High-Vision Library to perform these tasks with high accuracy and fast speed. This Library will make a huge difference in the motivation of Students, Developers and Online Machine Learning and Computer vision community. With its high Accuracy and fast speed it has enough potential to attract the community and with its easy to use interface it will be very easy for the community to apply AI to their use cases in real life.

1.3 Aims and Objectives

The objective of “High-Vision” is to improve the current AI for Students and Developers, there is still lots of research in this field and it is one of the hottest fields. Our Objective is to provide the latest and best methods to our community and even extend this implementation to perform a number of other tasks as well like 3D Object Detection, 3D Image Segmentation, Panoptic segmentation etc.

1.3.1 Aims of the project

Following are the main aims of this project:

- Exposure to Advance Artificial Intelligence and Computer Vision
- Make AI easy to Use
- Spread our knowledge to everyone
- Sharpen the coding skills
- Engage in the Learning and Developing Process
- Develop new skills
- Adaptive to Different Problems
- Willing to try New Methods and Techniques
- Team work

1.3.2 Objective of the project

The purpose of the project is to learn to program a computer to "understand" a scene or features in an image. Typical goals of this project include: The detection, segmentation, localisation, and recognition of certain objects in images (example: humans and cars).

1.4 Feasibility Study:

Feasibility means making test of a project that it will help us to make a decision whether it's viable to go through the project will able or not. And studying in feasibility studies in the project tell us it's development purpose. The famous layman's term explains that project test. If the project passes the test then it is viable for development.

Feasibility studies has 4 area:

1. Resources
2. Time
3. Finance
4. Technology

The Feasibility of “High-Vision” Library is checked using these four areas.

Technology:

The development of this project is proposed with advance technology. The Project is based on a Package or Library which is easy to use for everyone. This Project is developed using Python Programming Language and it uses Machine Learning Libraries like TensorFlow, Keras, OpenCV and Numpy on backend. It implements the State-Of-The-Art Algorithms like YOLOv4, TinyYOLOv4, CDCL, MaskRCNN, and DeepLabv3+.

Finance:

This Finance area is used to decide that it is feasibility to invest required amount of money or not. It is help us to save money waste or extra money. There should be no problem to invest money in project because all the money will use technology which is already open source from a number of developers and it extends those implementations and combines them in one easily accessible package. The hardware requirement for this project is not very high for standard use cases, it can easily be used on Windows operating system with normal specs like 4GB RAM and 2.3+ Processor.

Time:

Time is the most important factor in development of a project, this dimension measures whether it is worth spending time on the project with respect to its outcomes or not. Time plays very important role because project should be built in required time and it is built within a limited time period.

Resources:

This dimension checks how much and how many types of resources the project will need. For this particular Project, resources are not very high, with standard Machine Learning Libraries like TensorFlow, Keras, Scikit-learn, OpenCV and Numpy, these libraries are open source and

free to use for everyone and a Windows Operating System with Standard hardware specifications. With these resources this library can be easily used by anyone.

1.5 Expected Outcome

The main outcome of this project is easy to use interface, a better environment with higher accuracy and faster speed for a number of different tasks. User can input the path of an image and gets an output image with object detected and segmented whatever the task performed, similarly user can input the path of a video and gets an output video with detected objects, and it provides detection through live camera streams as well, user input a camera feed and it shows the detected objects in real time as well as saves the video for later use.

1.6 Report layout

Report layout describes a summary of all the chapters. A brief summary of all chapters is given below:

- Chapter 1: Describes an introduction of the High-Vision, Motivation, Aims and Objectives, Feasibility study and the Report layout.
- Chapter 2: Describes the background, the related works, Comparative Studies and Challenges of the High-Vision.
- Chapter 3: Describes system requirements, Jupyter Notebook, Spyder IDE and PyCharm IDE, Step to be followed.
- Chapter 4: Algorithms, State of art, YOLO Algorithms , CDCL Algorithms, Mask RCNN, Deep lab V3 Plus
- Chapter 5: Describes the Implementation and testing of High-Vision, Functional Testing, Unit testing and Outputs.
- Chapter 6: Conclusion, Future enhancement
- References

CHAPTER 2

BACKGROUND

2.1 Related Works

Open MMLab: provides different packages for different tasks like MMCV, MMDetection, MMDetection 3D, MM Editing, MM Segmentation, MMLAction2. It provides many algorithms for object detection, image segmentation. It even provides models for 2d detection as well as 3d detection. It provides models for action recognition in videos. Many People use these packages to perform the specified tasks. Open MMLab is built using PyTorch library provided by Facebook. It provides packages for image and videos editing purposes.

