## A

**MAJOR PROJECT REPORT**

**ON**

**Object Detection And Recognition Using Webcam With Voice Using YOLO Algorithm**

*A project report submitted to the*

***Jawaharlal Nehru Technological University***

*In partial fulfilment for the award*

***Bachelor of Technology***

***In***

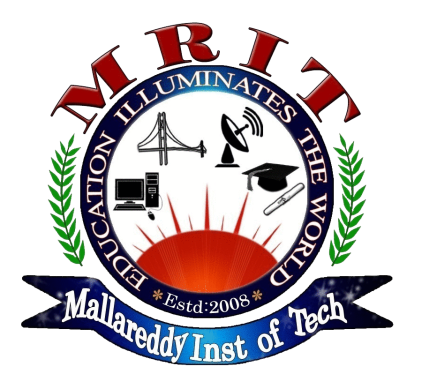
## COMPUTER SCIENCE AND ENGINEERING

***Submitted by***

### B. DEEKSHITH -19RJ1A0517

***Under the esteemed guidance of***

**Dr.KRN KIRAN KUMAR**



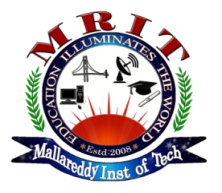
**MALLA REDDY INSTITUTE OF TECHNOLOGY**

### (Affiliated to JNTU, Hyderabad | Approved by AICTE, New Delhi )

### Accredited by NBA, Certificated by ISO 9001:2015

### Maisammaguda, Dhullapally, Via: Kompally, Hyderabad-500100

### 2019 - 2023

 **MALLA REDDY INSTITUTE OF TECHNOLOGY**

### (Sponsored by Malla Reddy Educational Society)

### Accredited by NBA, Certificated by ISO 9001:2015

### Approved by AICTE & Affiliated to JNTU, Hyderabad

Maisammaguda, Dhullapally Post, (Via: Kompally), Secunderabad - 500100.

**CERTIFICATE**

This is to certify that major project work entitled **“Object Detection And Recognition Using Webcam With Voice Using YOLO Algorithm”** is a bonafide work carried by **B. Deekshith (19RJ1A0517)** of **COMPUTER SCIENCE AND ENGINEERING DEPARTMENT** in **MALLA REDDY INSTITUTE OF TECHNOLOGY** and submitted to **JNT UNIVERSITY, Hyderabad** in the partial fulfillment of the requirement for the award of **BACHELOR OF TECHNOLOGY**.

|  |  |  |
| --- | --- | --- |
| **Project Guide** | **Project Coordinator** | **Head of Department** |
| **Dr.KRN KIRAN KUMAR** | **Ms.K.BHAVANI** | **Dr.KRN KIRAN KUMAR** |

**External Examiner**

**DECLARATION**

I hereby declare that the project entitle **“Object Detection And Recognition Using Webcam With Voice Using YOLO Algorithm”** submitted to **Malla Reddy Institute of Technology**, affiliated to Jawaharlal Nehru Technological University Hyderabad (**JNTUH**) for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** is a result of work done by me.

It is further declared that the project report or any part there of has not been previously submitted to any University or Institute for the award of degree.

### Deekshith - 19RJ1A0517

**ACKNOWLEDGEMENT**

I am happy to express my deep sense of gratitude to the **Principal** of the college **Dr. K. Srinivas Rao, Professor, Malla Reddy Institute of Technology** for having provided me the adequate facilities to pursue my project.

I would like to thank **Dr.KRN Kiran Kumar, Professor and Head, Department of Computer Science and Engineering, Malla Reddy Institute of Technology,** for having provided the freedom to use all the facilities available in the department, especially the laboratories and library.

I also grateful to my project coordinator, **Ms.K.Bhavani**, **Assistant Professor**, **Department of Computer Science and Engineering**, **Malla Reddy Institute of Technology**, for extending her support and assisting me throughout my project work.

I very grateful to my project guide, **Dr.KRN Kiran Kumar**, **Professor and Head, Department of Computer Science and Engineering**, **Malla Reddy Institute of Technology**, for his extensive patience and guidance throughout my project work.

I sincerely thank all the **Teaching and Non-teaching Staff** of the department of **Computer Science and Engineering**, also like to thank our **Classmates** for their timely suggestions, healthy criticism and motivation during the course of my project work. I thank my **Parents** who were the backbone behind my deeds.

Finally, I express my immense gratitude with pleasure to all **individuals** who have either directly or indirectly contributed to my need at right time for the development and success of my project work

# ABSTRACT

In this challenging evolution, the primary task in detecting the objects requires a computer vision that deals over indoor and outdoor classes. Over the past decades, this zeal requires more attentiveness. Previous implementation techniques involve in object detection with a strategy of single labelling. In this regard, a multi-label approach using machine learning and vision technologies, and accurate response can be acknowledged based on its accuracy and effectiveness. In the proposed work, we solve the existing system problem by using classification/clustering techniques that are used to reduce the recognize time of multi objects in less time with best time complexities. The model used to assist the visually impaired people can independently recognize objects which are near to them. The reverence, combined with the study, confounded the inception of these machine learning algorithms for visually impaired persons in assisting the accurate navigation, including indoor and outdoor circumstances.

Object recognition is an area of computer vision that looks for meaningful objects in images and movies (by creating a bounding box). Due to visual impairment, it restricts their movement in unfamiliar places. Here we will convert the image to text and then the text to voice in this project. Using this paper,we suggest an intelligent system that detects everyday common objects and produces voice feedback to inform them about the location of the object. First variations of YOLO (You Only Look Once) algorithm are compared and then the best one is used according to result we get it by training it on COCO dataset. After the object is detected, we calculate its spatial location then the voice output is generated by using text-to-speech(gTTS) module so that the visually impaired person will be to understand unfamiliar surroundings.

|  |  |  |
| --- | --- | --- |
|  | **INDEX** |  |
|  | **TABLE OF CONTENTS** | **PAGE NO** |
|  | **LIST OF FIGURES** | **VI** |
|  | **LIST OF SCREENSHOTS** | **VII** |
| **1** | **INTRODUCTION** | **1** |
| **2** | **LITERATURE SURVEY** | **3** |
|  | 2.1 WHAT IS YOLO?  2.2 SOFTWARE ENVIRONMENT  2.3 WHY TO CHOOSE PYTHON | 5  6  7 |
| **3** | **SYSTEM ANALYSIS**  3.1 EXISTING SYSTEM  3.2 PROPOSED SYSTEM | **11**  11  12 |
| **4** | **FEASIBILITY STUDY**  4.1 ECONOMICAL FEASIBILITY  4.2 TECHNICAL FEASIBILITY  4.3 SOCIAL FEASIBILITY | **13**  13  13  14 |
| **5** | **SYSTEM REQUIREMENTS** | **15** |
| **6** | **SYSTEM DESIGN**  6.1 WORKING OF YOLO ALGORITHM  6.2 INTRODUCTION TO UML  6.3 UML DIAGRAMS | **16**  16  21  23 |
| **7** | **IMPLEMENTATION**  7.1 METHODOLOGIES  7.2 MODULES  7.3 SAMPLE CODE  7.4 MAIN CODE | **29**  29  33  35  38 |
| **8** | **SYSTEM TESTING**  8.1 UNIT TESTING  8.2 INTEGRATING TESTING  8.3 FUNCTIONAL TEST  8.4 SYSTEM TEST  8.5 WHITE BOX TESTING  8.6 BLACK BOX TESTING | **41**  41  41  42  42  43  43 |
| **9** | **INPUT DESIGN AND OUTPUT DESIGN**  9.1 INPUT DESIGN  9.2 OUTPUT DESIGN | **44**  44  45 |
| **10** | **SCREENSHOTS** | **46** |
| **11** | **CONCLUSION** | **48** |
| **12** | **FUTURE WORK** | **49** |
| **13** | **BIBLIOGRAPHY** | **50** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  | **LIST OF FIGURES** |  |
| **FIGURE NO** | **FIGURE NAME** | **PAGE NO** |
| 2.1 | EVALUATION OF YOLO ALGORITHM | 06 |
| 6.1 | WORKING OF YOLO | 16 |
| 6.2 | EXAMPLE IMAGE WITH 3X3 GRIDS | 17 |
| 6.3 | ELEMENTS OF LABEL Y | 17 |
| 6.4 | BOUNDING BOX AND CLASS VALUES OF  GRID 1 | 18 |
| 6.5 | BOUNDING BOX AND CLASS VALUES OF  GRID 6 | 18 |
| 6.6 | AN EXAMPLE IMAGE FOR ANCHOR BOX | 19 |
| 6.7 | ANCHOR BOXES | 20 |
| 6.8 | ANCHOR BOX PREDICTION VALUES | 21 |
| 6.9 | USE CASE DIAGRAM | 24 |
| 6.10 | SEQUENCE DIAGRAM | 26 |
| 6.11 | DATA-FLOW DIAGRAM | 28 |
| 7.1 | YOLO ARCHITECTURE | 31 |
| 7.2 | PROCESS MODEL | 33 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  | **LIST OF SCREENSHOTS** |  |
| **Screenshot No** | **Screenshot Name** | **Page No** |
| 10.1 | RUN.BAT FILE | 46 |
| 10.2 | OPENING THE RUN.BAT FILE | 46 |
| 10.3 | DETECTING BOTTLE AND PERSON | 47 |
| 10.4 | DETECTING BOOK | 47 |

# INTRODUCTION

People with visual impairment face many difficulties in their lives. Recent data released by the World Health Organization (WHO) in 2020 shows that approximately 2.2 billion people worldwide suffer from imaginative and predictive disabilities. Detecting and identifying common devices in the area seems like a heroic project for the visually impaired. This is because they rely on different human beings; blind people rely on them and their sense of touch and smell to detect things, which can be particularly debilitating and sometimes dangerous. White cane is the most popular blind navigation tool. It is further enhanced by including ultrasonic and IR sensors to remove obstructions inside the visually impaired person's space and detect vibrations or sound. Although this method proved useful for the movement of blind people, it provided very little information about the environment. To better understand the user environment, it is important to know the objections and categories, along with identification and audio feedback.

Neural networks, especially CNN, have shown encouraging outcomes, especially in image object detection, type, and popularity. Most of the authors used the feed-forward neural network to offer sound cues about the purchased product. Implemented within the phone-based, real-time, complete disability detection and class gadget detection method, including the removal and monitoring the hobby element through a multi-scale Lucas-Kenned set of principles. Includes estimates of historical movement using homo graphic variations and a collection. The clustering method then uses classification, using the Histogram of Oriented Gradients (HOG) signifier in the bag of visual words (BoVW). Finally, a review of Electronic Travel Aids (ETA) constructed for visually compromised navigational aids is provided. Different ETAs, strengths, and weaknesses are mentioned and compared in terms of function. It also highlights that no state-of-the-art device has all the necessary features and that no generation needs to update the cane but completes it through appropriate warnings and comments.

First, youth is the future of all nations. we can build a more liveable place. As technology advances, traditional methods are being replaced by modern methods. But this evolution is what this world wants. Much research has been done to solve the inconveniences of daily life. Humans can identify objects and its locations in the matter of milliseconds. One of the biggest inconveniences that blind people experience in their daily lives is finding information about objects around them and mobility issues. Simple items are difficult for them to recognize and identifying objects of the same dimension and size is difficult. In this article, we propose an object detection using YOLO (You only looks once) algorithm to detect objects and gTTS module to convert the detected object with its location to voice as feedback. We

use YOLOv4-tiny which is a lightweight and faster object detection based on YOLO-v4. Predicting classes and boundary boxes are done quickly for whole frame, so that predictions can be seen in the whole context of frame. After the object is detected and identified it then finds its spatial location in each frame.It then will then be recastinto audio segment using Google-Text-To-Speech(gTTS) module. The YOLO-v4-tiny network is trained on COCO dataset containing approximately 123,287 hand labeled images and is classified into 80 different categories.

# LITERATURE SURVEY

Few products are being developed using in-deep learning and image processing to help the visually impaired. Below are some of the more common related jobs in this section. An Android mobile application was developed for the visually impaired, where live feeds from the camera were fed as input to a trained system learning model. The version uses image processing and item detection to capture instance tools that have been used to train versions on object classification. The software aims to help blind people move around without difficulty and alert them if they encounter any obstacles. A similar prototype tool has been developed, equipped with imaginative and present binocular sensors. These binocular sensors capture images at a set frequency from which the information maxima are determined by stereo image quality assessment (SIQA). These snapshots are sent to the cloud for similar processing. Detection and automatic results can be provided using a convolutional neural network-based primarily on large records. Through image analysis, cloud computing will return the requested information to the user so that the user can make cheap decisions about further actions. A mobile alert system has been developed for the visually impaired with a state-of-the-art and creative work-tracking infrastructure that uses web centers in the network to capture it. Mobile devices and marked points are determined by GPS function. The purpose of the cellular device is to enable visually impaired people to navigate distances according to the marked points. This research aims to explore and estimate the distance and relative functions of blind people around the devices around them, especially parked bicycles.

The proposed tool uses a single-shot Multi-Box Detector (SSD) to locate parked bicycles using various learning algorithms. It is based on the principle of a triangle comparing the original motorcycle with the motorcycle image to estimate the distance of 2 to 5 meters of the motorcycle using a hollow chamber device. The study provides relevant information, including length and grip protection on flat floors. Pipelines combine multi-factor cloud instances, tabletop aircraft detection, device detection, and full version estimation through a robust estimator. In this bodywork, the benefits of recent deep insights (e.g., RCNN, YOLO) are used, which will be a green path for the detection project, while the geometry-based strategy estimates the full 3D version. ۔ Relevant heritage objects no longer need to be separated (or separated) from the surrounding landscape for this study.

Mona and Riyadh proposed their dissertation in 2020 on "Retina Net Assessment for Internal Object Discovery for Assisted Navigation for the Blind and Visually Impaired." In these articles, the feature of Computer Vision is to accurately locate objects inside the house. People with visual impairments can be helped by visiting the features of the CNN Framework.

To identify specific devices first, we want to find the pixels found in the images. Unfortunately, if the conditions of the lighting fixture are incorrect, it isn't easy to locate and locate the fixture with high accuracy. Next, the algorithm needs to extract the photographic capabilities with a selective spectrum to detect internal objects, which can be implemented through the retina net. Enabling small object detection networks through the Region Proposal Network (RPN) means sub-sampling to capture photographic records. With 152 samples, the resort achieved an average accuracy of 83.1%, and Dennis Net achieved an average accuracy of 79.8% with 121 samples.

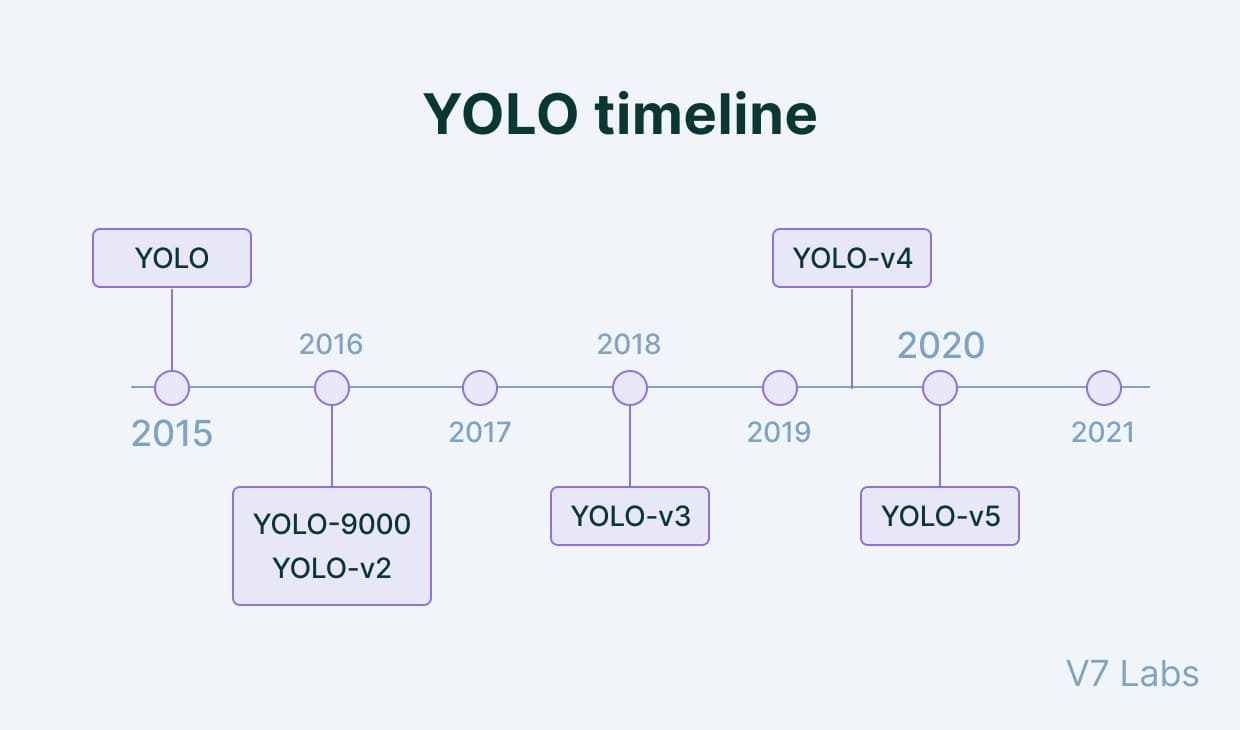
Han Hu and Jiayuan proposed their paintings in 2018 on "Relationship Networks for Object Detection." Based on the relationship models, the work allocated the same amount of work, considering its characteristics. It eliminates duplication and achieves the exact requirements. Since objects are connected on a 2D scale, use elements instead of phrases. In addition, versions are categorized as extras below the geometric and actual weight. Xiangrong and Allen proposed their paintings in 2005 on "A one-time waterfall for real-time object detection: with applications for the visually impaired." In this work, the main objective is to identify the complexities of time and its accuracy based on it. Various tests are performed by the module capture system that detects textual content in images that can advance for the visually impaired. The beauty of the model can be measured by F.P. And F.N. quotes. The functionality of the algorithm's choice can be implemented through a set of classifiers and schooling images. It can use smart telescopic systems for people with vision problems. In micro-screen images, the image forcefully represents itself, leaving behind the positive aspects of the image.

Alice Tang and Xiwan proposed their work in 2018 on "Serial Cerebral Angiography Automated Registration: A Comparative Review." Over time, performed fine-tuning to identify and reverse the disorder, primarily based on this particular work in the medical field. First, consider its effectiveness and accuracy. Magnetic resonance imaging (MRI) and computed tomography (C.T.) scans are analyzed in an image processing algorithm that can better evaluate than the DSA. Although the DSA establishes a diagnosis of several neurovascular conditions used during surgical procedures, based on these considerations, we can conclude that the system was designed based on patients diagnosed with ischemic stroke. Is. Wei and Xia proposed their articles in 2015 on "HCP: A Flexible Convolutional Neural Network (CNN) Framework for Multi-Label Image Classification." In this article, a version of CNN offers the best performance for image classification with a single tag. Due to its complexity, multiple labeling is an open plan for educational image design. The same image is taken as input to make assumptions, and it is shared with CNN to rank people through maximum polling. Single colors evaluate image hypotheses that different groups can represent.