Detectron2: It is provided by Facebook and it is next generation API for Object Detection, Instance Segmentation, Panoptic Segmentation, Pose Estimation and Dense Pose. It is provided by Facebook AI Research Team and it provides many State-Of-the-Art Algorithms designed by Facebook AI Researchers and by others. It uses PyTorch on its backend. It is one of the most famous APIs out there for AI and Computer Vision Tasks.

TensorFlow Object Detection API: is provided by TensorFlow itself, and uses TensorFlow on its backend which is open sourced by Google in 2015 and it is the most famous and mostly used AI and Machine Learning Library among the community. It provides the solution for Object Detection and Image Segmentation and can be easily extended to many projects.

2.2 Comparative Study

“High-Vision” is very similar to above mentioned APIs, and it even extends further to implement the solution for other tasks like Human Pose Estimation and Human Part Segmentation. This library is similar in working and much more easier to use than above mentioned packages, above mentioned libraries use command line scripts to work and are much difficult to setup, while

“High-Vision” is very easy to install and use, it uses simple python scripts to generate quality results.

2.3 Scope of Project

As we consider “High-Vision” will be the easiest library to use and for multiple tasks, it is for students and developers but possibly for the users who use Windows Operating System, Python and some libraries mentioned earlier. So, they can use it easily, user will need to have an internet connection to download some files automatically, if user doesn't have those files. Answering the user error and unpleasant moment to make sure the stability of the library, proper error handling is done in code.

2.4 Challenges

With new Platforms, updating operating systems, updating backend libraries, maintaining an effective library or package across internet is never easy.

2.4.1 What do we think?

The main challenge we faced was about the setting up the different models, setting up the code to work independently of any other model and combined working of whole library. Handling of user entered options was also a very critical part, as it can cause the whole system to terminate so handling that part was also crucial and challenging, after that we have automated the part where user have to provide the weights file, considering that it will ease the user, we have uploaded the weights file on github and on instantiating the class it automatically downloads the weights of models based on which model user want to use. We have focused on the community which use TensorFlow and Keras with Python. In result, we got involved in deep mathematics behind those algorithms and different code implementation paradigms like Object Oriented Programming, Functional Programming, and we got involved in how to make the code implementation more efficient and stable.

We believe that after going through all the process from writing a script to developing a robust AI library, we learnt how to:

1. Implement an idea
2. Developing New Methods and Strategies
3. Research on the Idea
4. Not to be afraid of Changing the Code
5. Do whatever it takes to finish in time

We went through a lot of articles about AI and computer vision and different available techniques and models and what will be future of AI. So we decided to make a project on AI to take part in changing the future. We have learnt so much from this project the how an API or library is developed using AI from scratch and what difficulties it involves as expert's point of view.

2.4.2 Time:

At the last, we can call our biggest test in this project has been the main factor of time for us and we all are working hard whole day and concentrating close to it. To concur on arrangements and get together was perhaps the most testing portions of the work, as the actual task was an intriguing subject and it was enjoyable to take care of business, however we took in a considerable amount from this cycle to be very much organized and all around arranged.

CHAPTER 3

REQUIREMENT SPECIFICATIONS

3.1 System requirements

Table 3.1 : System requirements

Criterion	Description
OS Version	Preferred: Microsoft® Windows® 7/8/10 (32-bit or 64-bit) Other Operating Systems can also work e.g. (MacOS) and (Linux)
RAM	Minimum 4GB RAM, 8GB for fast speed
Disk Space	Minimum 2GB, preferred 4GB
Processor	Minimum: CPU core i3, Preferred: CPU core i7
Python Version	Minimum: Python 3.5, Preferred: Python 3.6/3.7

3.1.1 Jupyter Notebook:

The Jupyter notebook is a famous platform and web application where we can share document and create live code, narrative text, equation, visualizations of code and text. This platform also does data cleaning, statistical model, transforms the data, numerical, machine learning things and more.

3.1.2 Spyder IDE:

It is a scientific open-source environment where we can write Python and it is design for engineer, scientist and data analysis. And this platform is absolute free. It has some useful things

like analysis, debugging, advance in editing. This development tool use data exploration, interactive execution, beautiful visual with capabilities of scientific package.