The extraction technique produces predictive effects that can use to maximize polling. When reviewing I-FT and HCP models, the HCP version improves device performance by 5.7%. Rum and Assam launched their paintings in 2018 in a combination of "RGB (red-green-blue) image and related photography intensity (RGBD) images" of a complex network of complex values ​​for the blind. This work uses multiple versions for visually impaired people to face objects more spectacularly in an interior area. This model has more than one label at a time. CVNN and multi-label techniques combine the image with labels that instantly match objects classes.

**2.1 What is YOLO?**

You Only Look Once (YOLO) proposes using an end-to-end [neural network](https://www.v7labs.com/blog/neural-network-architectures-guide) that makes predictions of bounding boxes and class probabilities all at once. It differs from the approach taken by previous object detection algorithms, which repurposed classifiers to perform detection. Following a fundamentally different approach to object detection, YOLO achieved state-of-the-art results, beating other real-time object detection algorithms by a large margin.While algorithms like Faster [RCNN](https://www.v7labs.com/blog/recurrent-neural-networks-guide) work by detecting possible regions of interest using the Region Proposal Network and then performing recognition on those regions separately, YOLO performs all of its predictions with the help of a single fully connected layer. Methods that use Region Proposal Networks perform multiple iterations for the same image, while YOLO gets away with a single iteration. Several new versions of the same model have been proposed since the initial release of YOLO in 2015, each building on and improving its predecessor. Here's a timeline showcasing YOLO's development in recent years.



**Fig 2.1 Evaluation of YOLO Algorithm**

### 2.2 SOFTWARE ENVIRONMENT

Python is a high-level, interpreted scripting language developed in the late 1980s by Guido van Rossum at the National Research Institute for Mathematics and Computer Science in the Netherlands. The initial version was published at the alt. Sources newsgroup in 1991, and version 1.0 was released in 1994.

Python 2.0 was released in 2000, and the 2.x versions were the prevalent releases until December 2008. At that time, the development team made the decision to release version 3.0, which contained a few relatively small but significant changes that were not backward compatible with the 2.x versions. Python 2 and 3 are very similar, and some features of Python 3 have been backported to Python 2. But in general, they remain not quite compatible.

Both Python 2 and 3 have continued to be maintained and developed, with periodic release updates for both. As of this writing, the most recent versions available are 2.7.15 and 3.6.5. However, an official End of Life date of January 1, 2020 has been established for Python 2, after which time it will no longer be maintained. If you are a newcomer to Python, it is recommended that you focus on Python 3, as this tutorial will do.

### 2.3 WHY TO CHOOSE PYTHON

If you are going to write programs, there are literally dozens of commonly used languages to choose from. Why chooses Python? Here are some of the features that make Python an appealing choice.

### 2.3.1 Python is Popular

Python has been growing in popularity over the last few years. The 2018 Stack Overflow Developer Survey ranked Python as the 7th most popular and the number one most wanted technology of the year. World-class software development countries around the globe use Python every single day.

According to research by Dice Python is also one of the hottest skills to have and the most popular programming language in the world based on the Popularity of Programming Language Index.

Due to the popularity and widespread use of Python as a programming language, Python developers are sought after and paid well. If you’d like to dig deeper into Python salary statistics and job opportunities, you can do so here.

### 2.3.2 Python is interpreted

Many languages are compiled, meaning the source code you create needs to be translated into machine code, the language of your computer’s processor, before it can be run. Programs written in an interpreted language are passed straight to an interpreter that runs them directly.This makes for a quicker development cycle because you just type in your code and run it, without the intermediate compilation step.

One potential downside to interpreted languages is execution speed. Programs that are compiled into the native language of the computer processor tend to run more quickly than interpreted programs. For some applications that are particularly computationally intensive, like graphics processing or intense number crunching, this can be limiting.

In practice, however, for most programs, the difference in execution speed is measured in milliseconds, or seconds at most, and not appreciably noticeable to a human user. The expediency of coding in an interpreted language is typically worth it for most applications.

### 2.3.3 Python is Free

The Python interpreter is developed under an OSI-approved open-source license, making it free to install, use, and distribute, even for commercial purposes. A version of the interpreter is available for virtually any platform there is, including all flavours of Unix, Windows, macOS, smartphones and tablets, and probably anything else you ever heard of. A version even exists for the half dozen people remaining who use OS/2.

### 2.3.4 Python is Portable

Because Python code is interpreted and not compiled into native machine instructions, code written for one platform will work on any other platform that has the Python interpreter installed. (This is true of any interpreted language, not just Python.)

### 2.3.5 Python is Simple

As programming languages go, Python is relatively uncluttered, and the developers have deliberately kept it that way. A rough estimate of the complexity of a language can be gleaned from the number of keywords or reserved words in the language. These are words that are reserved for Python 3 has 33 keywords, and Python 2 has 31. By contrast, C++ has 62, Java has 53, and Visual Basic has more than 120, though these latter examples probably vary somewhat by implementation or dialect. Python code has a simple and clean structure that is easy to learn and easy to read. In fact, as you will see, the language definition enforces code structure that is easy to read.

But it’s not that simple for all its syntactical simplicity, Python supports most constructs that would be expected in a very high-level language, including complex dynamic data types, structured and functional programming, and object-oriented programming or GUI programming. Additionally, a very extensive library of classes and functions is available that provides capability well beyond what is built into the language, such as database manipulation special meaning by the compiler or interpreter because they designate specific built-in functionality of the language. Python accomplishes what many programming languages don’t: the language itself is simply designed, but it is very versatile in terms of what you can accomplish with it.

### 2.3.6 Conclusion

This section gave an overview of the Pythonprogramming language, including:

* A brief history of the development of Python
* Some reasons why you might select Python as your language of choice.

Python is a great option, whether you are a beginning programmer looking to learn the basics, an experienced programmer designing a large application, or anywhere in between. The basics of Python are easily grasped, and yet its capabilities are vast. Proceed to the next section to learn how to acquire and install Python on your computer.

Python is an open-source programming language that was made to be easy-to-read and powerful. A Dutch programmer named Guido van Rossum made Python in 1991. He named it after the television show Monty Python's Flying Circus. Many Python examples and tutorials include jokes from the show. Python is an interpreted language. Interpreted languages do not need to be compiled to run. A program called an interpreter runs Python code on almost any kind of computer. This means that a programmer can change the code and quickly see the results. This also means Python is slower than a compiled language like C, because it is not running machine code directly. Python is a good programming language for beginners. It is a high-level language, which means a programmer can focus on what to do instead of how to do it. Writing programs in Python takes less time than in some other languages. Python drew inspiration from other programming languages like C, C++, Java, Perl, and Lisp. Python has a very easy-to-read syntax. Some of Python's syntax comes from C, because that is the language that Python was written in. But Python uses whitespace to delimit code: spaces or tabs are used to organize code into groups. This is different from C. In C, there is a semicolon at the end of each line and curly braces ({}) are used to group code. Using whitespace to delimit code makes Python a very easy-to-read language.

### 2.3.7 Python use [change / change source]

Python is used by hundreds of thousands of programmers and is used in many

places. Sometimes only Python code is used for a program, but most of the time it is used to do simple jobs while another programming language is used to do more complicated tasks. Its standard library is made up of many functions that come with Python when it is installed. On the Internet there are many other libraries available that make it possible for the Python language to do more things. These libraries make it a powerful language; it can do many different things.

Some things that Python is often used for are:

* Programming
* Game programming

# SYSTEM ANALYSIS

**3.1 EXISTING SYSTEM**

Low vision or blindness is one of people's top ten most common disabilities. Unfortunately, India is home to one of the most visually impaired populations in the world. This paper provided a unique framework to help visually impaired people find and recognize objects and is a great way to move around freely and keep track of their surroundings. The report uses a broadcast to provide insight into unpaired snapshot detection (SSD) approaches to locate and classify objects that can be examined when human faces and banknotes are detected. , Inception was completed using the v3 form. The entire SSD detector is based on the modified PASCAL VOC 2007 dataset, which introduces a completely new feature for coin detection. In addition, the discrete models of Inception v3 can recognize human faces and banknotes, making the framework scalable and adaptable to a person's ability. Ultimately, it can give the frame a visually impaired position within the sound layout. That changed the Delivery Forex Elegance (MAP) SSD independent detector rating to 67.8%, and the accuracy and forex credentials of the Inception v3 male and female models were 92.5% and 90.2%, respectively.

### 3.1.1 DISADVANTAGE OF EXISTING SYSTEM

* While YOLO is known for its high accuracy in object detection, it may not perform as well for smaller objects. This is because the algorithm focuses on detecting objects at a certain scale and may struggle to detect objects that are too small or too large.
* The YOLO algorithm may have difficulty with objects that are partially occluded or hidden from view. This can result in false negatives or incorrect predictions.
* The YOLO algorithm has been trained on a limited set of object classes, which may not include all the objects that need to be detected for a particular application. This can limit its usefulness for certain use cases.

### 

### 3.2 PROPOSED SYSTEM

In this mission, we combine SSD300 (Single Shot Detector) with the Starter version to search and evaluate currency banknotes. SSD 21 can recognize lessons, but it will not recognize currency bills, so we have introduced a broader layer for finding currency bills, but its accuracy is not sufficient. We removed the capabilities from the SSD and then returned to training with INCEPTIONV3. , Which can achieve accuracy. More than 97%. To find the currency, we used the INDIAN OLD NOTES dataset because the new Forex Note dataset is unavailable, so we train the SSD and start with the old Forex note, which this new model Can usually find 21 classes, of which 1 is more excellent. Forex, so now SSD General Configuration 22 can detect and identify.

### 3.2.1 ADVANTAGES OF PROPOSED SYSTEM:

* One of the main advantages of the YOLO algorithm is its ability to perform object detection in real-time. This makes it well-suited for applications that require fast and accurate detection, such as autonomous driving and surveillance systems.
* The YOLO algorithm is known for its high accuracy in object detection, particularly for small objects. Its ability to detect objects at different scales and aspect ratios also contributes to its accuracy.
* YOLO uses a single neural network for both object detection and classification, which makes it a lightweight and efficient solution that can run on low-power devices.

# FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

* ECONOMICAL FEASIBILITY
* TECHNICAL FEASIBILITY
* SOCIAL FEASIBILITY

### ECONOMICAL FEASIBILITY

### 

### This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

### 4.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

### 4.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

# SYSTEM REQUIREMENTS

## 5.1 Hardware Requirements

* System : Pentium IV 2.4 GHz
* Hard Disk : 256 GB.
* Monitor : 15 VGA Colour.
* Mouse : Logitech.
* Ram : 512 MB.

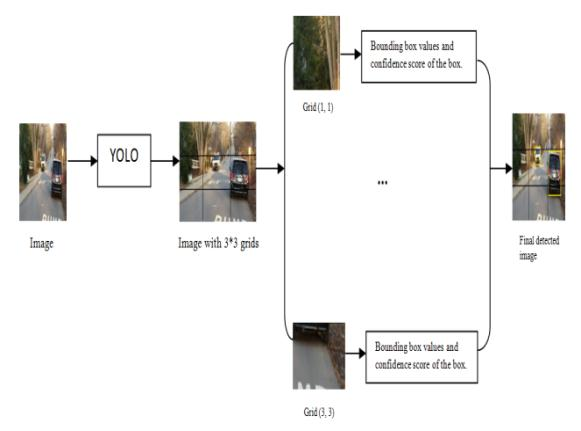
## 5.2 Software Requirements

* Operating system : Windows 7 Professional.
* Coding Language : python 3.7

# SYSTEM DESIGN

**6.1 WORKING OF YOLO ALGORITHM**

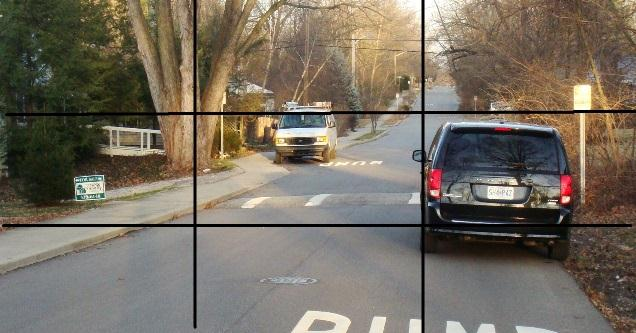
First, an image is taken and YOLO algorithm is applied. In our example, the image is divided as grids of 3x3 matrixes. We can divide the image into any number grids, depending on the complexity of the image. Once the image is divided, each grid undergoes classification and localization of the object. The objectness or the confidence score of each grid is found. If there is no proper object found in the grid, then the objectness and bounding box value of the grid will be zero or if there found an object in the grid then the objectness will be 1 and the bounding box value will be its corresponding bounding values of the found object. The bounding box prediction is explained as follows. Also, Anchor boxes are used to increase the accuracy of object detection which also explained below in detail.



**Fig 6.1 working of YOLO**

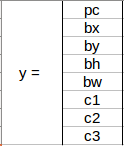
**6.1.1 Bounding box predictions**

YOLO algorithm is used for predicting the accurate bounding boxes from the image. The image divides into S x S grids by predicting the bounding boxes for each grid and class probabilities. Both image classification and object localization techniques are applied for each grid of the image and each grid is assigned with a label. Then the algorithm checks each grid separately and marks the label which has an object in it and also marks its bounding boxes. The labels of the gird without object are marked as zero.



**Fig 6.2 Example image with 3X3 grids**

Consider the above example, an image is taken and it is divided in the form of 3 x 3 matrixes. Each grid is labelled and each grid undergoes both image classification and objects localization techniques. The label is considered as Y. Y consists of 8 values.

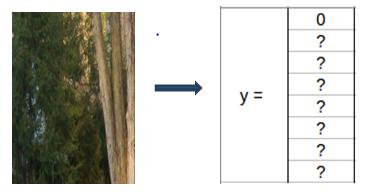


**Fig 6.3 Elements of label Y**

Pc – Represents whether an object is present in the grid or not. If present pc=1 else 0.

bx, by, bh, bw – are the bounding boxes of the objects (if present).

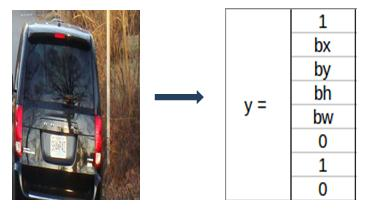
c1, c2, c3 – are the classes. If the object is a car then c1 and c3 will be 0 and c2 will be 1. In our example image, the first grid contains no proper object. So it is represented as,



**Fig 6.4 Bounding Box and Class Values of Grid 1**

In this grid, there exists no proper object so the pc value is 0.

And rest of the values are doesn’t matter because there exist no object. So, it is represented as ?. Consider a grid with the presence of an object. Both 5th and 6th grid of the image contains an object. Let’ consider the 6th grid, it is represented as.



**Fig 6.5 Bounding Box And Class Values of grid 6**

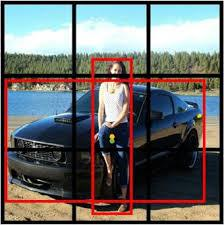
In this table, 1 represents the presence of an object. And bx, by, bh, bw are the bounding boxes of the object in the 6th grid. And the object in that grid is a car so the classes are (0,1,0). The matrix form of Y in this is Y=3x3x8. For the 5th grid also the matrix will be little similar with different bounding boxes by depending on the objects position in the corresponding grid. If two or more grids contain the same object then the

center point of the object is found and the grid which has that point is taken. For this, to get the accurate detection of the object we can use to methods. They are Intersection over Union and Non-Max Suppression. In IoU, it will takes the actual and predicted bounding box value and calculates the IoU of two boxes by using the formulae, IoU = Area of Intersection / Area of Union. If the value of IoU is more than or equal to our threshold value (0.5) then it's a good prediction. The threshold value is just an assuming value. We can also take greater threshold value to increase the accuracy or for better prediction of the object. The othermethod is Non-max suppression, in this, the high probability boxes are taken and the boxes with highIoU are suppressed. Repeat this until a box is selected and consider that as the bounding box for that object.

**6.1.2 Accuracy Improvement**

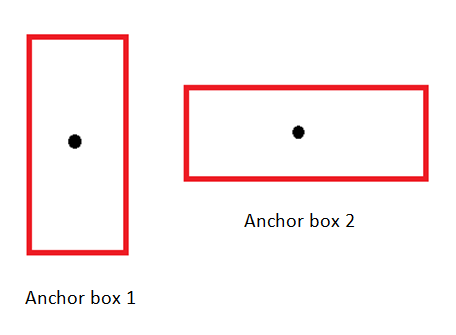
**6.1.2.1 Anchor Box:**

By using Bounding boxes for object detection, only one object can be identified by a grid. So, for detecting more than one object we go for Anchor box.



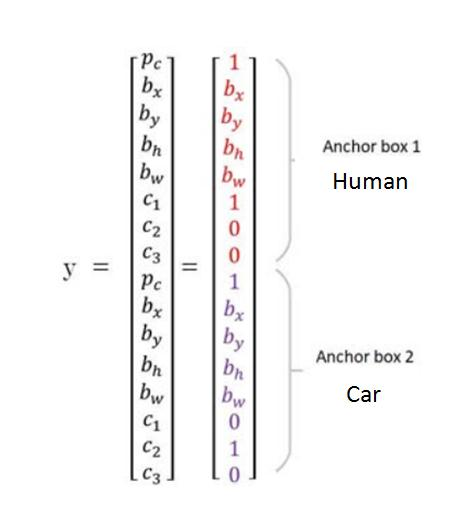
**Fig 6.6 An Example Image for Anchor Box**

Consider the above picture, in that both the human and the car’s midpoint come under the same grid cell. For this case, we use the anchor box method. The red color grid cells are the two anchor boxes for those objects. Any number of anchor boxes can be used for a single image to detect multiple objects. In our case, we have taken two anchor boxes.



**Fig 6.7 Anchor boxes**

The above figure represents the anchor box of the image we considered. The vertical anchor box is for the human and the horizontal one is the anchor box of the car. In this type of overlapping object detection, the label Y contains 16 values i.e, the values of both anchor boxes.



**Fig 6.8:Anchor box prediction values**

Pc in both the anchor box represents the presence of the object. bx, by, bh, bw in both the anchor box represents their corresponding bounding box values. The value of the class in anchor box 1 is (1, 0, 0) because the detected object is a human. In the case of anchor box 2, the detected object is a car so the class value is (0, 1, 0). In this case, the matrix form of Y will be Y= 3x3x16 or Y= 3x3x2x8. Because of two anchor box, it is 2x8.

**6.2 Introduction to UML**

Integrated Language Modeling enables the software engineer to construct an analytical model that employs a modelling model guided by a set of syntactic, semantic, and pragmatic rules. The UML system is depicted using five separate concepts that explain the system from a very diverse approach. Each perspective is specified by a series of diagrams, such as:

**6.2.1 Buyer Copy representation**

The Above way illustrates structure from a buyer angle. This analytical presentation specifies the handling situation in distinction to that perspective with respect to buyers.

**6.2.2 Structure Copy representation**

The Above copy represents not only infobot also performance can be obtained within schema. This Copy representation is a figure of vertical structures.

**6.2.3 Behavior Copy representation**

This represents observable fluctuations as chunk of system, reflecting its collective interact ion between the different skeletal fundamentals defined in buyer copy along with the aspect of the structural figure.