3.1.3 PyCharm IDE:

JetBrains developed a hybrid platform for the IDE in Python. It is popular for using python development application. The Python IDE of PyCharm use in some famous organization like Facebooks, Amazon, Pinterest, Twitter. PyCharm can be run in the Linux, windows, Mac OS. It can help the programmer to develop software in short time because it contains useful package and thing which is requirement for programmers and developers. So, High Vision should be installed in the PyCharm.

3.2 Steps to be followed

1) Please download and install Python version 3.x

<https://python.org>

2) Please install the following as well using pip command:

TensorFlow:

It is a software library which is open source and it is famous for using the many programming task and data flow. It is used in machine learning field and neural networks because it has math and other important library. It is developed by Google. It is also use for research. It launch Apache License 2.0 on November 9th ,2015

The 1st version released February 11,2017. It is mainly dependent on Google Brain second generation system. It can be used both in CPU and GPU for multiple time. This platform is available Linux, mac os, Windows, Mobile computing system like android and IOS in the 64bit .

The TensorFlow architecture is easy to use for development purpose in desktops, clusters and mobile gadget also. It has useful stateful dataflow graph, data arrays which is related to tensors.

pip install TensorFlow command

1. **NumPy:** It is Python programming language library, metrics, multi-dimensional arrays. It is Numeric successor and invented by Jim Hugunin with several developers. 2005 Travis Olphant created NumPy. There are some features like computing Numarray into Numeric, with modification. NumPy is a software which is open source and it has many contributor.

pip install NumPy command

2. **SciPy:** contains various features like optimizations, linear algebra, integration and interpolation and much more. It is based on NumPy array feature and can be classified as a part of NumPy stack which can be used for the similar applications like MATLAB and Octave.

pip install SciPy command

3. **OpenCV:** It is developed by Intel and real time computer vision. It get support from Willow Garage after the ITEEZ. This has cross platform and any one can use it freely using the open-source license.

pip install opencv-python [command]

4. **Pillow:** This is free Python programming library and it is frewhich can be used to open, save and edit different format of images in windows or other operating systems.

pip install pillow [command]

5. **Matplotlib:** It is a plotting library of Python programming language and NumPy. It gives us an object oriented API to use general purpose graphical user interface toolkits such as Tkinter, wxPython or Qt.

pip install matplotlib [command]

- 6. h5py:** It includes some feature like Low Level and High level interface from some python library. The low level interface can be called as a complete cluster of HDF5 API and the high level interface uses python and NumPy concepts that can give or support access to HDF5 databases, groups and files.

pip install h5py [command]

- 7. Keras:** It is a neural network library which is available as open source and it is written in python. We can run this on the top of Tensor Fow or Microsoft Cognitive Toolkit as well. It is specially used to deal with deep neural networks and it is very used friendly and extensible.

pip install keras [command]

CHAPTER 4

ALGORITHMS

4.1 State of art

First of all we have to know what is state of art and why it is used in this project.

State of art is algorithms which is used in machine learning in CV(image classification, Object detection , Semantic segmentation) , NLP(Sentiment analysis, Language modeling, text classification, machine translation, question answering), RS(Recommended system) and SR(Speech recognition).

Here we used state of art for the object and image classification in our project.

We used YOLO(You only look once) algorithms, CDCL(Cross Domain Complementary), Mask RCNN, Deep lab V3plus.

We will discuss those algorithms briefly and why it is used in our project.

4.2 YOLO (You Only Look Once) Algorithms

When we see a object in our eyes and we get to know what is that object look like, what is the object is , also information and location about that object. But how this is happened all those stuffs? Actually when we see a object it captures the information and send it to the brain. Brain will decode that information and give the meaning full draw a inference about that object. Is it look simply this process? But actually, it is not. Brain did huge process beyond our thinking. But what about if we want to do same thing using machine? If we want that machine will recognize the object, classify the object, also detect the object information then we have to do some process to success it. It isn't easy but using some algorithms, trainig data set successfully, using data a lot we can successful can do it in the machine learning.

Yolo algorithms is a algorithm which is based on regression part of machine learning and it not only detect the interesting thing on the image but also it predicts the object class and use bounding box in whole image to specify the object location properly.

Before we discuss more about Yolo Algorithms, we have to know first the difference in among the classification, localization, detection.

Image classification is labelling the object in the image. It is supervised learning problem in machine learning. It is label or classify what is the object class in the image.

Localization: Localize the the particuler object using bounding box in the image. Example: In a image there is a car and it will locate the specific car using bound box in the whole image.

Object detection: Object detection is nothing but detect multipule object in a single photo with classification and localization. Where classification and localization only locate one specific object in a image.

But in our project we used classification, localization and object detection also.

When we give image to the machine it will go through some convention layer then it process in that deep neural network layer and give us output.

To understand the yolo algorithms we have to understand this algorithms aim to predict the class of the object and the bounding box will be specify the object location. To understand bounding box we have to know this thing:

1. Center of the box(B_x, B_y)
2. Width (BW)
3. Height (BH)
4. Values of corresponding class of the object(C)
5. PC (The probability present of the object in the image)

Classification:

Here we used some example:

Car= C_1

Motorbike=C2

Person=C3

$Y=(PC, B_x, B_y, B_h, B_w, C_1, C_2, C_3)$

Here is the photo:



Figure 4.2.1 (Example 1)

We can see clearly in the photo that there is an object present that is a car. P_c means the chance of the present of the object in the image. If an object is present in the image, the P_c value will be 1, and if no object is present in the image, then the P_c value will be zero.

So in that photo, a car means an object present in the image, so the P_c value will be 1.

What is the object is that also matters for C_1, C_2, C_3 . Here, a car is present, so C_1 will be 1, and the rest of C_2, C_3 will be zero.

So, $Y=(1, B_x, B_y, B_h, B_w, 1, 0, 0)$

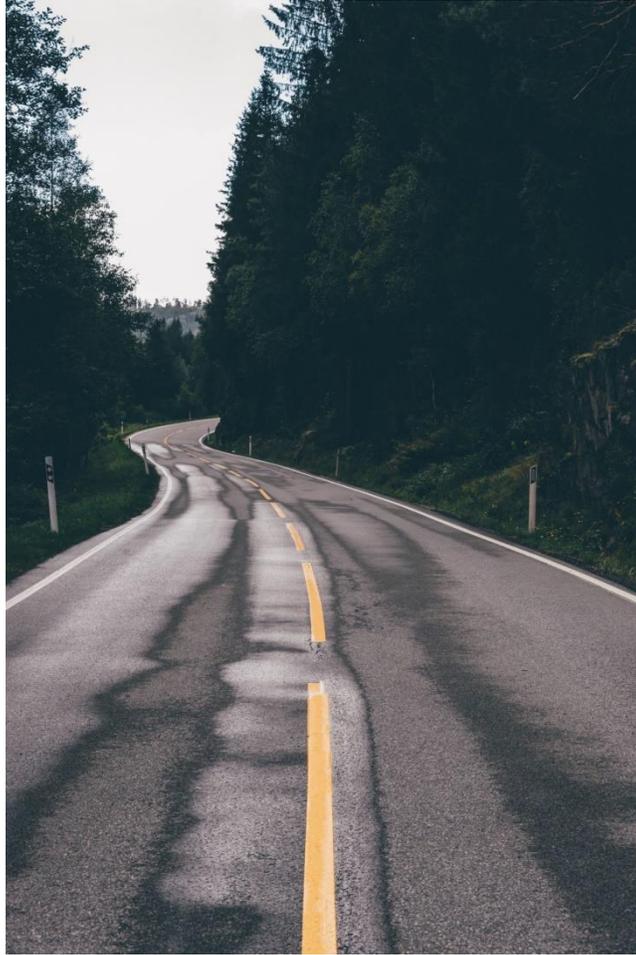


Figure 4.2.2 (Example 2)

In this photo we can see not object present. So the PC value will be Zero. And the BX,BY,BH,BW will be depending on object. So there is no object present so there is no value for BX,BY,BH,BW. Also have no value for C1, C2,C3

Loss Function: Loss function is finding the error algorithms output and given target. Loss function is calculated classification loss and localization loss.

$$\begin{aligned}
& \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
& + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
& + \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\
& + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\
& + \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
\end{aligned}$$

1. The 3 λ constant to take in the account more the one but aspect of the loss function. λ_{coord} means the highest order to more important in the first term.
2. *BB box predict* each of the grid in the cell and *CC predict* each the number of class. So YOLO is $S^2 \cdot S \cdot (B \cdot 5 + C)$
3. The real value box $C_{ij} \wedge C_{ij}$ is *intersection over union of the bounding box*.

Explain the general fact:

1. This term used to detect bad localization of center of cell.
2. It detects inaccurate height and weight of the bounding box
3. If confidence score is near the 00 then there is no object at all.