**6.2.4 Model View Performance**

This figure represents the, order as well as behavior which means a part to the plan are presented which are structured earlier.

**6.2.5 Viewing the Environmental Model**

This aspect shows, Skeletal as well as observable features of that region where schema needs to be applied are shown. Integrated Modeling Language (UML) was established to provide deep understanding regarding standard sight articulations that is recognizable, and mathematically rich in planning, constructing, as well as the usage of sophisticated program structures in a dual way that is skeletally as well as morally. UML had many software development operations, just like action train in production. It can be comparable with plans that are adaptable in various disciplines as well as incorporates several sorts of drawings. In compilation, UML diagrams tells boundaries, skeletal, along with observable of a structure as well as its contents. UML nay a programming articulation but also has some stuff which may be helpful in composing cipher in multiple articulations with the help of UML blueprints. UML had plainspoken link with thing-balanced test as well as plan.

### 

### 6.3 UML DIAGRAMS

Figuring the schema, is biggest crucial component is to abduct aggressive behavior. In precise, the aggressive act indicates that acts of in which schema that works / functions.

### 6.3.1 Use Case Diagram

Therefore, stable act alone may nay enough for constructing schema model instead aggressive act can be much significant than stable act. In the UML the 5 charts accessible for depicting the aggressive environment as well as the consumption blueprint is one amongst them. At present we need to mention that use case diagram may naturally grew strong there should be domestic and private variables to create a link. Inner along with outer ones can be called actors. The use case blueprints consequently comprise the participants, the use cases as well as the interactions. Blueprints can be helpful in developing a sub-structure / system model for the function.

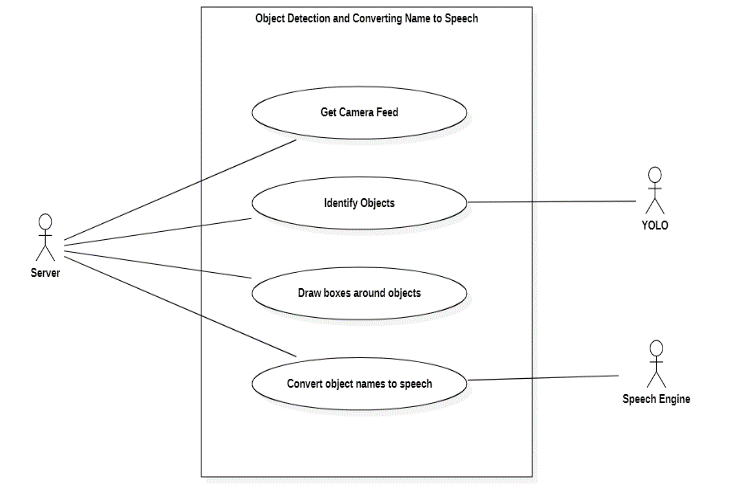
One-use case graphic catches specific schema performance. Therefore model all system numbers of utilised case drawings used. Character images are used to collect system requirements that incorporate inside and outside impacts. The above criteria were general design requirements. Therefore at the time of system study for compiling our performance the operating circumstances are altered and the characteristics detected. Sharply, the aims for employing case drawings is listed under:

a. For collecting the system demands.

b. For getting info regarding external representation about the system.

c. Analyze inside and outside factors that influences structures.

d. Perform connections between players' needs.



**Fig 6.9 Use Case Diagram**

**6.3.2 Sequence Diagram**

Sequential diagrams illustrate interactions between classes according to the exchange of messages across time. Also called event drawings. Sequential diagram is an excellent technique to illustrate and assess diverse working time scenarios. This may assist forecast how the schema may operate as well as highlight the authority the domain behave in the new modeling schema.

Objective of a sequence diagram is describing sequence of events, which may produce the intended result. Focus more on the way messages look than on the message per session. However, most of the sequence diagrams will explain about what messages are transmitted and the order in which they normally occur.

Sequence diagram is more often worn to communicate with diagram. Communication diagram - Interaction diagram can be used for demonstrating interaction behavior of schema. Because anticipating interoperability in the structure may be daunting endeavor, you will be employing multiple forms for communication diagrams for

catching different looks and characteristics of interoperability in the schema. Sequential blueprints - A sequence blueprint normally displays the communication amongst items in steady manner that is how this interaction takes volume. They might cause phrases occasion drawings or accident scenes that refer a sequence drawing. Sequential blueprints illustrate that the items for a system to function and in what sequence. Blueprints are extensively adopted by entrepreneurs as well as program builders to describe but also to comprehend the demands of new and current systems.

**6.3.2.1 Roles of Class or Participants**

The roles of the class specify how something would behave in background. Causing the UML thing mark for denoting domain lists, but cannot account item components.

**6.3.2.2 Activation or Activity Occurrence**

The start boxes reflect the time something you need to finish a job. If anything is busy executing the procedure or waiting for a reply message, put a tiny grey rectangle vertically in its health line.

**6.3.2.3 Messages**

Messages are arrows that indicate the relationship between things. Use half-arrow lines to illustrate inconsistent messages. Synchronized communications are delivered from something that would nay hold feedback in distinction to receiver previously maintaining with their responsibilities.

**6.3.2.4 Lifestyles**

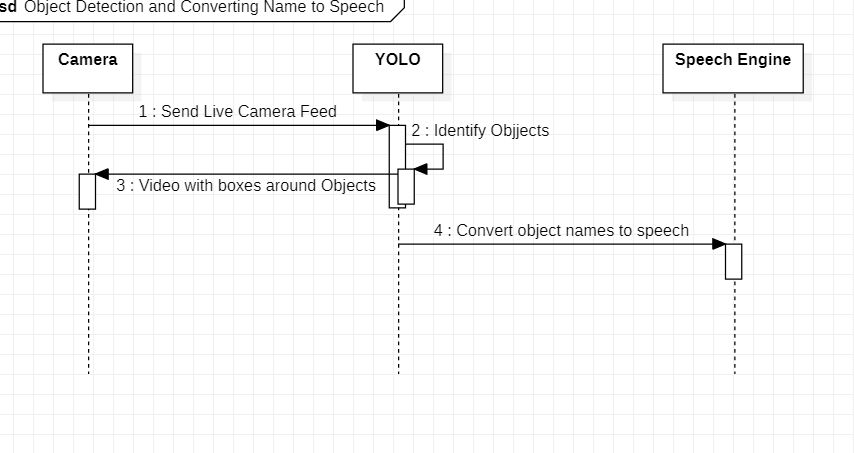
Lines of life are lines with a straight dash indicating the presence of an object over time.

**6.3.2.5 Destroying Things**

Items can be completed in advance using the "<< destroy >>" arrow pointing to X. This thing is taken out of memory. When the life line of that object ends, you can place an X at the end of its life line to indicate the event of destruction.

**6.3.2.6 Loops**

Repeating inwards a sequence blueprint can be represented like rectangle. Locate exit loop action in lower left corner of the square brackets []. Guard When you model the interaction of an object, there will be times when the condition must be met in order for the message to be sent to the object. Security guards are the conditions that need to be applied to all UML drawings to control the flow.

 **Fig 6.10 sequence diagram**

**6.3.3 Class Diagram**

Class drawings are the basic building component of all object-focused techniques. Class drawing might cause for depicting domains, likeness, link, relationships, as well as interactions. UML can be assessed for domain drawings. As the class is an OOP-based application structure, therefore as a class diagram it has the proper schema for depicting our domains, heritage, connections, as well as complex else OOPs behave in background. They outline many sorts of items and the interactions that exist between them.

The major objective of utilizing class drawings is:

1. This is the only UML that can appropriately represent the many components of the OOPs approach.

2. Appropriate planning as well as application inquiry endure rapid and efficient results.

3. This would be the foundation of distribution and section drawing. Each class is represented by a rectangle component comprising three word parts, characteristics and

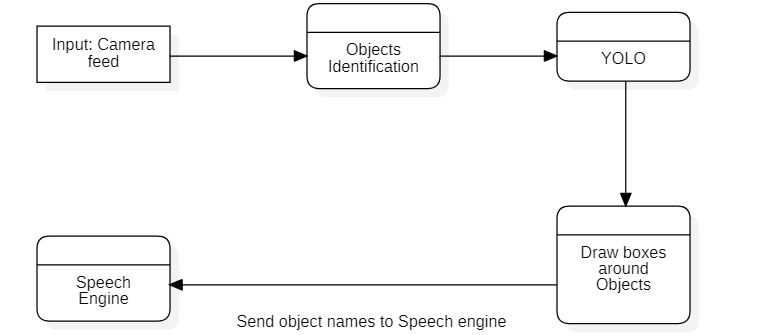
functionality.

Class drawing is a dry drawing. Represents a solitary representation of operation. Domain sketching would not only be adapted to view, understand, but also to express diverse system characteristics but also to generate useable software code software. The Domain blueprint explains characteristics as well as functioning of the class as well as the challenges laid forth in the system. Class drawings are commonly utilized in representation of competing systems because of their solitary UML blueprints, which those are adapted precisely for thing focused articulations. The Domain blueprints details an assemblage of domains, relationships, institutions, associations, along with pressure. Simply termed as architecture.

**6.3.3.1 Statistical Train Diagram**

A Statistical train diagram is a path for exhibiting data train with an action or structure. STD also adds advice on the output and input of every business and the action as such. The Statistical train Diagrams never dominates flow – nay prearrangement rules and nay convolutions. Certain data-planted functions can be represented by a flowchart. We can present Numerous ideas regarding the show casing Statistical train Diagrams. In every info stream, notably particular storage location is prerequisite in the process. A Classy Exemplification of the action can be made in a distinct Statistical train Diagrams, where it divides this action into smaller stages. Statistical train Diagrams is an apparatus for systematic inquiry and replication of data. While employing UML, a function blueprint usually plays the act of Statistical train Diagrams. A Certain type of info train system is ground-based Statistical train structure. Statistical train Diagrams are considered just as distorted Petri networks, since the places in aforesaid grids correspond to the memory of the data memory. Similarly, conversion semantics from Petri networks along with Statistics flow functions and functions from Statistical Train diagram should be grant equal .