Bounding box :

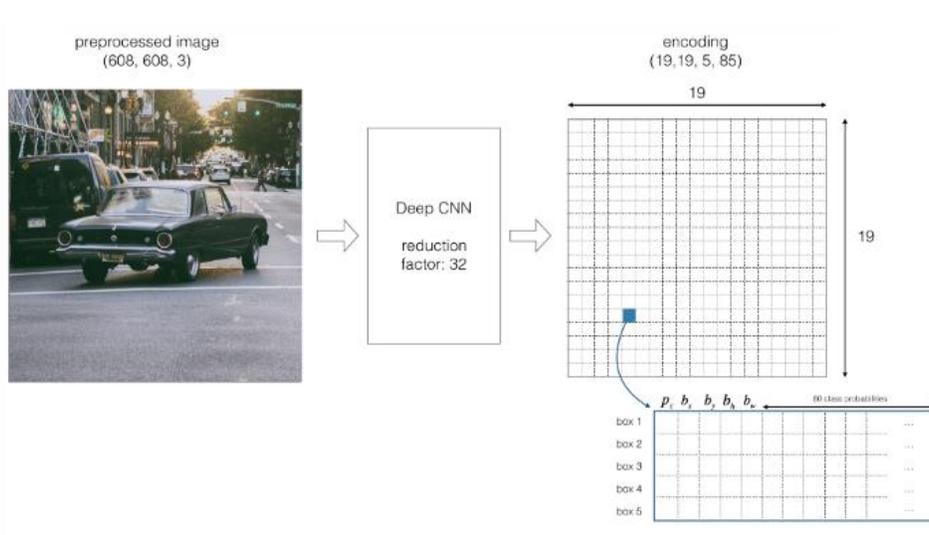


Figure 4.2.3 (Bounding Box)

In the bounding box of a image will 19X19 grid . Each cell has own PC,Bx,By,BH,Bw value. If one cell hasn't object present then pc value will be zero and rest of the other value will be zero. In the cell pc is have any object present or not, for bounding box(BX,BY,BH,BW) and for class labeling (C1,C2, C3).Each cell starting point is (0,0) and ending point (0,1). But we have always remembered that the object should be (Bx, Bh) <1 and (Bh, BW) >1

IOU: IOU means intersection over union. IOU we use when two boxes is in the one object. For calculating which box is correctly predict the object we calculate Area of intersection divided by area of union. Mainly it is counted overlap two boxes in one object.

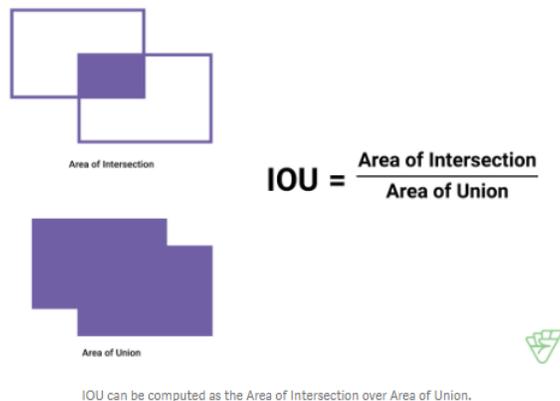


Figure 4.2.4 (Area of IOU)

Area of intersection is nothing but overlapping part of two boxes and area of union is total area both of the boxes.

If IOU is greater than and equal to 0.5 then we will consider it is a good prediction or bounding box is correct and if it is less than 0.5 then we will consider it is a bad prediction.

Non max suppression: First we have to predict the class probabilities then we should do non max suppression. It help us to get ride the not useful some anchor boxes and it depend the calculation of the class probabilities.

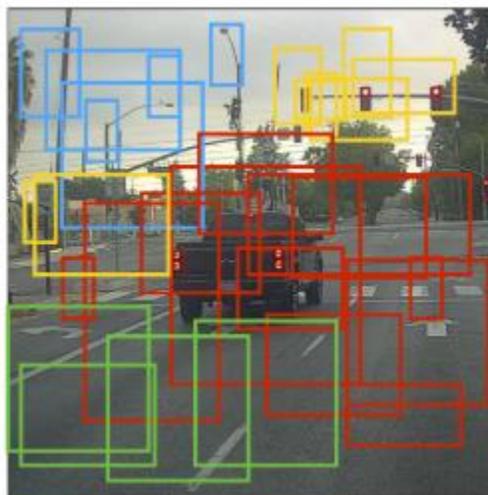


Figure 4.2.5 (Non max suppression)

But non max suppression need two things must that is

PC=highest

$IOU \geq 0.5$

Then we found high confident box among the weak boxes.

Anchor boxes: In the object detection there are large number of input image to decide the regions contains object and predicting of bounding box target more accurately we need anchor box. Because it helps to generates multiple bounding in different size and centring of the boxes. To localization and prediction of the object in YOLO algorithms is very much needed.

There are some facts of anchor box:

1. There present thousands of anchor box in the image.
2. Each anchor box try to predict offset which from the box of a candidate box
3. It is calculating the loss function depending on the ground truth
4. If the probability >0.5 then the prediction factor will be loss function

4.3 CDCL (Cross-Domain Complementary Learning)

CDCL algorithms mainly used to detect pose of multi-person part segmentation. In supervised learning with pixel of training data set to detect pose of multi person part segmentation is great achievement but labelling data of each pixel is very much expensive. So, to solve this problem we used synthetic data to avoid this problem.

.Requirements:

- 1.Synthetic data for labelling.
- 2.Domain adoption.
3. Multi task learning
4. Supervised and semi supervised part segmentation.

4.4 Mask RCNN

Mask RCNN is used for segmentation and object detection because it helps to generates the bounding box and segmentation mask for each object. It is used in keras, TensorFlow, python3.

Mask R-CNN

Mask R-CNN extends Faster R-CNN.

What's different in Mask R-CNN and Faster R-CNN?

1. Mask R-CNN is a additional branch of the segmentation detection of ROI.
2. Two outputs in Faster R-CNN but Mask R-CNN has three outputs.
3. Last object is object mask.

What's similar between Mask R-CNN and Faster R-CNN?

1. Mask R-CNN and Faster R-CNN are branch of regression because both use to classification and bounding box.
2. Both architecture use Resnet 101
3. Both use Region Proposal Network (RPN) to generate Region of Interests(RoI)

How does Mask R-CNN work?

Mask R-CNN model is divided into two parts

1. Region proposal network (RPN) to proposes candidate object bounding boxes.
2. Binary mask classifier to generate mask for every class

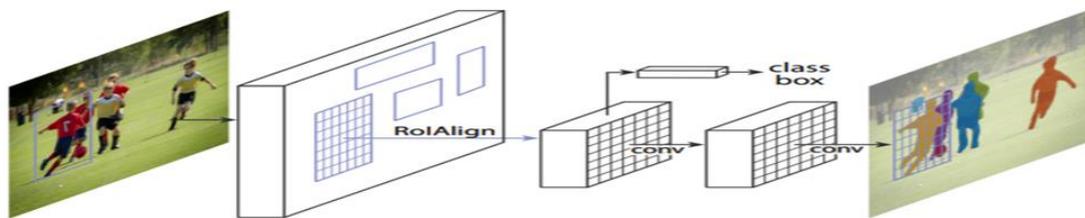


Figure 4.4.1 (Mask R-CNN)

3. Image is run through the CNN to generate the feature maps.
4. RPN(Region Proposal Network used by CNN to generate multiple ROI under binary classification.
5. The RoI Align network outputs multiple bounding boxes rather than a single definite one and warp them into a fixed dimension.
6. Warped features fed fully to make connection the layer which use to make classification using SoftMax of boundary box prediction for regression model
7. This Warped features again fee in the mask classifier which have two CNN output binary mask for each and every ROI

4.5 Deep lab V3 plus

For the semantic segmentation it has task to assign the every pixel label in the video or image. So we use the open source Deep Lab which is implanted in TensorFlow. It modelled by Google.

DeepLab-V3: Addition the deeplab3 to decoder the segmentation task.