STD is a system replicaing tool, the most popular and important presentations in data flow comparisons. STD allows us to capture the system as a network of active processes, connected to each other by “pipes” and “catch tanks” of data. It is a systematic, graphic representation of external organizations, logical storage, data sinks and data flow in the system. You can also call STD as follows: bubble chart, bubble diagram, actionreplica, and flow function diagram.



**Fig 6.11 Data-flow Diagram**

# IMPLEMENTATION

## 7.1 METHODOLOGIES

An in-depth, novel framework for the visually impaired using an ancient fusion of parallel CNNs that outperforms more sophisticated techniques for the popularity of hobbies. Both Google Net and CNN's Alex Net complement each other to identify unique features of the same magnificence. As a result, each of them is fed an input video feed and a support vector. magnificence beauty ratings are combined using a machine (MVS). Another proposed new method using CNN consists of a recurrent neural community (RNN) and a SoftMax classifier for detail detection and shade thresholding for hue, saturation, and intensity (HSI) of color reputation. . An approach that blends innovative and predictable wearable strategies with the profound talent of the visually impaired. The device uses regression-based mechanisms to scan items without preferred dates, handles sudden camera movements, and uses You Only Look Once (YOLO) to perceive the item.

A mobile application designed for blind men and women. You can draw in different ways: online and offline, based on the community connection between people. The online mode uses the fastest RCNN to make predictions in strong situations and YOLO for instant results. However, using the Haar and Histogram of Gradient (HOG) functions in a distinct identity module serves this purpose in the offline mode. CNN is designed to use the ImageNet dataset for the reliability of pre-qualified articles. A new DLSNF (usually a deep learning-based sensory navigation framework) based on the YOLO framework is proposed to develop a sensory navigation device based on the NVIDIA Jetson TX2. SqeezeNet, a moderate CNN pre-qualified model, scored higher and lower arithmetic delays. Squeeze Net was developed by changing the weight of the last convolution layer, replacing the corrected linear unit (ReLU) as an activation function with LeakyReLU, and adding a batch normalization layer.

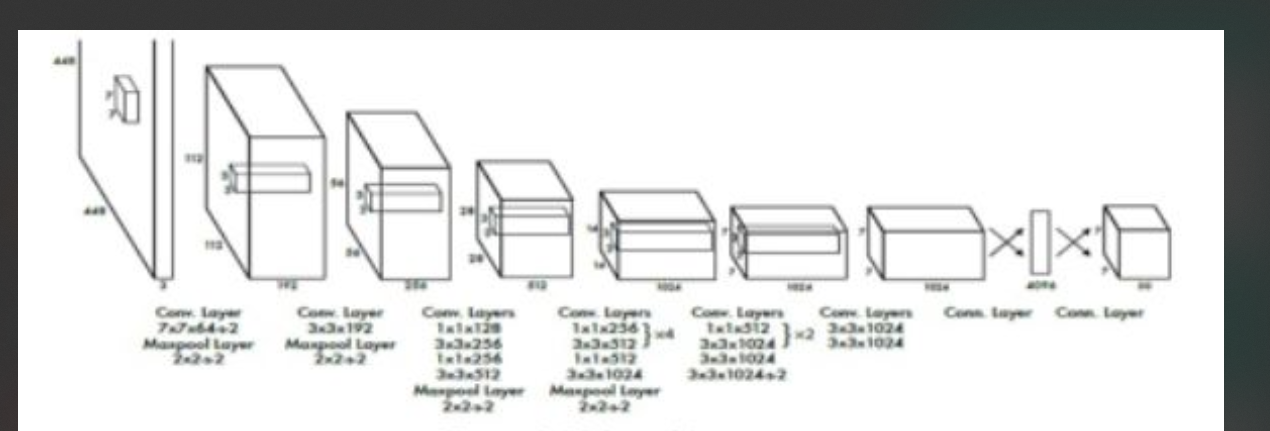
**7.1.1 Object Recognition**

Though similar approaches to object identification, object discovery and object recognition operate differently. Even though both are widely used for images and video. Object detection is considered a subset of object recognition in processing. Object detection and recognition are often employed in a wide range of sectors, from personal security to workplace productivity. It is used for autonomous driving systems, machine inspection, surveillance, security, and image retrieval, among other computer vision applications. In general, the text to speech conversion capability cannot be recommended for non-operating system devices. Thus, Android- or iPhone-based smartphones are the most popular choice among smart phone users who are blind or visually impaired****.****

Object discovery is the phenomenon of finding instances of items in both still images and films that contain the objects in question. Bounding boxes and information about the identified items' locations inside the frame are highlighted. Technology related to both image processing and computer-aided vision is object detection. It classifies and identifies a wide range of things from movies and digital photos, including people, animals, and cars. Multiple things in a video or digital picture can be swiftly classified using object detection. Although object detection has been around for a while, it is currently more prevalent than ever across a variety of sectors. The object detecting system has been put into action using a variety of techniques.

* + - 1. **YOLO Algorithm**

The YOLO algorithm was initially proposed by Joseph Redmon and his colleagues. In2015, he released a paper on YOLO under the heading "You Only Look Once" Real-Time recognition, and it became immensely successful right away. CNN is followed by YOLO. When making predictions, the algorithm only "looks once" at the image since there is only one propagation that occurs throughout the neural network. Comparedto other methods of object identification, the YOLO model is the fastest and most effective. The main benefit of YOLO is its quickness. There are 45 frames per second in this. The model is constructed in a concise manner to acquaint its network with an abstract description of things.  The primary goal of object detection is to identify one or more specific things in audio visual or digital pictures. Contrarily, object class recognition classifies items into a certain category or class. Every thing has unique qualities that make it easier to distinguish it from other objects in movies or pictures. Additionally, it sets them apart from other classes. Object detection is the process of finding and defining objects, such as people, animals, objects, cars, and so on. Combining the YouOnly Look Once (YOLO) architectural algorithm with the COCO dataset results in a quick and effective deep learning technique for object recognition.



**Fig 7. 1 YOLO Architecture**

YOLO is made for comprehensive image processing and steadily raises the effectiveness of object detection. Frame identification is seen as a regression issue in this situation. In order to quietly store specialised information about groups and their looks, YOLO employs the entire background throughout training and testing periods while the networked is concentrated on recent images. To continuously predict all bouncing boxes throughout all groups for a picture, it uses features from the full image. The method divides the informational image into a SxS example. The matrix cell can identify the point and choose the certainty scores for those containers when the focal point of an object falls within a network cell.

**7.1.2 Estimation of Image Position**

We must generate a bounding box for each identified object in order to approximate the position of the image. Using the specific bounding box's height and width in relation to the image frame. To estimate the position of an object within a bounding box, 5 values are used. The position of the item is shown by the first 4 values:  bx, by, bw, bh, BC the fifth value, specifies how much of a box an object occupies.

BC=pr\*IOU

Where IOU = Intersection over Union

Pr   = Object existing in box

This estimates the predictability and likelihood of the box containing an object of any class. If there is no object in the box, BC=0, otherwise BC=1.

**7.1.3 Voice Generation**

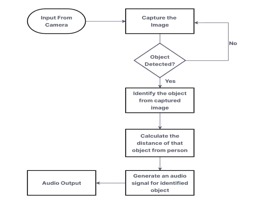
When the system locates the desired object, voice guiding is a feature that offers information to specific users, such as those who are blind, in a convenient manner. It is crucial to alert the blind person heading in the way of the presence of an object when it has been discovered.PYTTSX3 is an essential part of voice generation module. Text to speech conversion can be done using the Python library Pyttsx3. Python version 2 and 3 are both compatible with this package. A straightforward tool for text to speech conversion is Pyttsx3. We also used Google Text to Speech (GTTS) for voice alerts. Google Text to speech contain many inbuilt English accents for the users who are from different parts of the world. It is very easy to use, it converts the text into audio which can be saved as a mp3 file It also supports many regional languages which is also useful for those who do not able to understand English.

**7.1.3.1 Dataset**

Here we used COCO dataset. COCO stands for common object in context. The COCO dataset contains challenging, high-quality visual datasets for computer vision. The images in the dataset were gathered from commonplace locations that provided specific information. In real-world situations, multiple items or things may be contained within the same frame, and each one needs to be distinguished as a distinct object and properly segmented. The identification and segmentation of the objects visible in the photographs are contained in the COCO dataset. We used this information to develop our item recognition and detection technology for persons with disabilities. This dataset contains approximately 90 objects or items.

**7.1.3.2 Process Model**

The working of this system is represented in the below process model. Input is taken from the user's camera to capture the images. The system checks if any objects are detected in the image. If an object is detected, the system identifies the object. Then the system calculates the distance of the object from a person. Based on the calculated distance, the system generates an audio output.



**Fig 7.2 Process Model**

## 7.2 MODULES

**7.2.1 Python**

It has a large number of libraries for applications such as Scikit-learn for machine learning, OpenCV for computer vision, TensorFlow for neural networks, etc.

**7.2.2 OpenCV**

Techniques for Object Recognition in Images and Multi Object Detection and segmentation is the most significant and testing central undertaking of Computer vision. It is a basic part in numerous applications, for example, image search, scene understanding, and so far. However, it is as yet an open issue because of the assortment and multifaceted nature of item classes and foundations. The most effortless approach to identify and fragment an item from a picture is the shading based techniques. The term and the foundation ought to have a critical shading distinction so as to effectively portion objects utilizing shading-based strategies.

**7.2.3 YOLO**

Provides a framework that allows detection of objects in near real time speeds. For deployment in a mobile device we are using Tiny YOLO, which is a lightweight YOLO framework for mobile and edge devices.

**7.2.4 TensorFlow**

An open source software library and framework used for neural network and machine learning applications. It has a flexible architecture and can be deployed on servers, laptops, mobile phones, edge devices, etc.

**7.2.5** **TensorFlow Lite**

TensorFlow for mobile devices has a separate smaller and lighter framework that utilises the less powerful computing power of smartphones and edge devices called TensorFlow Lite. This smaller framework allows for easy deployment of machine learning applications on the smaller and less powerful devices.

**7.2.6 Keras**

Keras is a deep neural network library that runs on top of other frameworks such as TensorFlow. It is user friendly and allows for the training of machine learning models with relative ease using neural networks. Keras contains a host of add-on features for neural networks such as layers, optimizers, activation functions, etc. making it a useful library for many applications.

**7.2.7 Numpy**

YOLO typically requires images to be pre-processed before feeding them into the network. NumPy can be used to resize, normalize, and manipulate image data, such as converting images from different formats to arrays, applying transformations, or adjusting pixel values. NumPy provides a wide range of array operations that can be useful in YOLO. For example, YOLO involves calculations like anchor box adjustment, non-maximum suppression (NMS), or bounding box manipulation. NumPy's array operations can efficiently handle these computations.

Once the YOLO algorithm has made predictions, post-processing steps are necessary to interpret the results. NumPy can be used to process and manipulate the predicted bounding box coordinates, class probabilities, and confidence scores to obtain the final detection results.NumPy can be utilized for visualizing the results of the YOLO algorithm. By manipulating the predicted bounding box coordinates and overlaying them on the original image, NumPy can assist in generating visually appealing and informative object detection output.

**7.2.8 Playsound**

The playsound module is a Python library that provides a simple interface to play audio files. It is commonly used for basic audio playback in Python applications. For example, you can use it to play a sound effect or a music file.

**7.2.9 GTTS**

The gtts (Google Text-to-Speech) library is primarily used for converting text into speech or audio. While it is not directly related to the YOLO (You Only Look Once) algorithm, it can be used in combination with YOLO to generate spoken descriptions of detected objects.

**7.3 Sample code:**

import numpy as np

import argparse

import cv2 as cv

import subprocess

import time

import os

def detectObject(CNNnet, total\_layer\_names, image\_height, image\_width, image, name\_colors, class\_labels, Boundingboxes=None, confidence\_value=None, class\_ids=None, ids=None, detect=True):

if detect:

blob\_object=cv.dnn.blobFromImage(image,1/255.0,(416,416),swapRB=True, crop=False)

CNNnet.setInput(blob\_object)

cnn\_outs\_layer = CNNnet.forward(total\_layer\_names)

Boundingboxes,confidence\_value,class\_ids=listBoundingBoxes(cnn\_outs\_lay er, image\_height, image\_width, 0.5)

ids = cv.dnn.NMSBoxes(Boundingboxes, confidence\_value, 0.5, 0.3)

if Boundingboxes is None or confidence\_value is None or ids is None or class\_ids is None:

raise '[ERROR] unable to draw boxes.'

image, cls = labelsBoundingBoxes(image, Boundingboxes, confidence\_value, class\_ids, ids, name\_colors, class\_labels)

return image, cls, Boundingboxes, confidence\_value, class\_ids, ids

def labelsBoundingBoxes(image, Boundingbox, conf\_thr, classID, ids, color\_names, predicted\_labels):

cls = []

if len(ids) > 0:

for i in ids.flatten():

# draw boxes

xx, yy = Boundingbox[i][0], Boundingbox[i][1]

width, height = Boundingbox[i][2], Boundingbox[i][3]

class\_color = [int(color) for color in color\_names[classID[i]]]

cv.rectangle(image, (xx, yy), (xx+width, yy+height), class\_color, 2)

text\_label = "{}: {:4f}".format(predicted\_labels[classID[i]], conf\_thr[i])

cv.putText(image, text\_label, (xx, yy-5), cv.FONT\_HERSHEY\_SIMPLEX, 0.5, class\_color, 2)

cls.append(predicted\_labels[classID[i]])

return image, cls

def listBoundingBoxes(image, image\_height, image\_width, threshold\_conf):

box\_array = []

confidence\_array = []

class\_ids\_array = []

for img in image:

for obj\_detection in img:

detection\_scores = obj\_detection[5:]

class\_id = np.argmax(detection\_scores)

confidence\_value = detection\_scores[class\_id]

if confidence\_value > threshold\_conf:

Boundbox = obj\_detection[0:4] \* np.array([image\_width, image\_height, image\_width, image\_height])

center\_X, center\_Y, box\_width, box\_height = Boundbox.astype('int')

xx = int(center\_X - (box\_width / 2))

yy = int(center\_Y - (box\_height / 2))

box\_array.append([xx, yy, int(box\_width), int(box\_height)])

confidence\_array.append(float(confidence\_value))

class\_ids\_array.append(class\_id)

return box\_array, confidence\_array, class\_ids\_array

def displayImage(image):

cv.imshow("Final Image", image)

cv.waitKey(0)

**7.4 Main code**

import numpy as np

import cv2 as cv

import subprocess

import time

import os

from ObjectDetection import detectObject, displayImage

import sys

from gtts import gTTS

from playsound import playsound

import os

from threading import Thread

global class\_labels

global cnn\_model

global cnn\_layer\_names

playcount = 0

def deleteDirectory():

filelist = [ f for f in os.listdir('play') if f.endswith(".mp3") ]

for f in filelist:

os.remove(os.path.join('play', f))

def speak(data, playcount):

class PlayThread(Thread):

def \_\_init\_\_(self, data, playcount):

Thread.\_\_init\_\_(self)

self.data = data

self.playcount = playcount

def run(self):

t1 = gTTS(text=self.data, lang='en', slow=False)

t1.save("play/"+str(self.playcount)+".mp3")

playsound("play/"+str(self.playcount)+".mp3")

newthread = PlayThread(data, playcount)

newthread.start()

def loadLibraries(): #function to load yolov3 model weight and class labels

global class\_labels

global cnn\_model

global cnn\_layer\_names

class\_labels = open('model/yolov3-labels').read().strip().split('\n')

#reading labels from yolov3 model

cnn\_model=cv.dnn.readNetFromDarknet('model/yolov3.cfg', 'model/yolov3.weights')

#reading model

cnn\_layer\_names = cnn\_model.getLayerNames()

#getting layers from cnn model

cnn\_layer\_names=[cnn\_layer\_names[i[0]-1]for i in cnn\_model.getUnconnectedOutLayers()]

#assigning all layers

def detectFromImage(imagename): #function to detect object from images

#random colors to assign unique color to each label

label\_colors = np.random.randint(0,255,size=(len(class\_labels),3),dtype='uint8')

try:

image = cv.imread(imagename) #image reading

image\_height, image\_width = image.shape[:2] #converting image to two dimensional array

except:

raise 'Invalid image path'

finally:

image, \_, \_, \_, \_ = detectObject(cnn\_model, cnn\_layer\_names, image\_height, image\_width, image, label\_colors, class\_labels)#calling detection function

displayImage(image)#display image with detected objects label

def detectFromVideo(): #function to read objects from video

global playcount

#random colors to assign unique color to each label

label\_colors = np.random.randint(0,255,size=(len(class\_labels),3),dtype='uint8')

try:

video = cv.VideoCapture(0)

frame\_height, frame\_width = None, None #reading video from given path

except:

raise 'Unable to load video'

finally:

while True:

frame\_grabbed, frames = video.read() #taking each frame from videoz

if not frame\_grabbed: #condition to check whether video loaded or not

break

if frame\_width is None or frame\_height is None:

frame\_height, frame\_width = frames.shape[:2] #detecting object from frame

frames, cls, \_, \_, \_, \_ = detectObject(cnn\_model, cnn\_layer\_names, frame\_height, frame\_width, frames, label\_colors, class\_labels)

data = ""

if len(cls) > 0:

for i in range(len(cls)):

data+=cls[i]+","

cv.imshow("Detected Objects", frames)

if len(cls) > 0:

speak("Detected Objects = "+data, playcount)

playcount = playcount + 1

if cv.waitKey(5) & 0xFF == ord('q'):

break

if \_\_name\_\_ == '\_\_main\_\_':

loadLibraries()

deleteDirectory()

detectFromVideo()

# SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

**8.1 UNIT TESTING**

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

## 8.2 INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

## 8.3 FUNCTIONAL TEST

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be

Accepted.

Invalid Input : identified classes of invalid input must be

rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs

must be exercised.

. Systems/Procedures : interfacing systems or procedures must be Invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

## 

## 8.4 SYSTEM TEST

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

## 8.5 WHITE BOX TESTING

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

## 8.6 BLACK BOX TESTING

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it. The test provides inputs and responds to outputs without considering how the software work

# INPUT DESIGN AND OUTPUT DESIGN

## 9.1 INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

1. What data should be given as input?
2. How the data should be arranged or coded?
3. The dialog to guide the operating personnel in providing input.
4. Methods for preparing input validations and steps to follow when error occur.

### 9.1.1 OBJECTIVES

Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus, the objective of input design is to create an input layout that is easy to follow.