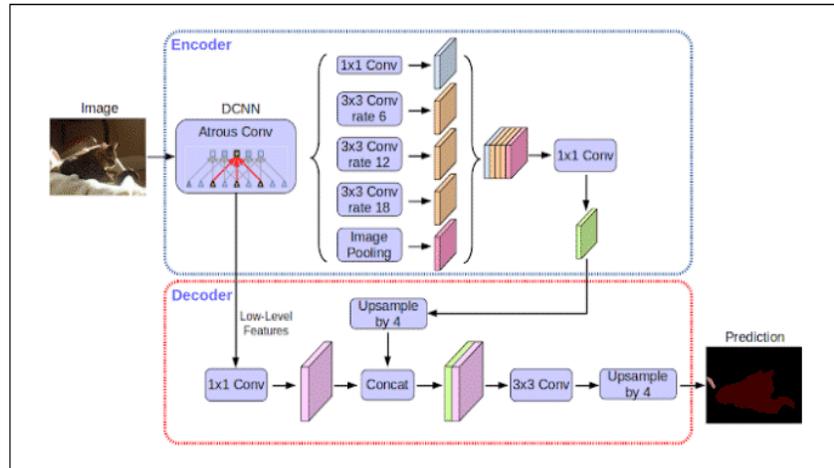


Figure 4.5.1 (Architecture of DeepLab-V3)

The architecture of the latest version of Deep Lab (DeepLab-V3) is composed of two steps:

- 1. Encoder:** In this encoder step we have pre trained CNN which extract information from the image and object location
- 2. Decoder:** The encoding steps the extracted information used to create the output size of the original input image

CHAPTER 5

IMPLEMENTATION AND TESTING

5.1 Testing Methodology

By and large, testing is utilized to discover framework mistakes. A task's test can be completed by analyzing codes, plan and execution of the entire framework. Testing is inescapable to improve the nature of the framework. Looking into and testing code is another fundamental of programming that is frequently neglected in venture improvement. Testing is an indispensable piece of the framework improvement measure. The main Standard for software Testing is contained in the ANSI/IEEE standard 829/1983- Standard for software testing Documentation. Some software testing may also be performed by CAST (Computer Aided Software Testing).

5.1.1 Functional Testing

In functional testing analyzer needs to approve the application to see that all predetermined necessities of the client's whatever we have said in supplemental restriction framework have been joined or not

- There are two categories of functional testing:
- Positive functional testing: Application function is tested with valid input and also verify the output accurately
- Negative functional testing: It exercising application function use the invalid inputs from some unexpected operation condition.

5.1.2 Unit Testing

It is used detailed design and implementing of the project. It finds out the rational defects in the project

5.2 Output Samples

It takes the path of an input image and generates an output image based on the algorithm applied on images like shown below. It can perform same tasks on videos as well as live camera streams.

Object detection: It detects object from video and image which is related computer vision and image processing. It is currently using in many field

Instance segmentation: Detect instance of each object from image instead categorizing each pixel.

Body part segmentation: .Detect each body part of a person in the image and video example legs, head, and more

Pose detection/estimation: Pose detection means it detect pose of a person or object in the video and images.

Semantic segmentation: r efers to the way toward connecting every pixel in a picture to a class mark. These names could incorporate an individual, vehicle, bloom, household item, and so forth, just to make reference to a couple. We can consider semantic division picture order at a pixel level.



Figure 5.2 (Human pose detection)

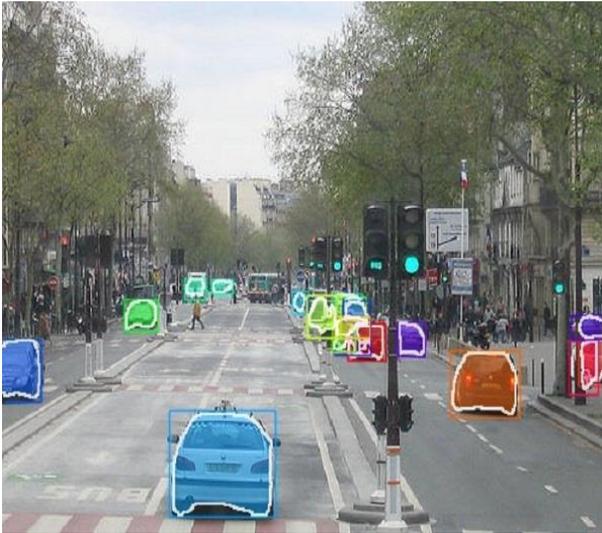


Figure 5.2.1 (Cars Detection)



Figure 5.2.2 (Instance segmentation detection)



Figure 5.2.3 (Body parts detection)



Figure 5.2.4 (Human pose detection - 2)



Figure 5.2.5 (Semantic Segmentation detection)

CHAPTER 6

CONCLUSION

6.1 Conclusion

By utilizing this proposal and dependent on test results we can distinguish protests all the more decisively and recognize the items separately, play out different complex errands like Object Detection, Instance Segmentation, Semantic Segmentation, Human Pose Estimation and Human Part Segmentation. This paper also provides experimental results on different methods for object detection and identification and compares each method for their efficiencies .