## 9.2 OUTPUT DESIGN:

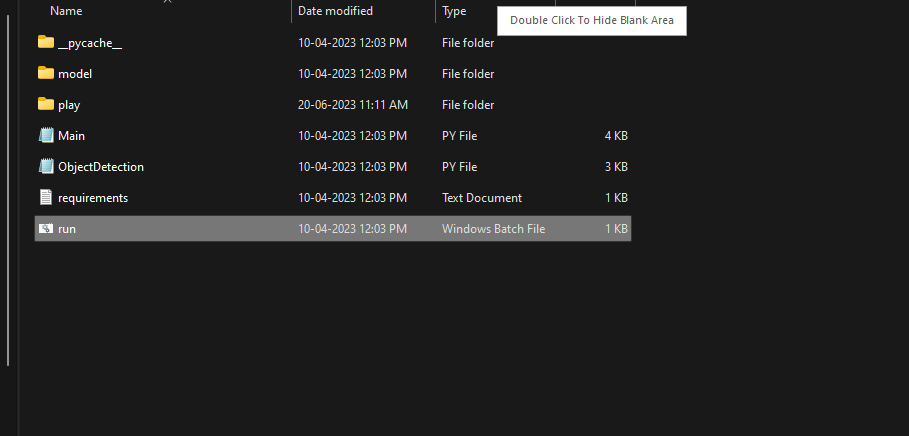
A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system’s relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each Output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system. The output form of an information system should accomplish one or more of the following objectives.

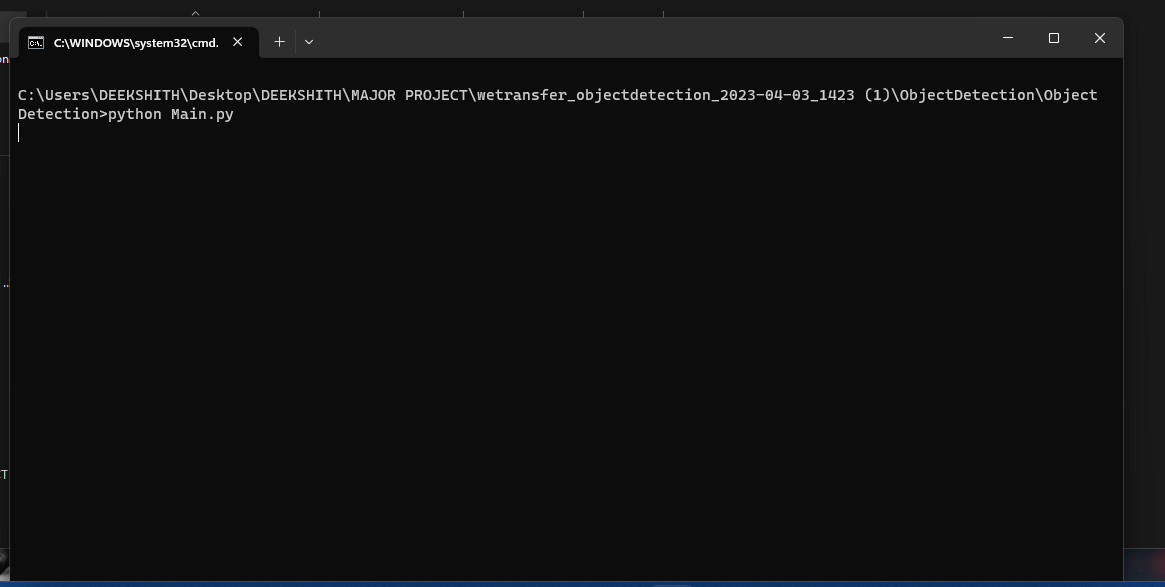
* Convey information about past activities, current status or projections of the Future.
* Signal important events, opportunities, problems, or warnings.
* Trigger an action.
* Confirm an action.

# SCREENSHOTS

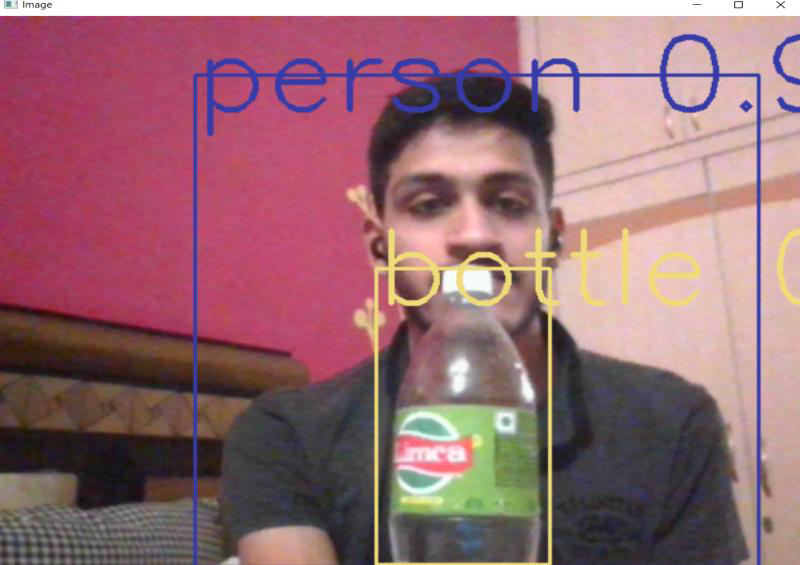
To run required project just double click on **‘run.bat’** file to get below screen



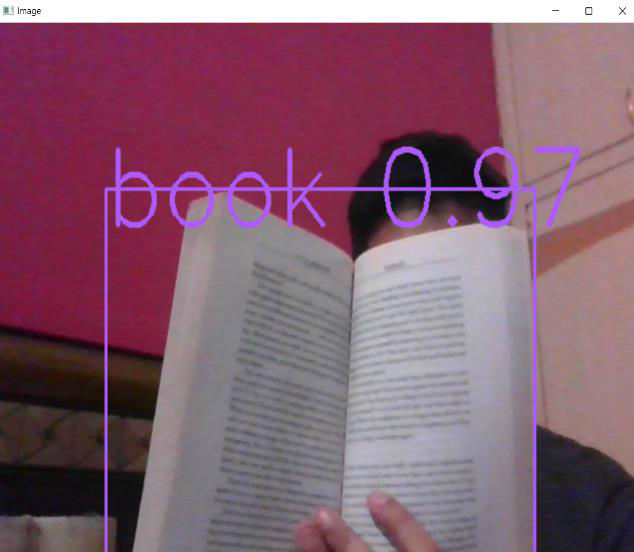
**Fig 10.1 run.bat file**



**Fig 10.2 Opening run.bat file**



**Fig 10.3 Detecting Bottle And Person**



**Fig 10.4 Detecting Book**

# 11.CONCLUSION

In this project we used image recognition, voice generation modules for the development of the project. As of now accuracy is good but in case if we want to increase the accuracy we have to train the model with more object/images in the dataset. This project is a small experiment which is useful for blind persons, can be able to find the objects which are surrounded by them, and they are in a position of taking care of themselves when they are outside. The ability of the blind person to stand alone and carry out tasks independently makes this blind assistance device useful for object detection by voice warnings. The device's camera serves as the blind person\'s virtual eye, capturing every detail of their environment. The voice alerts keep the person informed about the surroundings so that accidents are decreased. Reduced rely on other parties. There are so many people present in the world who are visually impaired and illiterate from different parts of the world .Some of them do not understand other languages except their local language in their local accent . so, one of the future scope for this project is to develop the idea in which voice alerts in such a way that they can use their own local language .

# FUTURE WORK

This study summarizes a feasible approach of real-time object detection and classiﬁcation from input video based on COCO dataset. As we focused on both high accuracy and low computational complexity, the expected result is obtained by applying YOLO model and TensorFlow API. This research opens the opportunity of detecting real-time objects in a very effective way combining both SSD for object detection and YOLO for object recognition. Our developed system also provides some quality enhancements to ensure faster response while object detection is performed on the camera. After analyzing the outcome, we can conclude that the proposed system gives better accuracy in real-time object detection even with low light and occlusion. We have also considered map value to measure a comparative quantitative performance of different deep learning-based techniques in detecting objects. Here, we see that our method gives satisfactory results in case of map. Besides, our system still needs improvement in accuracy. So, our intention is to apply region-based CNN method in detecting objects for ﬁne- tuning the model in the future. We hope this will increase the accuracy of object detection and work better for detecting small objects.

*  To improve the accuracy of the model.
*  To read texts and time from digital clock and handwriting recognition and give

speech output.

*  Face Recognition, capturing and saving of data should be implemented.
*  Using GPS and to make its use safe outdoors such as roads etc.
*  Also to make the model to give the description of the object if required.
*  Implementation of Hardware using Raspberry Pi.

# BIBLIOGRAPHY

[1] F. Jabeen, and A. M. Enriquez, 2015, “Feed forward neural network training based interactive shopping for blind,”, IEEE, pp. 1–6.

[2] R. Tapu and T. Zaharia, 2013, “A smart phone-based obstacle detection and classification system for assisting visually impaired people,”, pp. 444–451.

[3] R. Tapu, and E. Tapu, 2014, “A survey on wearable devices used to assist the visual impaired user navigation in outdoor environments,”, IEEE, pp. 1–4.

[4] J. Monteiro, R. Granada, 2017, “Virtual guide dog: An application to support visually-impaired people through deep convolutional neural networks,”, IEEE, pp. 2267–2274.

[5] R. Kumar and S. Meher, 2015, “A novel method for visually impaired using object recognition,”, IEEE, pp. 0772–0776.

[6] Tapu R, Mocanu B, 2017, “Seeing without sight-an automatic cognition system dedicated to blind and visually impaired people,”, pp. 1452–1459.

[7] B.-S. Lin, and P.-Y. Chiang, 2017, “Simple smartphone-based guiding system for visually impaired people”, p. 1371, 2017.

[8] C. D. Pai and K. Potdar, 2018, “A convolutional neural network based live object recognition system as blind aid,”.

[9] J.-C. Yang L, 2018, “A deep learning approach to sensory navigation device for blind guidance,”, IEEE, 2018, pp. 1195–1200.

[10] Y. Bazi and H. Alhichri, 2019, “Helping the visually impaired see via image multi-labeling based on squeezenet cnn,”, p. 4656, 2019.