6.1.1 future enhancements

- 1) The features removed is adequate for perceiving the article and denoting the area of the item. x the proposed classifier can perceive the item in less computational expense.
- 2) The proposed worldwide component extraction requires less time, contrasted with the customary element extraction technique
- 3) The presentation of the SVM-kNN is more prominent and promising when contrasted and the BPN and SVM.
- 4) The exhibition of the One-against-One classifier is effective n number.
- 5) In particular, the commitments towards this exploration work are as per the following

As a scope for future enhancement,

A) Features either the neighbourhoods or worldwide utilized for acknowledgment can be expanded, to build the productivity of the recognition system.

B) Image can be included in the Geometric properties

C) For Night time visual following, night vision mode ought to be accessible as an inbuilt element in the CCTV camera.

D) With the help of our thesis, we are making an effort to develop an algorithm that can provide a good initiation for future application or can act as a base for future application.

In this Thesis, I have used my laptop's camera to do object detection and visual tracking. This concept can be used in the applications such as Intelligent Robots, vehicles that can guide themselves and to enhance the security systems to detect the different or doubtful behaviour as well as the detection of lethal weapons, Look at the doubtful and suspicious movement of terrorists and enemies on borders and can have night vision capabilities and many similar applications.

REFERENCES

- [1] Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images via a spa-based representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 26,1475–1490. doi:10.1109/TPAMI.2004.108.
- [2] Andreopoulos, A., and Tsotsos, J. K. (2013). 50 years of object recognition: directions forward. *Compute. Vis. Image Underst.* 117, 827–891. doi:10.1016/j.cviu.2013.04.005
- [3] Bourdev, L. D., and Malik, J. (2009). “Poselets: body part detectors trained using 3d human pose annotations,” in *IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 – October 4, 2009 (Kyoto: IEEE)*,1365–1372.
- [4] Erhan, D., Szegedy, C., Toshev, A., and Anguelov, D. (2014). “Scalable object detection using deep neural networks,” in *Computer Vision and Pattern Recognition Frontiers in Robotics and AI* www.frontiersin.org November 2015

Turnitin Originality Report

Processed on: 26-Jan-2021 16:08 +06
 ID: 1494621760
 Word Count: 6286
 Submitted: 1

OBJECT DETECTION WITH HIGH-VISION By Fahima Yesmin 171-15-8994

Similarity Index 22%	Similarity by Source Internet Sources: 20% Publications: 5% Student Papers: 17%
---	---

[exclude quoted](#) [exclude bibliography](#) [exclude small matches](#)

mode:

Change mode [print](#) [refresh](#) [download](#)

5% match (Internet from 17-Jul-2020) https://www.researchgate.net/publication/337464355 OBJECT DETECTION AND IDENTIFICATION A Project Repo	✕
3% match (student papers from 04-Nov-2019) Submitted to Daffodil International University on 2019-11-04	✕
3% match (Internet from 28-Aug-2020) https://towardsdatascience.com/computer-vision-instance-segmentation-with-mask-r-cnn-7983502fcad1?gi=c0b2a001df75	✕
2% match (student papers from 29-Oct-2019) Submitted to Manchester Metropolitan University on 2019-10-29	✕
2% match (student papers from 25-Jan-2021) Submitted to Daffodil International University on 2021-01-25	✕
2% match (Internet from 03-Jul-2020) http://dspace.daffodilvarsity.edu.bd:8080	✕
1% match (student papers from 05-Apr-2018) Submitted to Daffodil International University on 2018-04-05	✕
<1% match (Internet from 28-Nov-2005) http://computervision.wikicitities.com	✕
<1% match (student papers from 05-Apr-2018) Submitted to Daffodil International University on 2018-04-05	✕
<1% match (student papers from 12-Jan-2021) Submitted to Daffodil International University on 2021-01-12	✕
<1% match (Internet from 09-Sep-2017) http://uobrep.openrepository.com	✕
<1% match (student papers from 09-Jun-2020) Submitted to Galgotias University, Greater Noida on 2020-06-09	✕
<1% match (student papers from 22-Jun-2020) Submitted to K. J. Somaiya College of Engineering VidyaVihar, Mumbai on 2020-06-22	✕
<1% match (Internet from 15-Nov-2020) https://www.coursehero.com/file/56589261/vision-handoutpdf/	✕